# Deconvolution Template Document

## Introduction

This document outlines all of the pertinent information about the “Excel Deconvolution Template” which may be used for data processing. The template aims to provide a user friendly method of “deconvoluting” TPD data. An excel spreadsheet was chosen as the method for this to allow for easy integration into the existing data processing ecosystem at NRCan. Additionally everyone is generally familiar with excel and has it preinstalled on their computer.

The first section of this document outlines the fundamentals of deconvolution.

The second section provides an indepth description of the workbood.

The third section works through an example of using this workbook on real TPD data.

## Gaussian Function

A gaussian function is a special function which has the form shown below. This function finds widespread application in fields ranging from statistics to image processing. As you can see the function is completely defined by three parameters and in the worksheet these are the parameters we will be playing with to fit to our data set.

One important property of a gaussian curve is the area under the curve. This can give valuable infrmaon about the data recoded. For example if you have oxygen desorption data as a rate of flow vs time then the integral (area under the curve) will be the total oxygen desorbed. Finding the area under part of a gaussian curve is actually very difficult and requires numerical approximation and a lot of work, lucky we only care about the total area under a gaussian curve. This greatly simplifies the calculation and allows us to use the expression below:

Deconvolution is the process of taking some signal and reducing it to its constituent parts. For this problem we are looking to take the data from our TPD and rewrite that data as the sum of a set of new functions which will give use new insight into the original data. There are many ways to look at this problem; one conceptual way is analogous to Fourier analysis. There is also a parallel with wavelet theory. But in the end we won’t be using any fancy math so if this section doesn’t make too much sense, that’s ok, you can still use the software.

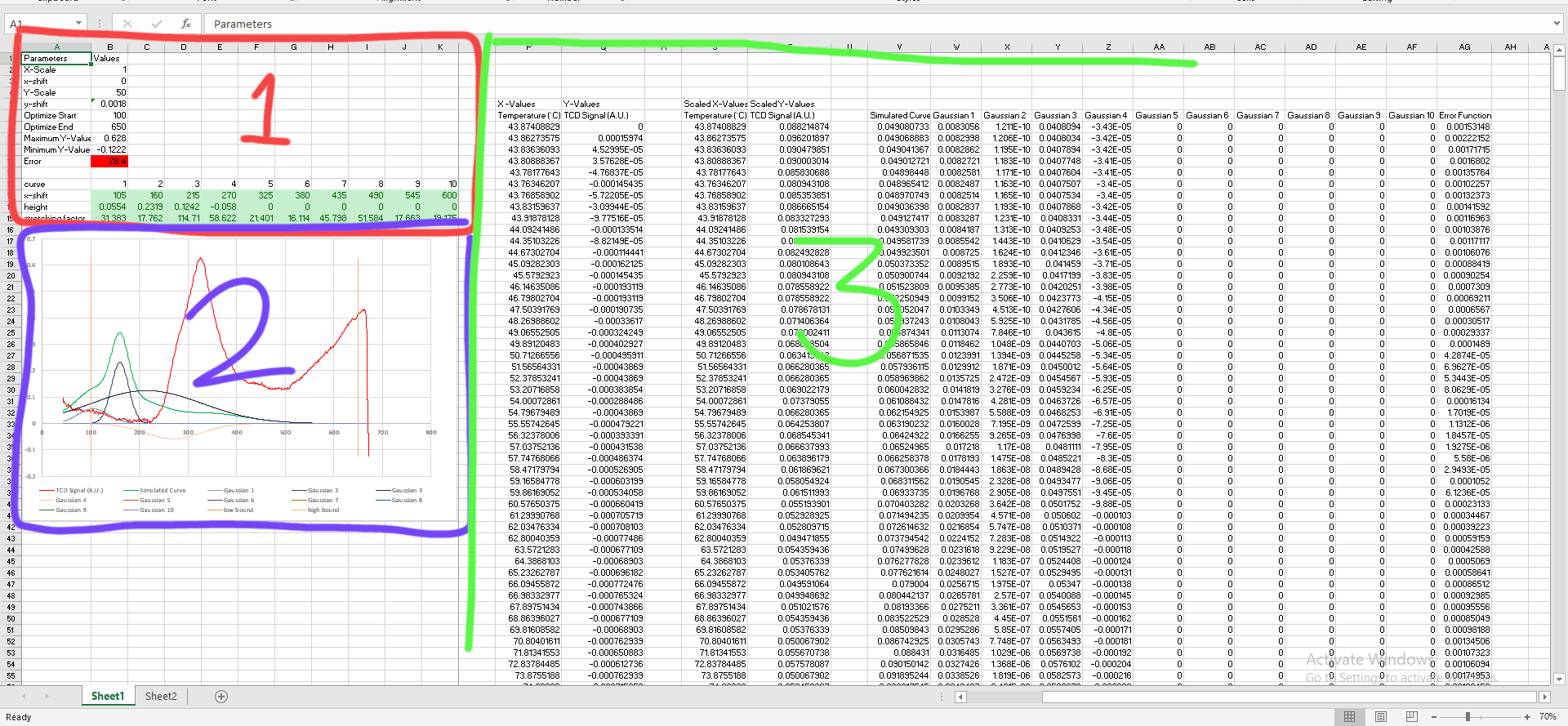
If our original data is f, then our aim is to rewrite the data as a linear combination of gaussian functions. An infinite set of gaussian functions produces a complete basis for any function. While this fact is true, it’s not super useful because we don’t have time to tune an infinite number of parameters. Instead the challenge becomes finding the set of parameters which contribute the most to the original function.

In Fourier analysis by taking the Fourier transform you also get the terms as a (infinite) series with the most significant term first and subsequent terms being less and less important. The challenge we face is that we don’t know which terms are the most important and which terms are the least important. That is what the workbook does, it takes a finite number of gaussians and optimizes their parameters maximizing the overlap between the sum of the basis gaussians and the original function.

Deconvolution as an optimization problem

# Worksheet Design

I will now describe the different sections of the worksheet and explain why there are there.



You can see that the workbook is broken into three sections;

1. This section houses all of the parameters which can be adjusted in order to change how the fit will be done.
2. This section has a graph which allows you to see the fit as you vary the parameters of interest.
3. This section has the actual values of the raw data, and of the fitted data.

The basic idea for this spreadsheet is that the data which has to be worked on is pasted in cell P6 as indicated by the “Paste Data” box and arrow. The x and y data is then linearly transformed into the next two columns. The linear transform allows for you to change the offset of the data, and scale the values as needed. The x-axis shift is unlikely to be used but is there to keep the spreadsheet organized and allow for that flexibility in the future.

The next block of columns houses simulated data. This is takes the parameters for each gaussian function and applies them at the scaled x values. The sum of all of these gaussians becomes the simulated curve. All of this data is plotted together on the graph.

By varying the parameters in section 1 the original data can be fit with the gaussian peaks. A reminder of what each parameter is for is included adjacent to the parameter.

|  |  |
| --- | --- |
| Parameter | Purpose |
| X-Scale | In combination with X-Shift provides full linear adjustment of the X-Axis data. The scale value is multiplied to all of the raw x-data after the shift is applied. If the x data is g(x) then the shifted data is given by:  An example of when this would be useful is if the collected data was in L and there was a constant offset of 0.10L. Then by setting the y-scale (a) to 1000mL/L and the y-shift (x0) to -0.10L the data will automatically be adjusted to the correct units and offset. |
| X-Shift | See X-Scale. |
| Y-Scale | See X-Scale. |
| Y-Shift | See X-Scale. |
| Optimize Start | The data collected from the TPD will usually have some strange behavior at the high temperature range which is not of interest, and usually explainable by instrumentation limits. Similarly at low temperatures there may be some behavior which is again not the focus of the study. In these situations you want to fit a certain number of gaussian functions between the high temperature and low temperature area of interest. To allow you to do this there are two vertical bars in the graph (Section 2). These bars correspond to “Optimize Start” and “Optimize End”. Between these two bars the error is calculated and minimized. Outside of these bars the error is ignored. |
| Optimize End | See Optimize Start. |
| Error | This value is the sum of the square of the difference of the simulated and scaled data. More concisely it can be expressed as:  Where yn is the original scaled y-axis data, and y’n is the simulated y-axis data. This equation is summed over n points where n ranges from optimize start to optimize end. |

# Data Fitting Example

In this section I’m going to run through an example of working with some data to illuminate some key steps not yet mentioned as well as providing a reference for how it’s done.

1. Open the TPD data and the “Gaussian Deconvolution Template”
2. Copy the X and Y data which you want to work with into the X-Values and Y-Values columns. Notice that the parameter cells are preset with some defaults which may or may not work for you. Also the scaled data has now appeared on your graph.
3. Begin adjusting the parameters for the initial conditions. The curve number let’s you know which curve is being addressed, the x-shift tells you the centre of the gaussian peak. The height is the value of the gaussian at the top of the peak. Lastly the stretching factor is how wide broad the gaussian peak is. Playing with these parameters you should be able to get a relatively good fit, or at least a good starting point.
4. Next we will optimize the fit using excels equation solver. This is a plugin you may need to install yourself before you can use the template, however it is easy to do and a standard package.
   1. Go to the data tab in the excel ribbon
   2. Click solver
   3. Click the set objective box and enter the Error parameter into the field. This is the cell we are minimizing, so make sure to check “min” and not “max”, we want to minimize the error, not maximize it!
   4. Under the “changing variables” field you must under the range that includes all of the gaussians you want to optimize. Excel will apply a clever algorithem to vary these parameters finding an optimal value for all of the variables.
   5. Click solve.
5. Veryify the results make sense. There is some conditional formatting to indicate cells which may be have odd values, but that doesn’t mean that they are wrong. More importantly just look at the graph, is the fit good? Are the peaks well distinguished and reasonable in position and size? Does the placement make sense. In this example it looks like the background data which is not interesting is forcing the fourth gaussian to become a basckground compensation.
6. Adjust the parameters/number of gaussians, fit bounds, etc. Then optimize again. I am not happy with this fit so I will move the optimization ends such that the error calculation ignores both the data at the beginning and at the end focusing on the peak of interest in the middle. Refitting with these new parameters then produces a much better result.

