**Deriving Zenith Equations**

We’ve looked at the Direct Sun equation, what its terms mean and where they come from, as well as looking at how to derive the correction factors to N, so that we can use the equation with an instrument that hasn’t recently been sent off for recalibration.

However, the majority of the observations across a season will not be Direct Suns, but rather zenith observations (sometimes referred to as Umkehr observations). Whilst for Direct Suns the light reaching the instrument has been largely unperturbed in its travel, it is a different story for zenith observations. These observations take scattered light from the open sky, each photon of which will have had a different path through the atmosphere as they bounce off various obscurations and molecules in the atmosphere. This variety of paths through the atmosphere means that there is not a “one size fits all” equation for deriving column ozone from zenith equations for all Dobsons, as there is for Direct Suns. Rather, a relationship has to be derived between column ozone (which we obtain by looking at Direct Suns that are close in time to our Zenith measurements) and the factors which appear in our Direct Sun equation, N and mu. Once this equation has been derived, we can then apply it to Zenith observations that do not have Direct Sun measurements nearby.

Now that we understand the general principle behind deriving these equations, let’s delve into the details of how exactly to do this. Firstly, by looking at the current method of deriving the equations, then by looking at possible future methods.

**Current Method: SigmaPlot Regression**

The method currently used to derive the equations is created by using software called SigmaPlot. It is available from BAS, but each user requires their own license, so to install it you must contact [servicedesk@bas.ac.uk](mailto:servicedesk@bas.ac.uk) to get a license key. If for whatever reason SigmaPlot is not available, similar regression analyses can be performed with Excel or other software [1].

First, you need the data in a format where the regression equation can be performed. This requires the N values and mu values for the AD\* zenith measurements, as well as the column ozone values calculated from the Direct Sun measurements that are nearest in time to the N and mu values. You should also add a column that is the time difference between the Zenith measurement and the nearest Direct Sun measurement in seconds. The addition of this column allows you to filter out measurements beyond a certain point. Why is this useful? Well let’s say there has been 2 weeks of continuous cloud where we were unable to take any Direct Suns: for a Zenith observation in the middle of this time period, the nearest DS measurement would be a week away and is unlikely to be of use, so we should exclude a subset of our Zenith measurements whilst deriving our equation.

Once we have all of this data in a file, we can then import it into SigmaPlot and open the Regression Wizard. In the settings for the Regression Wizard we can define our variables from the columns in the sheet and tell it the form of the equation that we wish to use. You can see from the image below, we normally use a polynomial in N(AD)/mu, but with other terms (such as N(A)) included also.

After the form of the equation has been set, we can then set the wizard going and SigmaPlot will come up with the best equation it can to model the relationship between the terms you give it and column ozone. Once it produces this, it will inform you of the significance of each of the terms and if you see a term (or terms) of little importance, you can cut this term in order to simplify the equation.

Graphical user interface, application

Description automatically generated

Table

Description automatically generated

Using the equation given for AD, you can then repeat the process for CD and for CD at high mu values (which behaves slightly differently than at low mu values due to the umkehr effect – more on this later).

\* AD measurements are generally considered to be the “Gold Standard” of ozone measurements, so we derive these first.

**Future Method: Machine Learning**

Charles Simpson from the AI Lab and I have been working on the problem of deriving the Zenith equations using machine learning. This may offer better results than from normal regression, partly because it can look at many different types of model and choose whichever seems to fit the data best. It may also prove to be a useful tool for spotting anomalies in the data, the exclusion of which will result in a better equation.

Unfortunately both Charles and myself are leaving BAS within a month of each other – however we have handed over the project to James Byrne, who is both familiar with the instrument and is a part of the AI Lab at BAS.

[1] <https://statisticsbyjim.com/regression/regression-analysis-excel/> gives a good summary of how to do a regression analysis with Excel.