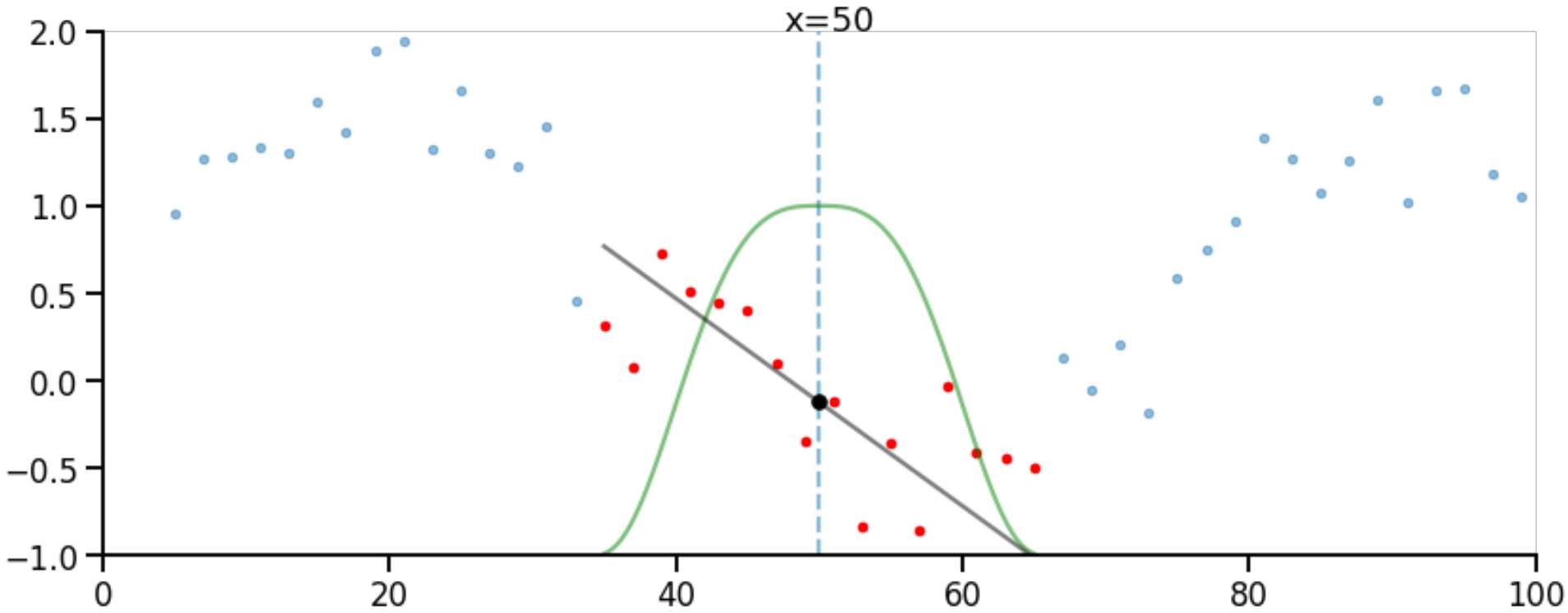
LOWESS (Practice)

Time series decomposition

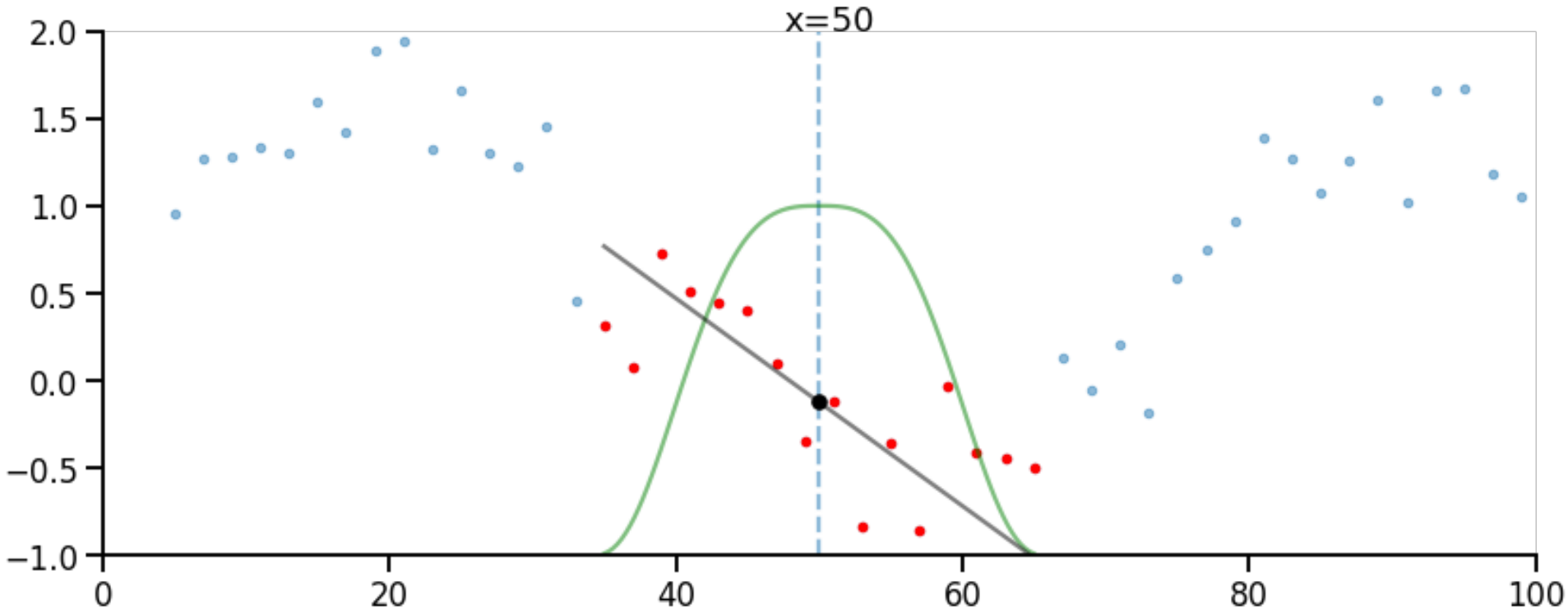
LOWESS

- LOWESS curve at x is given by the weighted robust linear regression



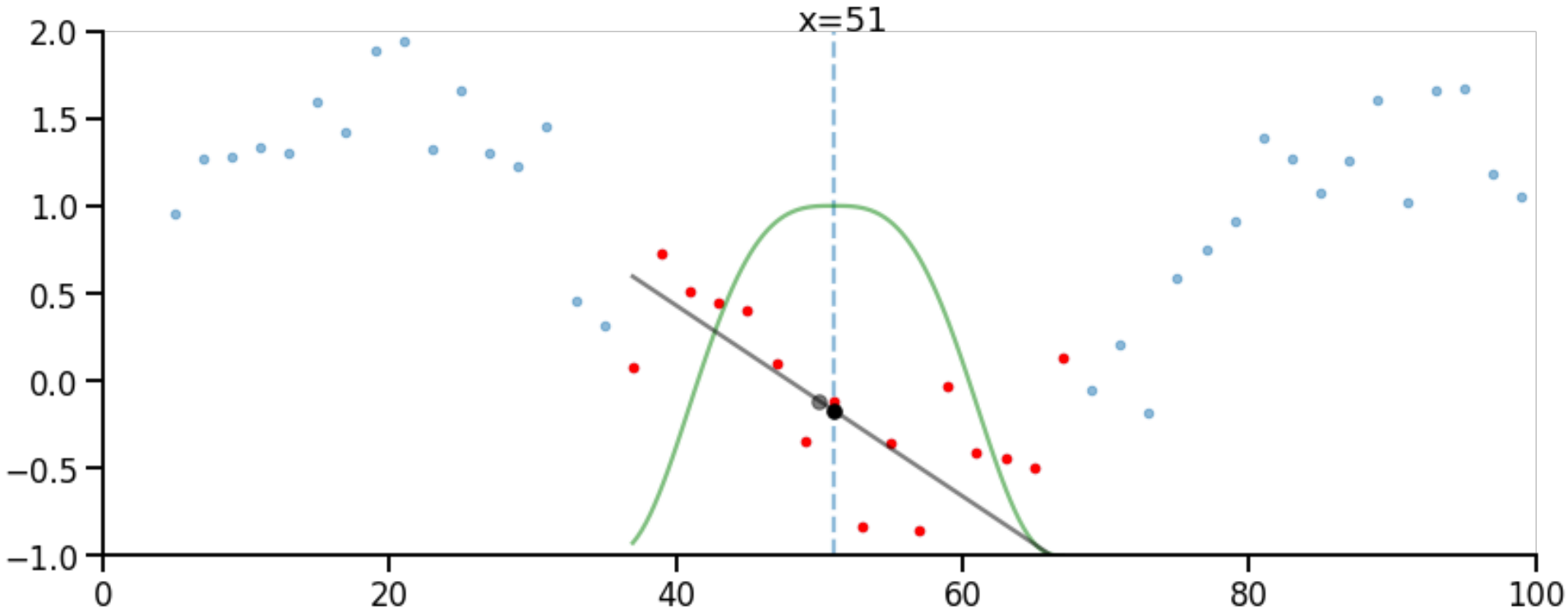
LOWESS

- Evaluate the same process across many x values to obtain a smooth fit



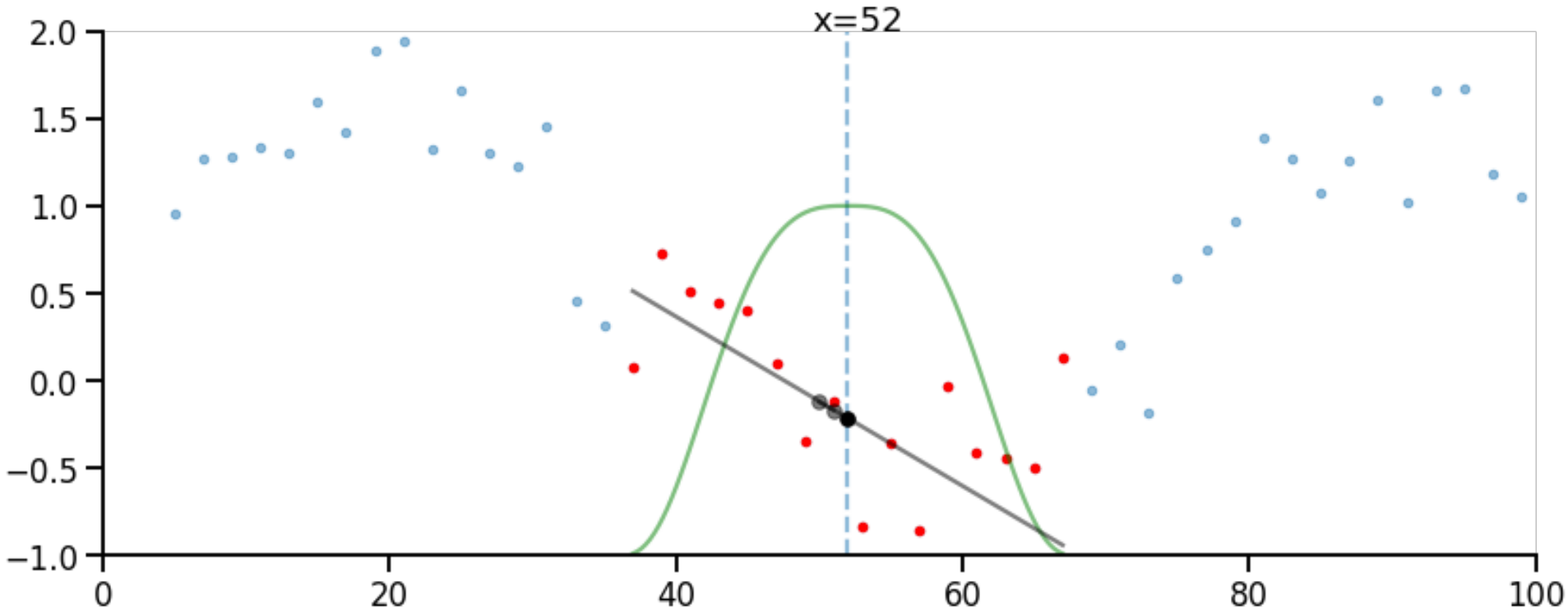
LOWESS

- Evaluate the same process across many x values to obtain a smooth fit



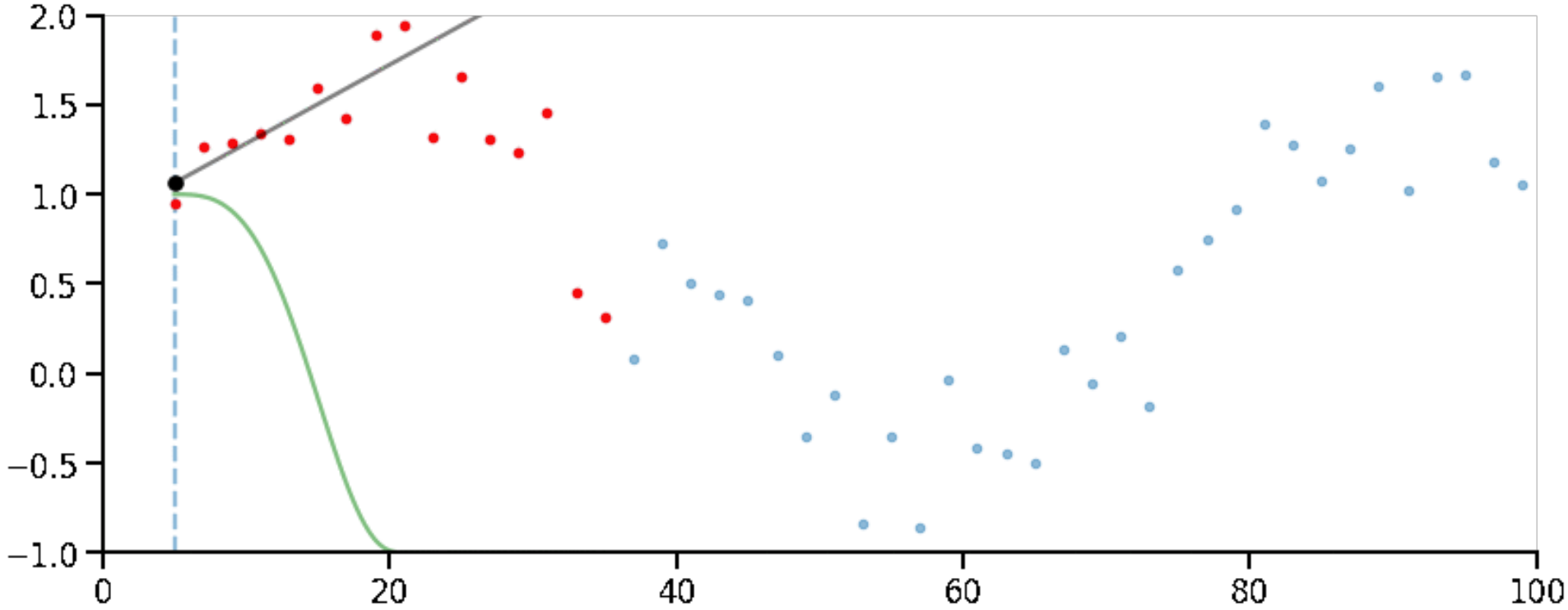
LOWESS

- Evaluate the same process across many x values to obtain a smooth fit



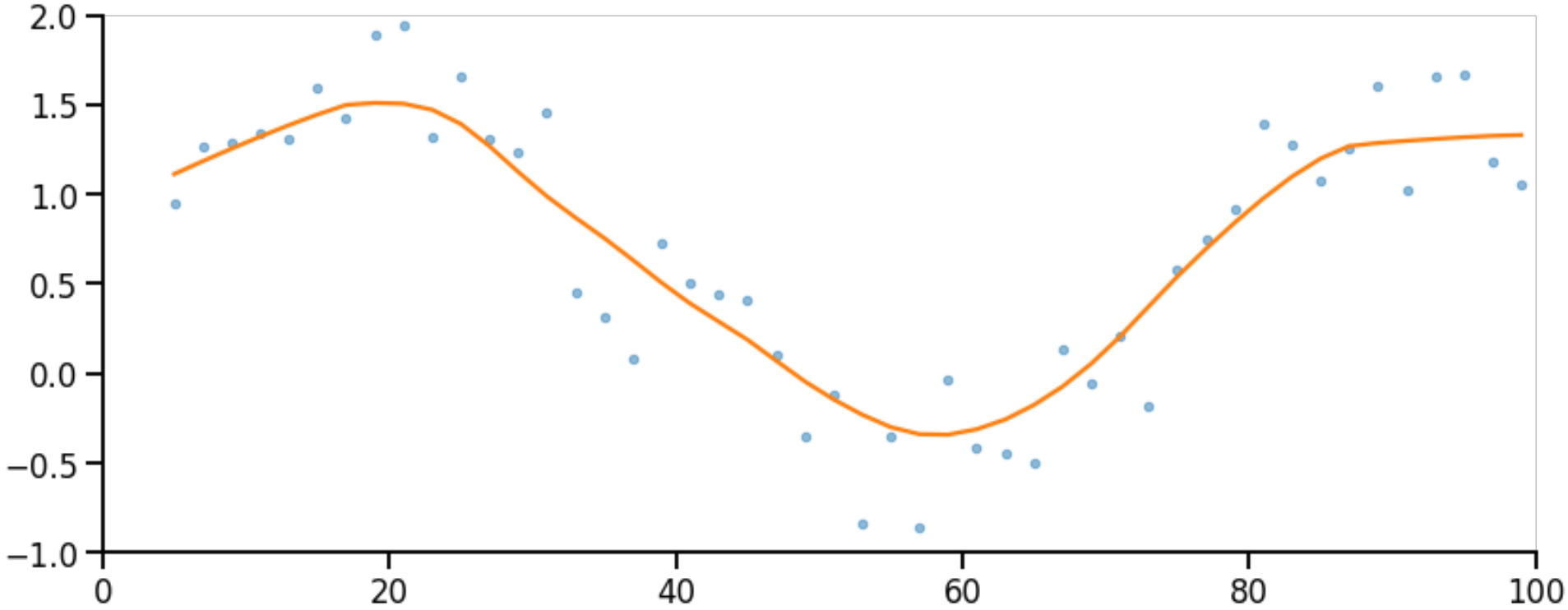
LOWESS

- Evaluate the same process across many x values to obtain a smooth fit



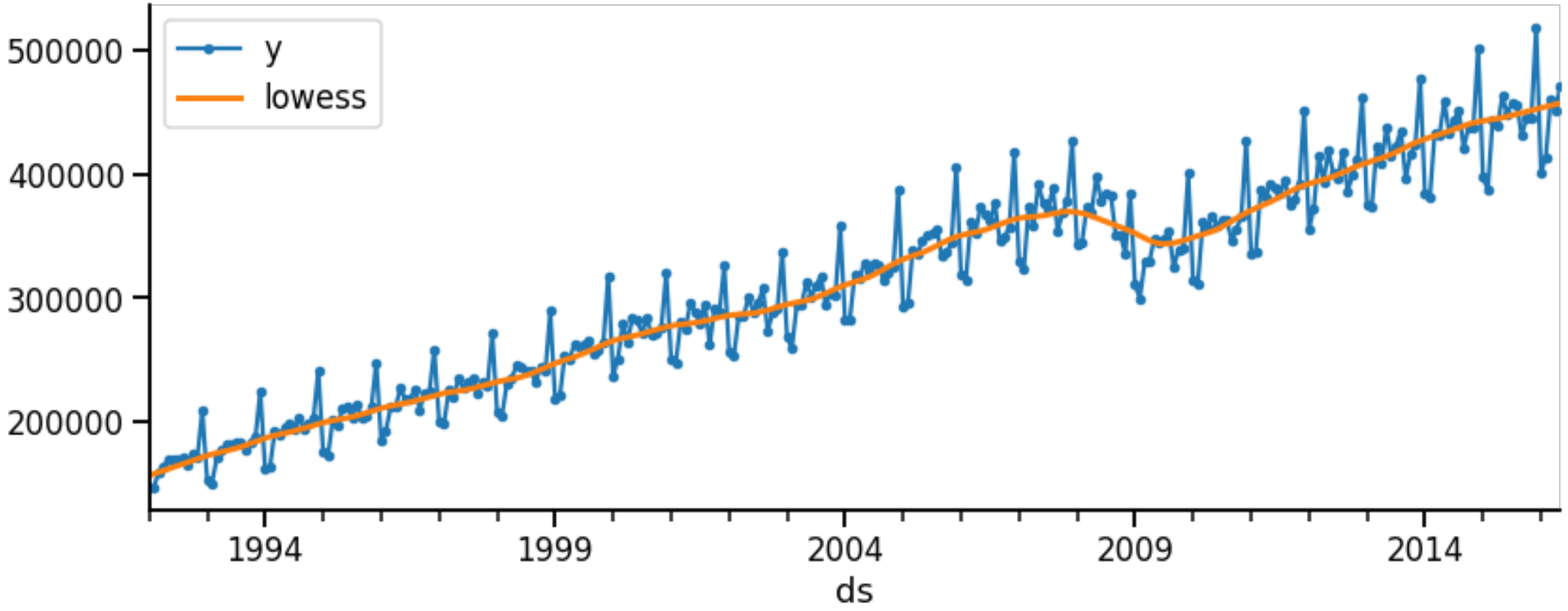
LOWESS

- Evaluate the same process across many x values to obtain a smooth fit



LOWESS

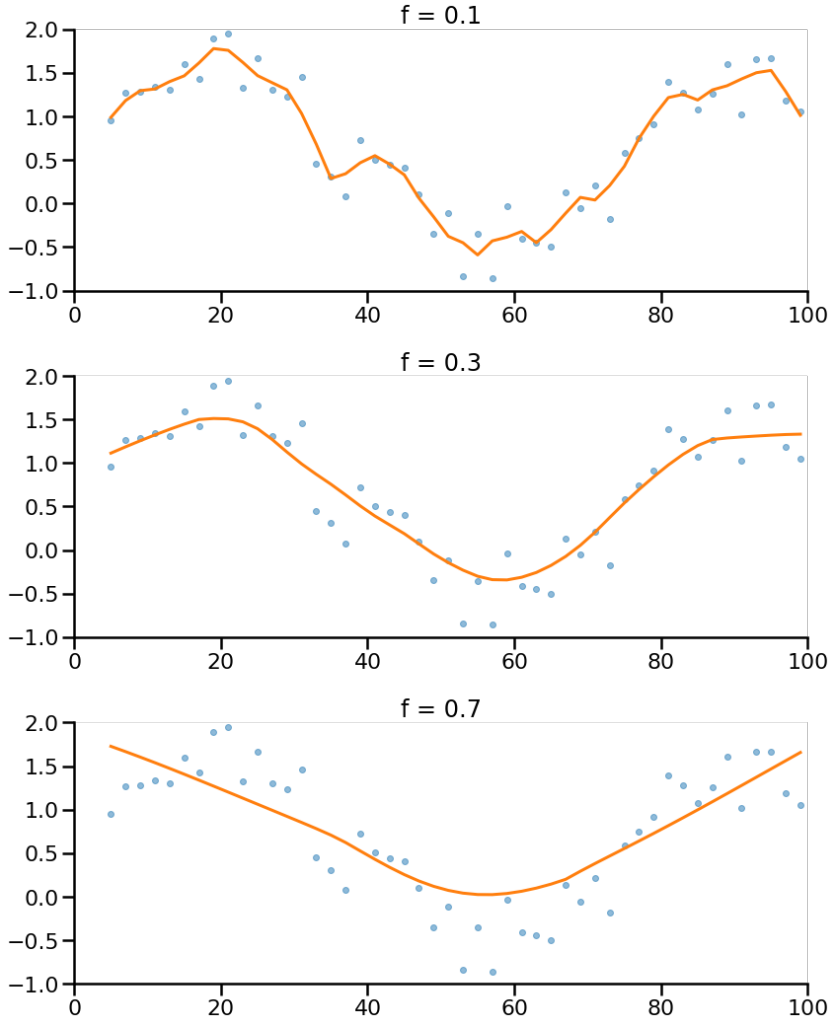
- The LOWESS fit can be used as an estimate of the trend of a time series



Parameters

- f: Fraction of data for window size
  - Determines smoothness of the fit
  - Good value depends on data
  - Trying to have as smooth a fit without distorting patterns in the data
  - Can set visually or using cross-validation

- i: Number of iterations for robust regression
  - Ensures robustness to outliers
  - Typically set to 2 or 3 [1]



[1] Cleveland, William S. "Robust locally weighted regression and smoothing scatterplots." *Journal of the American statistical association* 74.368 (1979): 829-836.

Implementation

statsmodels.nonparametric.smoothers_lowess.lowess

`statsmodels.nonparametric.smoothers_lowess.lowess`(*endog, exog, frac=0.6666666666666666, it=3, delta=0.0, xvals=None, is_sorted=False, missing='drop', return_sorted=True*)[source]

LOWESS (Locally Weighted Scatterplot Smoothing)

A lowess function that outs smoothed estimates of endog at the given exog values from points (exog, endog)

Parameters

endog : 1-D `numpy array`

The y-values of the observed points

exog : 1-D `numpy array`

The x-values of the observed points

frac : `float`

Between 0 and 1. The fraction of the data used when estimating each y-value.

it : `int`

The number of residual-based reweightings to perform.

```
# The input of lowess need to be numpy arrays and a numeric type

y = df['y']
x = np.arange(0, len(y)) # datetime is not a numeric type
                         # so we use a sequence of integers
                         # for the x-values for the purpose of fitting

# lowess returns an array where the first column
# are the x values and the second column are the
# values of the fit at those x values

res = lowess(endog=y, # the y values
             exog=x,  # the corresponding x values
             frac=0.1, # fraction of dataset to use in window
             it=3, # Number of iterations for robust regression.
                   # The default value of 3 is typically good enough [1].
            )

# Append the lowess curve to the original dataframe
df['lowess'] = res[:,1]

df.head()
```

ds	y	lowess
1992-01-01	146376	156224.385686
1992-02-01	147079	157578.127266
1992-03-01	159336	158916.267408
1992-04-01	163669	160244.261688
1992-05-01	170068	161568.493429

Discussion

Pros

- Robust to outliers
- No missing data at the edges
- Interpretable
- A good f will avoid over-smoothing the data
- No assumptions made about the data

Cons

- Slow to fit on large datasets
- Requires access to entire dataset to evaluate curve at any point
- Selecting f often requires manual inspection

Summary

LOWESS is a method to fit a smooth curve to a scatter plot

LOWESS can be used to extract the trend term in a time series