JMP Discovery Summit 2023

Using Graph Builder's Splines To Align Measurement Curves

Bernd Heinen
Statistician (Msc)
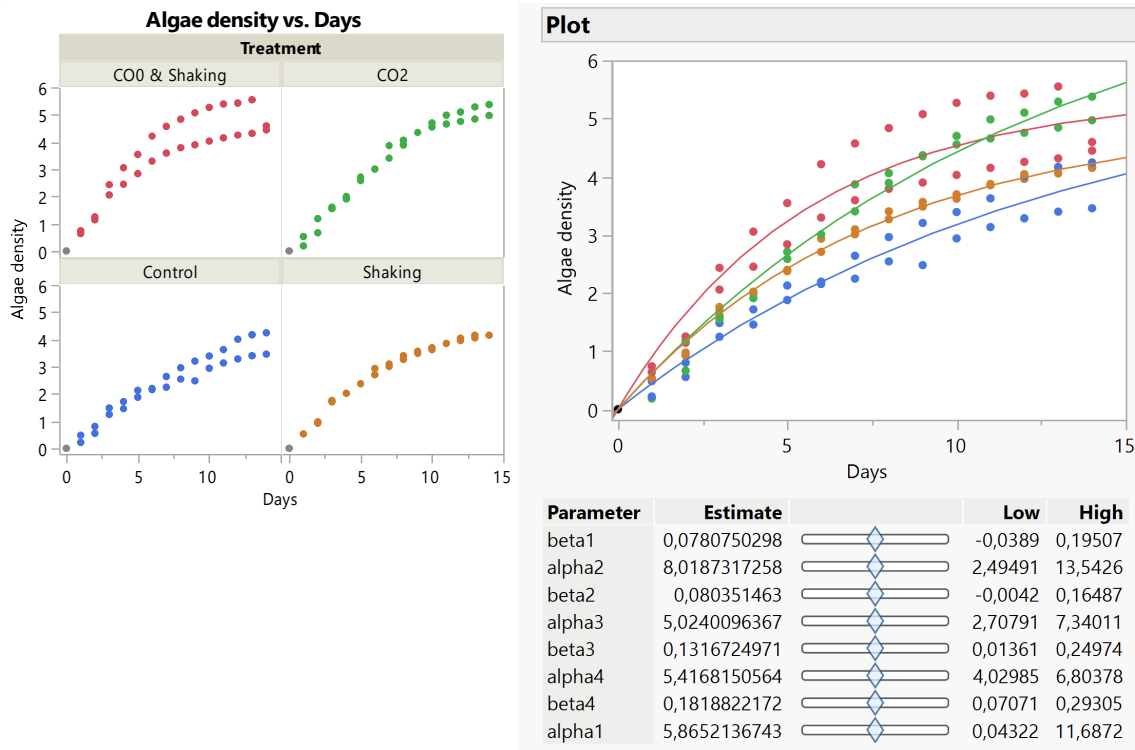Consultant, Coach and Trainer
bernd.heinen@stabero.com

Abstract

A frequent task in data analysis is aligning curves before a descriptive or root cause analysis. Often an additional complication occurs when the measurement intervals are not equidistant in the series to be compared. There is not one single value that quantifies the shift for a whole curve. Interpolation is the solution in cases like this. Simple linear interpolation may lead to numerous random errors; a spline interpolation is more robust. Since the Graph Builder exports the formula for the spline in its current shape, it became an easy, accessible tool for the alignment of curves. And as all steps can be programmed in JSL, it provides a framework for automating curve alignment. This presentation will describe the background, concept, and case study application for the alignment of curves.

In General

Curves are of interest in many different conditions and JMP has a lot of functionality to address all kinds of analytical tasks.

One example where the curve itself is the object of interest for the analyst:

Parameter	Estimate		Low	High
beta1	0,0780750298		-0,0389	0,19507
alpha2	8,0187317258		2,49491	13,5426
beta2	0,080351463		-0,0042	0,16487
alpha3	5,0240096367		2,70791	7,34011
beta3	0,1316724971		0,01361	0,24974
alpha4	5,4168150564		4,02985	6,80378
beta4	0,1818822172		0,07071	0,29305
alpha1	5,8652136743		0,04322	11,6872

These are growth curves for Algae development under different conditions. Data is from JMP's sample library.

A predefined or a custom formula can be fitted and very often the parameters of these formulas are of interest, because they have a technical meaning. They can be interpreted like growth rate, inflection point, limit or more.

JMP has the platforms

● Fit Curve
● Nonlinear
● Functional Data Explorer (Pro)

to address those problems.

Sometimes, curves are only used for visualizing trends or for interpolation. The last case is the scenario that shall now be discussed.

The Use Case

In an experiment two different devices are used to measure a variety of substances. Unfortunately, the clocks of the devices can not be synchronized. Each of the devices has one sensor that measures the same substance. The measurement series of these sensors can be used to synchronize all readings.
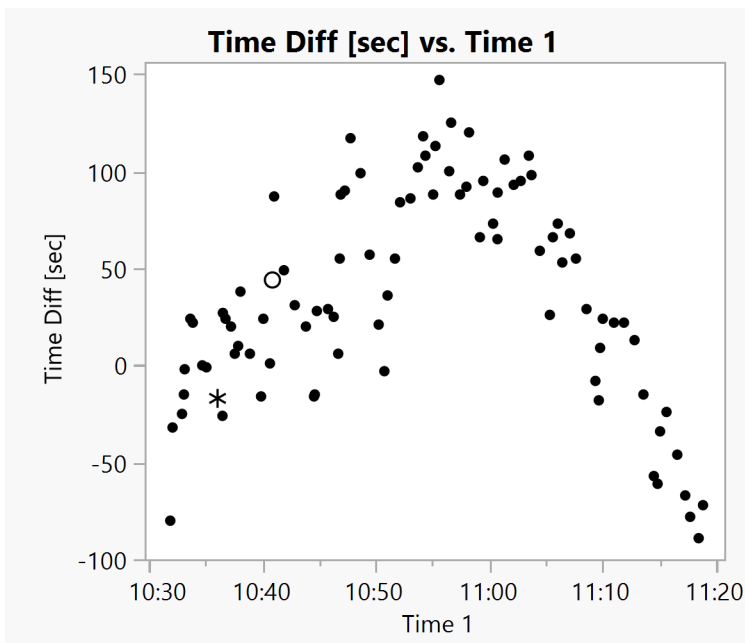
The data:
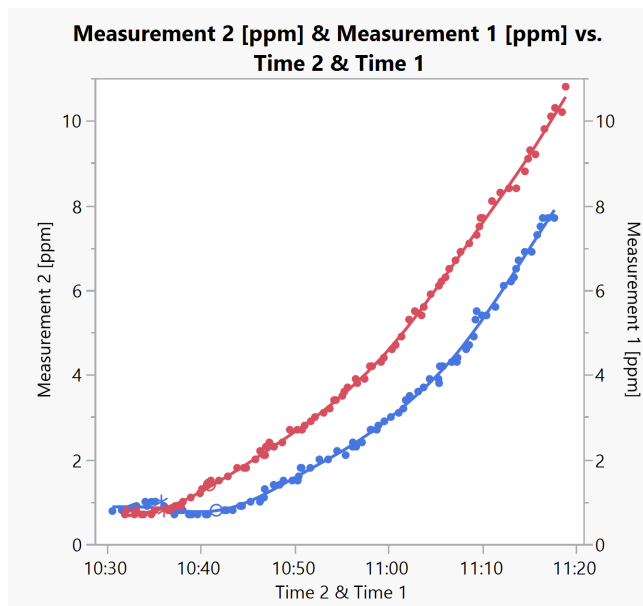
The red curve is the reference, the clock of the blue curve is delayed. The lines are not parallel and the time intervals are not constant, so that a simple time shift wouldn't work.

Is there another relationship between the time points?

Time 2 vs. Time 1
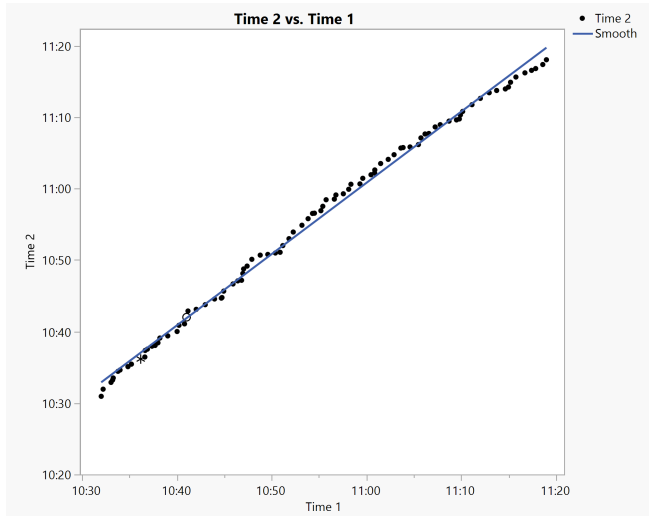
There are only very few data points on the regression line. Deviations occur in both directions. Obviously, there is no stringent relationship between the two time series.

Perhaps we cannot see a pattern with the bare eye. It may be easier to spot something if we look at the time difference within the rows of the data table:

Time Diff [sec] vs. Time 1

It gives the impression of a curve, but with large random variation. It is clearly NOT pointing out a simple time shift or other pattern.

What can we see in the data table?

This is a picture of the data table:

Beginning with row 22 "Measurement 2" takes the same development than "Measurement 1" starting in row 10.

The solution we came up with, is a "Paternoster shift".
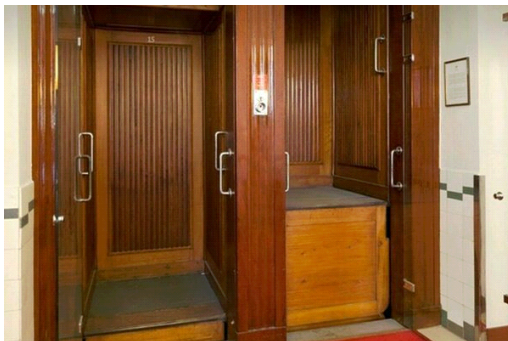
The two measurement series have nothing in common. If we want to shift one column "up" until the two series are aligned to the best extend possible, we need to answer two questions:

● How do we compare the two series?
● How many rows do we need to shift?

As there are no common time points we need to do some kind of interpolation.
Linear interpolation is the first thing that comes to mind.

Measurement 2 [ppm] & Measurement 1 [ppm] vs. Time 2 & Time 1
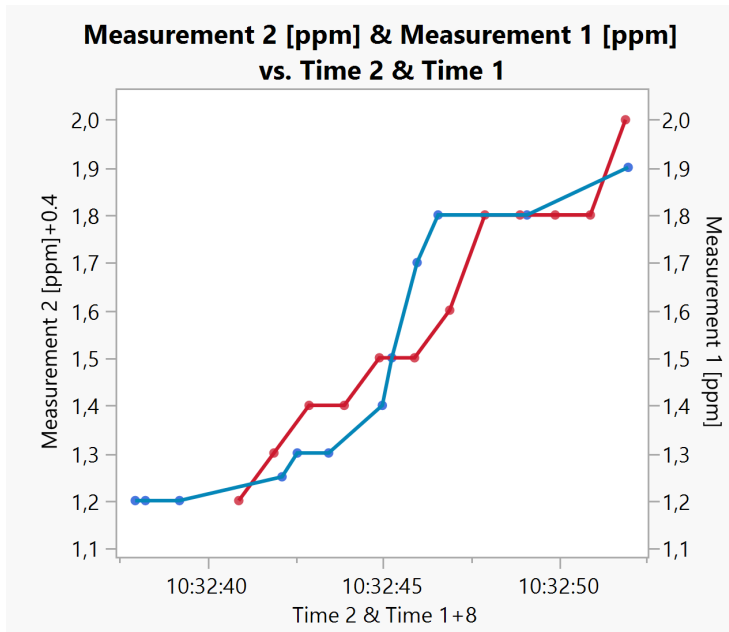
When we zoom into an arbitrary part of the interpolated curves, the disadvantages of linear interpolation become obvious. Where the interpolation builds a horizontal line, the reference to the X-value, usually time, is not unique. Where the ascent is very steep, small changes in the X-value result in large differences in Y.

Spline interpolation offers an elegant way to overcome these weaknesses.

Splines do not only interpolate between neighbors, they consider points farther away as well and they deliver smooth functions.

Fitting Splines in JMP

There are some platforms that allow fitting splines. These are:

● Fit Y by X

Fit Y by X lets you adapt the spline through a slider pretty much like the Smoother in Graph Builder. This platform exports the predicted values but not the prediction formula. Therefore it can't be used for the interpolation in the current task

● Fit Model

(knotted Spline Effect)
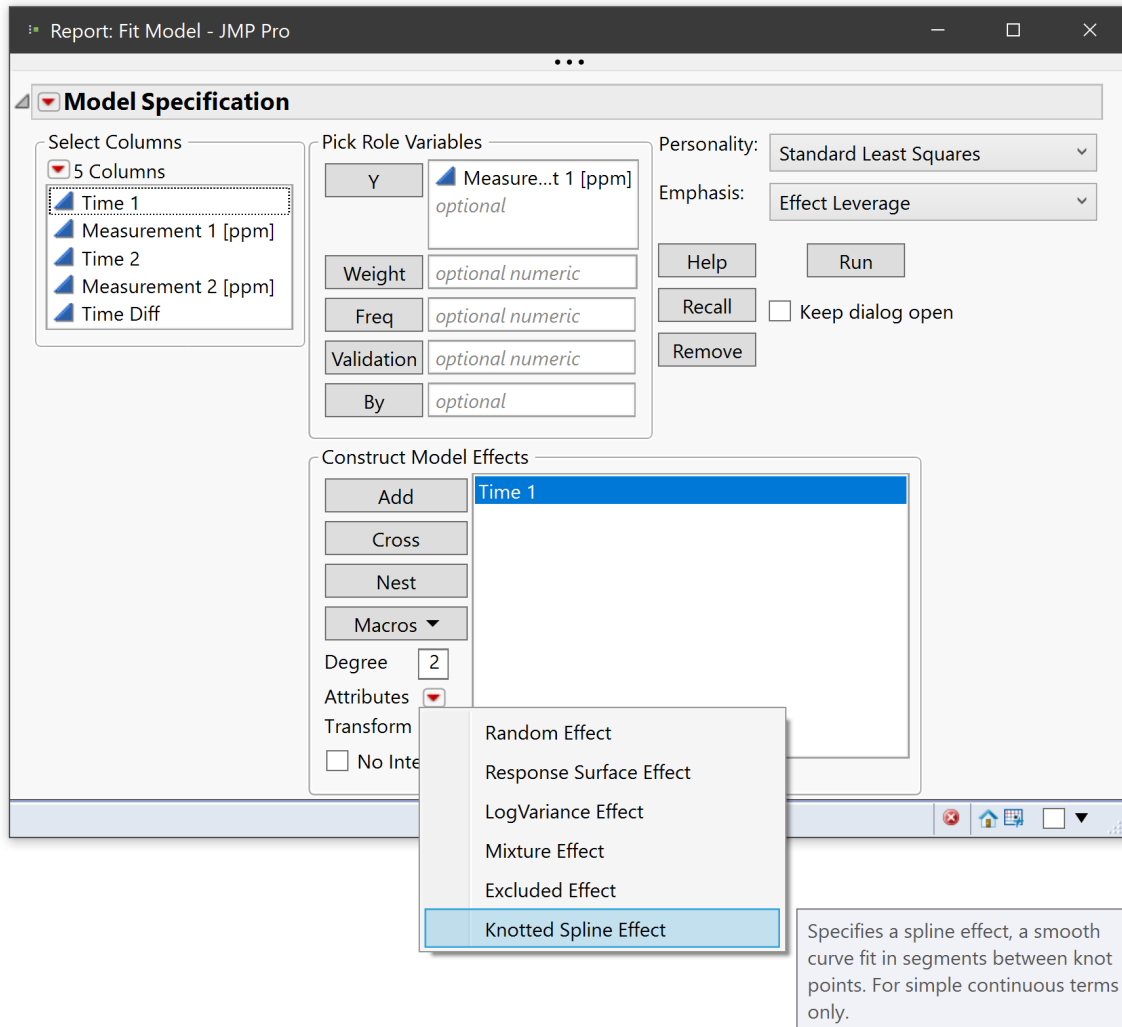
In Fit Model an effect can get the attribute "Knotted Spline Effect", and if one uses it, a window pops up that lets you specify the number of knots to be used. The higher the number, the closer the spline gets to the data. Fit Model is the only analytic platform that lets you export the spline function. Its disadvantage is that one needs to specify the number of knots before running it.
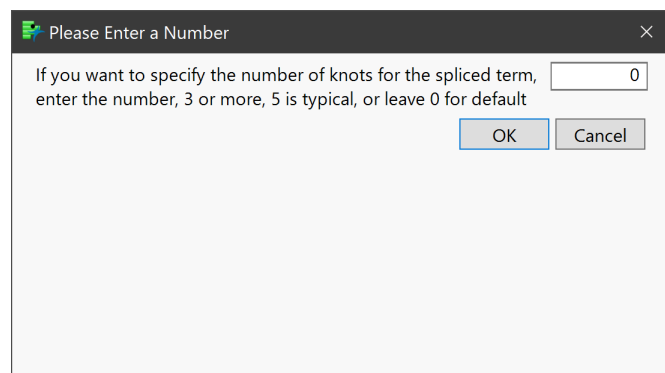
● Functional Data Explorer (Pro)

Fitting splines is the core of the Functional Data Explorer. It comprises a search mode for the best spline fit and also exports the spline function. But fine tuning the fits and exporting the function requires a few clicks and is not so intuitive as in other platforms. Clearly, the Functional Data Explorer is made for other tasks and it is not available for everybody.

Remains the Graph Builder
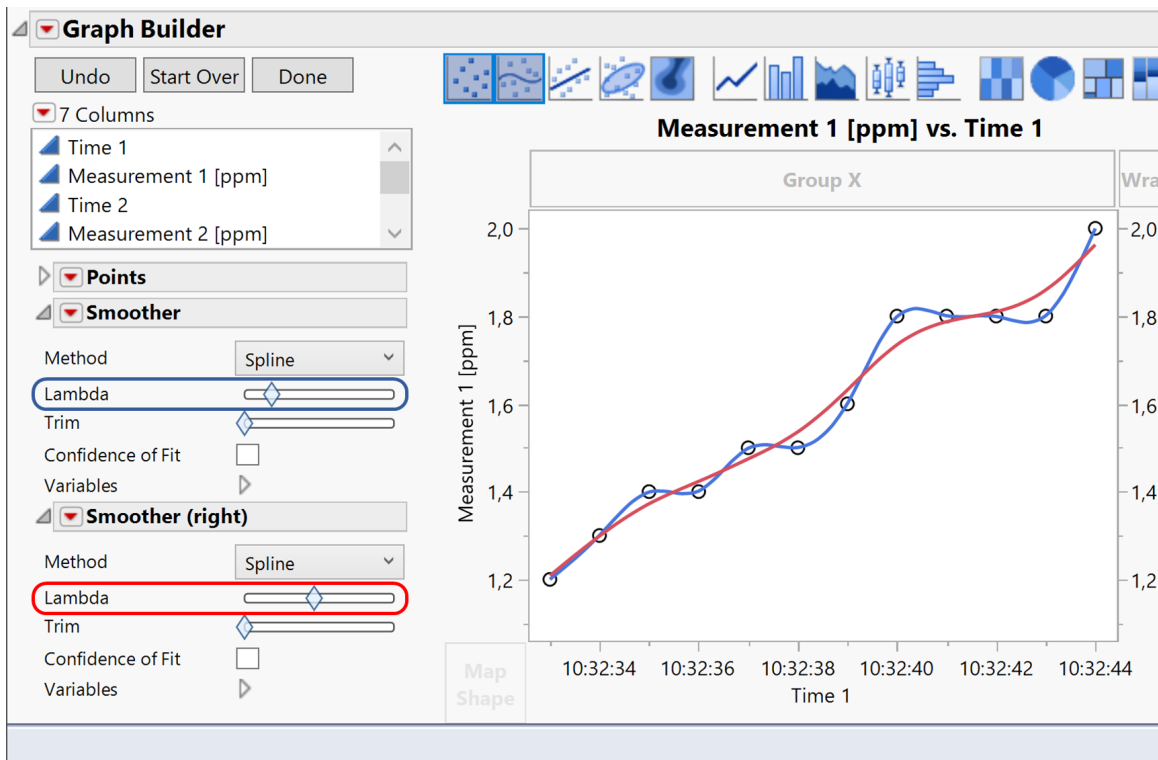
The Graph Builder surprises as an easy to use modeling platform.

The Lambda slider lets you determine the degree of fit. When you check the Confidence of Fit box, a confidence band is plotted, based on a bootstrap procedure. It is a help in determining an appropriate degree of smoothness: the spline should be within its own confidence band.

The Smoother's hot spot offers the option to Save the Formula and if the confidence box is checked, the confidence limits will be saved, too. In the data table column of the X variable (time) an arbitrary value (within the range of observations) can be appended and the interpolated value will be calculated.

How does this help for the alignment?

Splines could be fit for both measurements. Some arbitrary time values within the given range can be set and the time for one series can be changed until there is only a minimal difference between the measurement values.

A scripted solution

For the current case the data should not be changed but the time series be shifted relative to each other. And, a manual solution was no option. As a stopping rule for the stepwise move the absolute value of the cumulated differences is used If it reaches a minimum the process stops. All elements needed can be scripted in JSL. Here is a list of respective commands:

```
Names Default To Here(1);
dt = Data Table( "Interpolation Example" );

//*** Set up the graph ***
gb = dt << Graph Builder(
        Size( 531, 456 ),
        Show Control Panel( 0 ),
        Variables( X( :Time 1 ), Y( :"Measurement 1 [ppm]"n ) ),
        Elements( Points( X, Y, Legend( 9 ) ),
        Smoother( X, Y, Legend( 10 ), Lambda( 0.01 ) ) )
);

//*** Send comands to the Smoother ***
// Change the stiffness of the Smoother
gb << update element( 1, 1, 2, {Lambda( 0.1 )} );
/ Create formula column(s) in the data table
gb << update element( 1, 1, 2, {Save Formula()} ); /

//*** Get information from the current graph        ***
//*** (allows to get results from user interaction) ***
se = gb << Get Element( 1, 1, 2);// Get current specifications of the Smoother
lambda = Parse(Word(2, Char(se[5]), "()"));   // Extracts the value of lambda
```
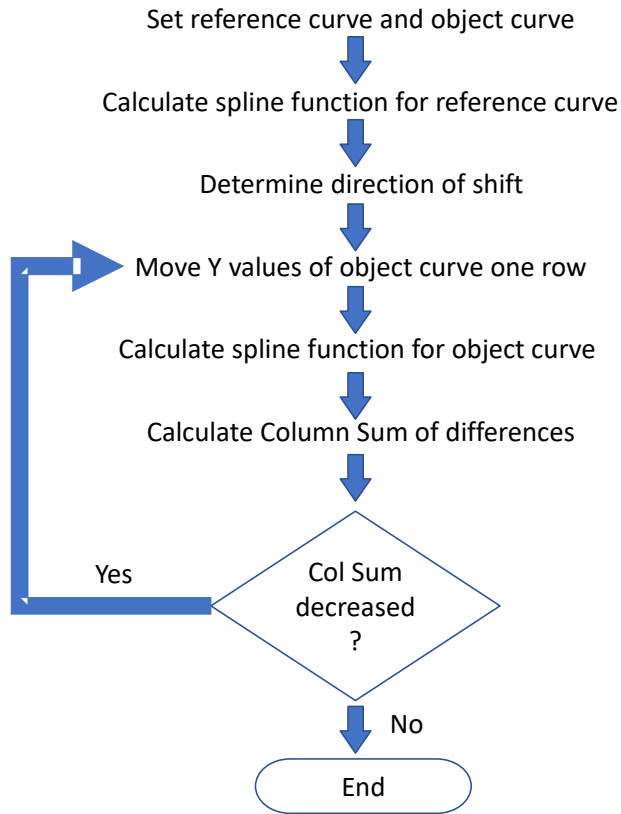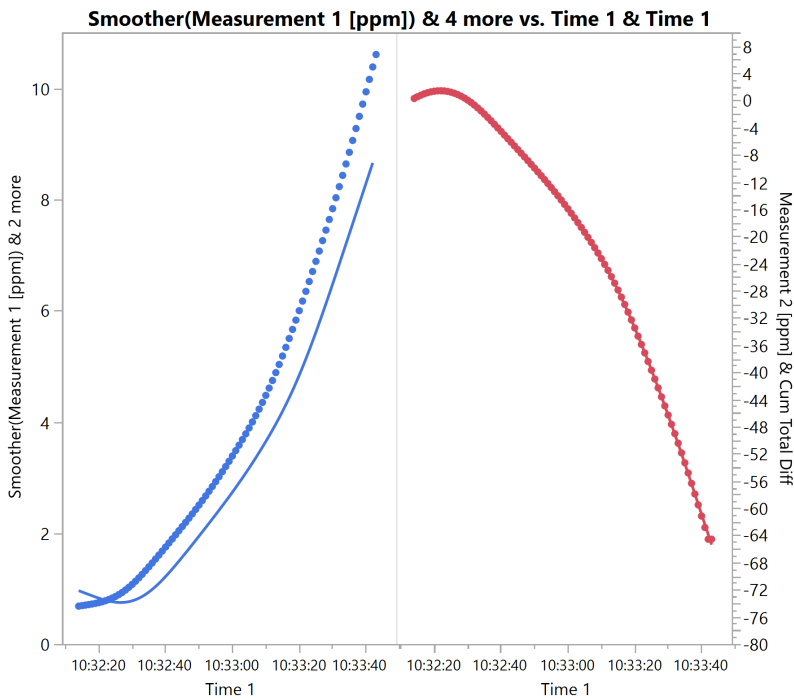
In production the whole process runs without visibility for the user. To demonstrate the steps and results every step updates some informative graphs.

Here is what shall be done to find the best shift:

Set reference curve and object curve

⬇

Calculate spline function for reference curve

⬇

Determine direction of shift

⬇

Move Y values of object curve one row

⬇

Calculate spline function for object curve

⬇

Calculate Column Sum of differences

⬇

Yes ← Col Sum decreased ?

⬇ No

End

The start situation of the process:

**Smoother(Measurement 1 [ppm]) & 4 more vs. Time 1 & Time 1**

The left graph shows the data points and the two spline functions. The dotted curve is the reference, the curve with the continuous line is the one that shall be shifted. Both curves reference the left Y-axis.
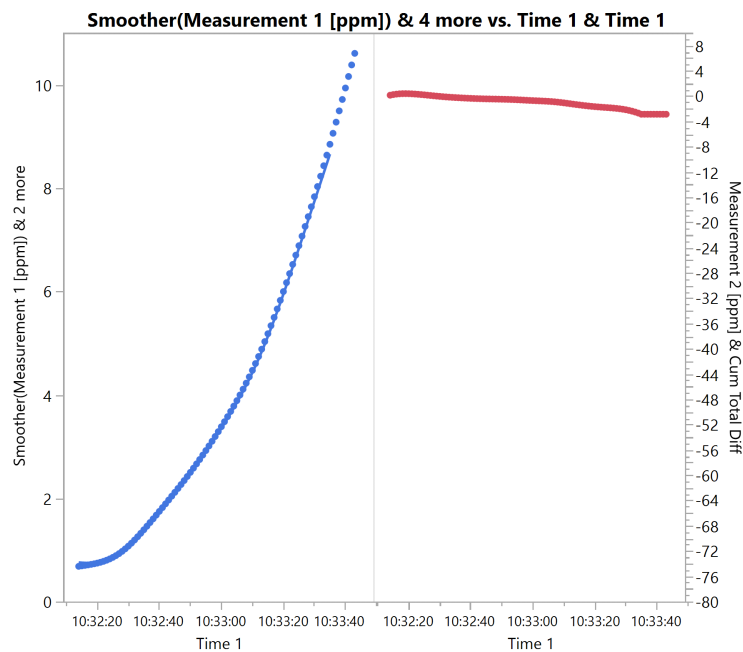
The red curve on the right shows the cumulated differences between the splines. Its scale is on the right Y-axis.

The graphs are updated with every step of the iterative process. One intermediate result looks like this:
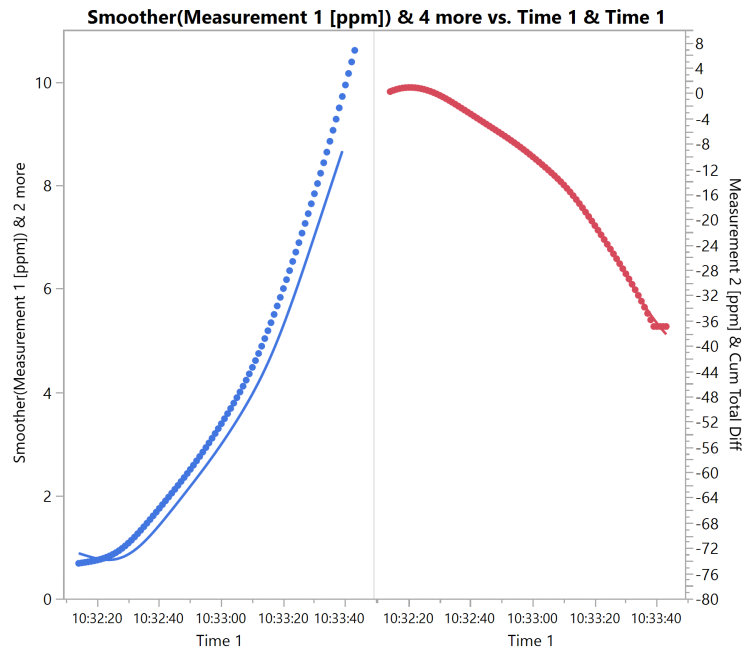
After the last iteration the curves look like this:

The difference of the absolute times of the two measurement devices was turned into a number of rows that one of the time series lagged behind the other one. After this lag was determined, the data of all sensors was shifted accordingly. The corrected data table is then the base for further analysis.