Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

Given a sequence of observations at equal increments of time or space, etc. a smoothed curve can be fitted in several alternative ways using running medians and/or moving averages. From the main SimFIT menu choose [Statistics], [Time series], [Data smoothing], then study the test file provided which is `g10caf.tf1`.

```
569
416
422
565
484
520
573
518
501
505
468
382
310
334
359
372
439
446
349
395
461
511
583
590
620
578
534
631
600
438
516
534
467
457
392
467
500
493
410
412
416
403
422
459
467
512
534
552
545
```

These are measurements of coal production in millions of tons per year in the USA from 1920 to 1968, and SimFIT provides the following smoothing options.

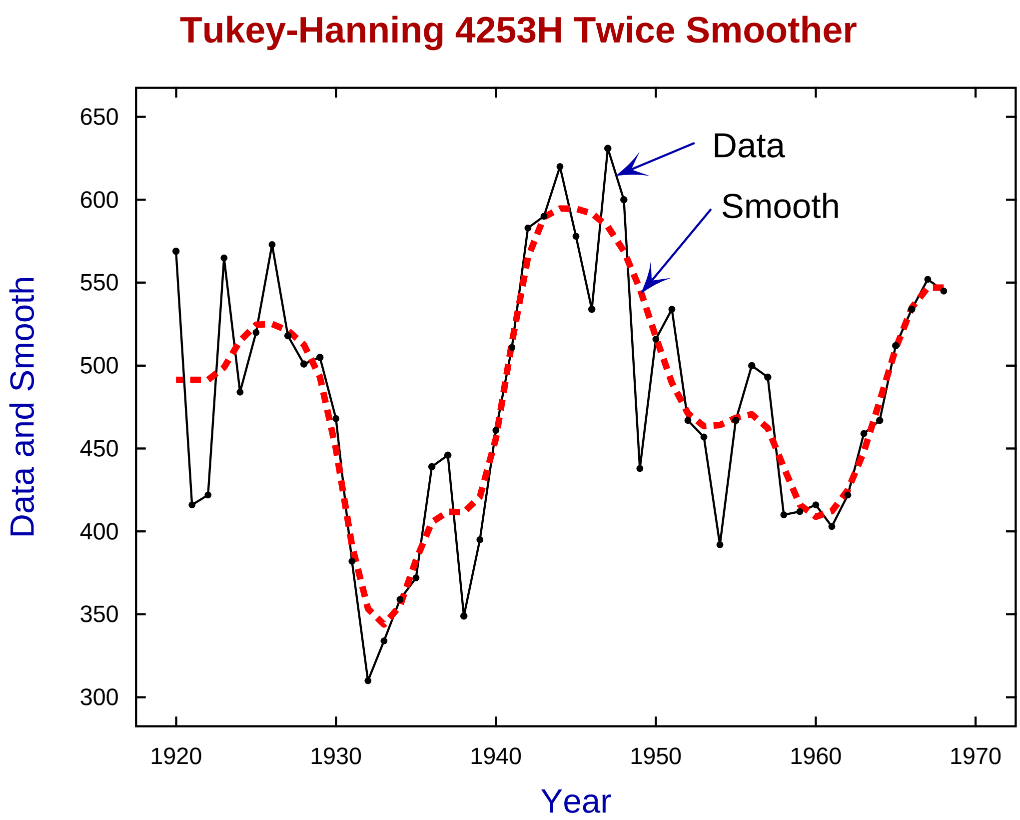
1. Running medians with span 4 then 2
2. Running medians with span 5
3. Running medians with span 3
4. Moving averages with span 3 using Hanning
5. The Tukey-Hanning 4253H twice smoother

The result of smoothing is to create a smoothed fit together with residuals, known as rough, that is

Data = Smooth + Rough

and the usual options are available to analyze and plot the smoothed fit and residuals.

The first four options may be investigated, especially with sparse data, but for most purposes the Tukey-Hanning 4253H twice smoother would be preferred as illustrated in the next graph.



Theory

1. **Running medians of even span (e.g. 4 followed by 2)**

   These are applied in pairs as there are complications due to lack of symmetry. For instance, a span of four generates n + 1 smooths from a sample size n as follows.

$$
\begin{aligned}
z_1 &= y_1 \\
z_2 &= (y_1 + y_2)/2 \\
z_3 &= \text{median}(y_1, y_2, y_3, y_4) \\
\ldots \\
z_{n-1} &= (y_{n-2} + y_{n-1})/2 \\
z_n &= (y_{n-1} + y_n)/2 \\
z_{n+1} &= y_n
\end{aligned}
$$

   Following this by a span of two restores the dimension of smooths to n in the following way.

$$
\begin{aligned}
z_1* &= (z_1 + z_2)/2 \\
z_2* &= (z_2 + z_3)/2 \\
\ldots \\
z_{n-1}* &= (z_{n-1} + z_{n-2})/2 \\
z_n* &= (z_n + z_{n+1})/2
\end{aligned}
$$

2. **Running medians of odd span**

   For instance a span of five proceeds as follows

$$
\begin{aligned}
z_1 &= y_1 \\
z_2 &= \text{median}(y_1, y_2, y_3) \\
z_3 &= \text{median}(y_1, y_2, y_3, y_4, y_5) \\
\ldots \\
z_n &= y_n
\end{aligned}
$$

   while a span of three is simply

$$
\begin{aligned}
z_1 &= y_1 \\
z_2 &= \text{median}(y_1, y_2, y_3) \\
z_3 &= \text{median}(y_2, y_3, y_4) \\
\ldots \\
z_n &= y_n
\end{aligned}
$$

3. **Moving averages**

   This usually involves a weighting scheme as with the Hanning of span three with binomial weights.

$$
\begin{aligned}
z_1 &= y_1 \\
z_2 &= 0.25y_1 + 0.5y_2 + 0.25y_3 \\
\ldots \\
z_n &= y_n
\end{aligned}
$$

4. **Endpoint rules**

   To correct for singular behavior at the extreme points SimFIT uses the correction

$$
\begin{aligned}
z_1 &= \text{median}(3z_2 - 2z_3, z_1, z_2)\\
z_n &= \text{median}(3z_{n-2} - 2z_{n-1}, z_n, z_{n-1})
\end{aligned}
$$

5. **The 4253H twice smoother**

   The previous simple smoothers may be of some use with sparse data, but the most widely recommended smoother is the Tukey-Hanning 4253H twice smoother. This applies running medians of four, then two, then five, then 3, followed by a Hanning filter of span 3. The rough (i.e. residuals) are then calculated but are then also subjected to the same sequence before being added to the smooth from the first pass, followed by calculating new rough.