PHY340 Data Analysis Feedback  
Group P02 doing Problem P2

# Data Analysis

The data analysis is not well explained, and suffers badly from a poor reading of the sources. In particular, the quantities from Carroll (1985) and from Liu and Jordan (1960) are treated as if they were the same thing, ***which they are not***: Car­roll defines *Kd* as *If*/(*Ir* + *If*), where *Ir* is the direct, and *If* the diffuse, horizontal irradiance *at ground level*, i.e. where *Z* is the solar zenith angle and *S'* is the solar constant, i.e. the extraterrestrial irradiance, whereas Liu and Jordan define (or *D*/*H*0 in their nota­tion). Clearly these differ by a factor of *Kt*, which is why one is much higher than the other: it makes no sense to treat them as if they were the same thing. The fact that the diffuse fraction you obtain from equation 5 is much higher than that obtained from equation 9 is not surprising: equation 5 gives you the fraction of sunlight *at ground level* which is diffuse rather than direct, whereas equation 9 gives you the fraction of sunlight *incident on the Earth’s atmosphere* which is subsequently received at ground level as diffuse light. Naturally you expect the latter to be lower. One should further note that equation 5 is Carroll’s expression appropriate to a cloudless sky: he quotes several different sets of coefficients (his table 4) for different levels of cloudiness. It is quite unlikely that the skies above Coventry were generally cloudless—Carroll’s measure­ments, in contrast, were carried out in Davis, California, where cloudlessness is probably the usual condition—and therefore equation 5 is probably inappropriate in any case.

It is completely unclear why you instead choose to do your calculations for “the day with the closest value to the monthly average” instead of using all days and averaging—you appear to be following Klein (and Liu and Jordan), but Klein was writing in 1976, before the widespread availability of computers: there is no reason why you could not do the appro9priate calculations for each day and take a genuine average. It is also not clear from your report why you record “the energy differences of every two consecutive values” (which is not what you mean: 1 and 48 are not consecutive; nor, for that matter, are they the values you should be subtracting—the first day’s collection would be cell 49 – cell 1, not cell 48 – cell 1—though fortunately this does not matter as there is no sunlight around midnight, so both values are always the same). Using a single day in­stead of the whole month is throwing away most of your data, and since you nowhere explain that the data files record *cumulative* energy readings, the reader has no idea why you are taking energy *differences*.

Your sunset hour angles are clearly wrong: at the latitude of Coventry (52.4°), the sunset hour angle varies from 124.2° at summer solstice (late June) to 55.8° at winter solstice (late Decem­ber). Your equation 3 is correct, but your numbers are complete garbage—they don’t even vary in the right direction (you appear to have concluded that the days are shorter in summer than in winter). Since the only code you have provided is the very trivial snippet used to read in the data and calculate daily irradiance, it is impossible to know what you did, but clearly you did not correctly code your equation 3. This may be why you obtain a negative efficiency for one of your measurements, which obvi­ous­ly cannot possibly be obtained by any valid calculation—the station does not record a nega­tive total energy for December, so it is simply not possible to obtain a negative efficiency using any sensible methodology. Indeed, the fact that many of your “direct beam” numbers are negative suggests that there is something very wrong with your calculations: clearly the direct compo­nent of the irradiance *can* be zero (on a completely cloudy day, which is not at all unlikely for winter in Coventry), but it *cannot* possibly be negative. Your values of *Kt* seem implausibly small—it is very unlikely that <10% of the incident sunlight makes it to ground level. You appear to have chosen your installations at random: in fact, the three you have chosen are nowhere near Coventry, so the horizontal irradiance data you were given are not relevant to them (only three of the installations appear in the PV data and in the list of installations around Coventry: 521, 1336 and 1572), so it is not clear that your efficiencies really mean anything anyway.

In fact, the values of *Kt* derived from the Coventry Met. Office data average, over the year, to 0.421±0.009, with little obvious seasonal variation though a good bit of variation from day to day, presumably due to weather conditions—see figure 1. This is reasonably similar to the over­all average of 0.463±0.011 from Klein’s Blue Hill data (his Table 4), although the latter have a clearer seasonal variation, probably because the climate of the eastern US is more extreme than that of the UK.

*Figure 1: Clearness index H/H0, calculated from Coventry weather station data.   
The red curve shows monthly averages.*

I am sure that your values of the diffuse fraction are wrong: it is not reasonable that the diffuse component always accounts for at least 80% of the total irradiance, as you obtained using your method 1, nor that it never accounts for more than about one-quarter, as you found using your method 2. The problem here is that your clearness indices are wrong; because they are all very small, you essentially always recover the constant in your chosen linear fit. I used Page’s ex­pres­sion from Klein (1976) (not Liu and Jordan’s, which Klein prefers, because the version given by Klein is not well-behaved at low *Kt*), and found values of the diffuse fraction averaging 0.524±0.010, which is con­sistent with Liu and Jordan’s figure 7 (and quite different from your values).

*Figure 2: the efficiency of Station 521.  
The red curve shows monthly averages.*

My overall efficiency plot for station 521 (which is actually in the vicinity of the Coventry weather station, unlike the ones you chose) is shown in figure 2: the average over the year is (11.80±0.12)%, but this is pulled down by poor weather from October to December: averaging from February to September gives about 12.5%.

You seem to have wound up, overall, with plausible final efficiencies, but I think this is a com­plete accident. Every calculation I have checked is wrong. The average below is a combination of two rather different marks: Alastair gave you the benefit of the doubt because your answers seemed to be of the right order of magnitude: he says “the group must have programmed the calculation correctly for “good” method since the results are within the expected range.” This turns out not to be the case: as noted above, the fact that the results are sensible appears to be purely fortuitous.

Last but not least, there is no serious attempt here to produce an estimated error. The “oh, it’s far too complicated” excuse is unacceptable: you do not *have* individual uncertainties, so the prob­lem of propagating them through a complex calculation simply does not arise. What you *do* have, but did not use, is *statistics*: each of your monthly averages has an easily calculable stan­dard deviation, leading to an easily calculable error on the mean. I did the whole analysis in Ex­cel, because I don’t have Python on my travelling laptop: I don’t recommend this (because some of the formulae are very long), but it wasn’t *difficult* (it took less than a day).

I grant you that Klein’s equation formatting, in particular, is *awful*, and it is quite possible that in some cases you may have misinterpreted it—it is not impossible that in some cases *I* have mis­interpreted it, although I did cross-check with Liu and Jordan where both sources gave the equation in question—but that’s not why your sunset hour angles are wrong, since your equa­tion 3 is correct.

Average mark for this section: 30.5/50

# Data Presentation

The presentation of the data is poor. The analysis is not well explained: since the angle between the Sun and the normal to the plane of the installation varies dramatically over the course of each day, it is not at all clear how you define your angle *θ* in equation 6, and you do not even attempt to define your angle *β* in equation 7. Your data are presented entirely in the form of long tables of numbers (quoted, in most cases, to a wholly unreasonable number of significant figures: given that your calculated efficiencies for station 368, for example, vary from 1.9% to 33.7% (neither of which is at all sensible, see figure 2), you surely should have recognised that quoting the overall average to 6 significant figures was not reasonable (and, even if you failed to calculate any other errors, you could have provided a standard error for this average). You offer no comparison with literature values, and therefore your claim that your second model “pro­duces more appropriate values” is unsupported. Klein (1976) and Liu and Jordan (1960) both provide copious plots and tables of data, but you do not use these to check your intermediate results: if you had, perhaps you would have noticed that your clearness indices were completely unrealistic and your diffuse fractions correspondingly implausible.

Most of all, there is no attempt to provide any sort of graphical presentation, although your table of data could easily be used for that purpose. This has probably served you well, as a matter of fact—Alastair would have noticed that your numbers were wrong if he had seen a plot of effi­ciency against date, since it should not change by a factor of 20—but normally it would cost you many marks. You have been told many times that you should provide graphical representations of your results whenever this is possible, and it was clearly possible here.

Average mark for this section: 10/30

# Style

You seem to have some idea of what a report should look like: your section headings are broadly ap­propriate, although generally there should ***not*** be a section headed “Error analysis” (unless there are *systematic* errors which need to be explained in detail, error handling either does not need to be discussed at all—for example, standard error propagation from raw data to physics results can simply be assumed—or needs to be discussed alongside the results). However, your abstract is dreadful (no explanation of what you mean by “first model” and “second model”), your introduction is worse (no explanation of what your data are, no discussion of where your stations are compared to Coventry (you might then have notices that station 382 is 101 km away), a rather self-pitying and frankly childish complaint about “long formulas” (that’s what computers are *for*: if you did need to propagate any errors—which you don’t, see above—it would be completely straightforward to work through the upper and lower bounds for each va­riable (it’s just 2*n* more columns on your spreadsheet, where *n* is the number of variables with errors that need propagating) and then add the results in quadrature), and no discussion of the reasons for your choice of methodology), and your method section is inadequate, as noted in the previous section. There are also numerous errors of English: in particular, it is *not* correct to have a capital letter after an equation, if the previous sentence is still continuing (“The values are calculated from *y* = *mx* + *c*, where *m* is the fitted gradient and *c* the intercept”—it makes no difference whether the equation is inline or centred on its own line). I know that Microsoft Word will capitalise the offending word automatically, unless you’ve had the sense to turn off that part of its autocorrect facility, but that just means that you have to take a few seconds to change it back. Some words are just wrong: in the Introduction I presume that you mean that the irradiance data are assumed to be accurate, since “reasoned” does not make sense (but such an assumption is unwarranted unless you take the trouble to select installations *that are actu­ally close to Coventry*: you cannot reasonably assume that the weather over the whole of Eng­land is well represented by Coventry).

Average mark for this section: 11.5/20

Overall average mark: 52%.