Atmos. Chem. Phys. Discuss., 12, C7667–C7674, 2012
www.atmos-chem-phys-discuss.net/12/C7667/2012/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Effects of cosmic ray decreases on cloud microphysics" *by* J. Svensmark et al.

J. Svensmark et al.

enghoff@space.dtu.dk

Received and published: 1 October 2012

Dear Eimear Dunne,

Thank you for your comments, which appear in italics below along with our answers.

> *When testing for significance, a confidence level should be chosen, and a confidence interval defined. Svensmark et al. (2012a) have used $[X_{[-15,-5]} - 2\sigma, X_{[-15,-5]} + 2\sigma]$ as their 95% confidence interval, where σ is the standard deviation formulated by Svensmark et al. (2012a) by averaging the standard deviations of '100 realizations of the mean of 5 randomly placed 36-day intervals, chosen from the 2000-2006 MODIS data, excluding the FD event intervals'. However, when working with a sample rather*

> *than a known population, they should instead estimate a confidence interval $[X - t_{\alpha/2,N}S, X + t_{\alpha/2,N}S]$ based on the sample, where $t_{\alpha/2,N}$ is Student's one-tailed t-statistic for confidence level $\alpha/2$ and sample size N, X is the sample mean and S is the sample standard deviation.*

Given the number of observations per studied epoch the difference between using $t_{\alpha/2,N}$ and σ is minimal, especially considering that we do have a very large sample number for the standard deviation since we sample the entire data set 100 times. Nevertheless we will, for the sake of discussion, use Student's t-test below. We will however keep the base at $X_{[-15,-5]}$ (see below).

> *Figure 1(b) shows the same data, but using the mean from the whole 36-day period instead of the 15 days prior to the Forbush decrease. The apparent increase in the confidence interval is due to a change in the scale of the y-axis.*

Using the entire 36-day period will contaminate the base level because of the potential change in the mean that a signal represents. We remain confident that a good measure of a base level is found *before* an actual Forbush decrease eg. at days [-15,-5]. We do however also acknowledge that the variance consequently changes in time, as shown in the comment by Laken et. al. (SC C962). This should (and will below) be considered when testing for significance. Also please refer to our final response where we apply another method of analysis which circumvents this issue.

> *Figure 1(c) uses the confidence interval $[X - t_{\alpha/2,N}S, X + t_{\alpha/2,NS}]$, instead of $[X - 2\sigma, X + 2\sigma]$.*

It is unclear which N (N, Neff or infinity?) and which α has been used in this and other plots, but we take it to be $N = 36$ and $\alpha = 0.95$ although it produces a slightly different confidence interval compared to your Fig. 1C. While the modification to σ is minimal with $N = 36$, the S of your Figure 1C increases quite a bit, which might well be attributed to the signal in the data, and as such this approach suffers from the same contamination problem as the setting of a base level, a problem which we feel is important to eliminate from the analysis.

In the following, we shall compute the minima in terms of the confidence interval adjusted by Student's t-distribution, while taking into account the temporal development of the standard deviation $\sigma(t)$. In order to do so, we represent the confidence interval as $\left[X_{[-15,-5]} - t_{\alpha/2,N}\sigma(t), X_{[-15,-5]} + t_{\alpha/2,N}\sigma(t)\right]$ such that the standard deviation is calculated for each day in the time series (using Monte Carlo) and the extrema which is tested is compared to the usual variance at the day that it occurs. This temporally varying standard deviation is fully representative of the situation plotted, and it does not allow itself to be contaminated by a potential signal in either base level or variance. As such it serves as a good interval for the significance test.

> *In Figure 1(d), superposed epoch analysis is used to adjust the data from each Forbush decrease, removing any linear trends and making the data comparable. The red line in Figure 1(d) shows the three-day running mean, as in Figure 1 of Svensmark et al. (2012a). The adjusted data points are also shown.*

The difference between a "Superposed Epoch Analysis" and the method of our paper seems to be the removal of a linear trend (and the variance and mean details discussed above). We agree that removing a linear trend from the events provides a better basis for comparison, even though the effect is minimal in this case on the shape of the graphs. It does however have an important effect on the temporal development of the variance, which is apparent when comparing Figure 1 of comment SC C962 (Laken et. al) to Fig. 1 of this comment, where the six parameters are shown along with 1 and

2 standard deviations found using the above method and with linear trends removed. Note that this narrows the confidence interval.

> *Figure 1(e) is the same as Figure 1(d), except that FD 2 from Table 1 in Svensmark et al. (2012a) has been omitted.*

Please refer to our reply to comment SC C962 (Laken et al) where we discuss the significance of adding and removing this single event.

> *Figure 1(f) accounts for autocorrelations within the data when calculating the confidence interval. The effective sample size for LCF is Neff = 7.61 when calculated according to the method specified in Svensmark et al. (2012b). Therefore, the size of the confidence interval increases by 36/7.61 = 2.76 times.*

We are not sure what is shown in your Figure 1f, since the curve doesn not appear to be CF with or without a removal of linear trend, and consequently it is hard to discuss. However we note that the confidence interval found by applying your correction for autocorrelation seems to be very large. Only 2 points (barely) exceed you dark grey area (corresponding to about 1 std. dev.). Another, less drastic, approach would be to simply use the reduced sample size in the t-test.

> *A more appropriate statistical analysis shows that the observed response in data is not statistically significant. Figure 2 here shows the equivalent of Figure 1 in Svensmark et al. (2012a), but using superposed epoch analysis to calculate the confidence interval (equivalent to Figure 1(d)) instead of Svensmark et al. (2012a)s approach. The effective sample size has not been accounted for in Figure 2 as it was in Figure 1(f), meaning that the*

true confidence interval is likely to be larger due to the presence of autocorrelations.

Figure 1 of this reply takes the same issues into account as your Fig. 2. In both representations the signal remains significant in several parameters above the 95% limit. This does not take into account the constraints of localization and sign which are explained in our comment AC C717 - these factors strengthen the confidence in the observed signal. Another way to take the constraint of sign (based on the physics we only expect the signal in one direction) into account is to make the t-test 1-sided instead of 2-sided. This is done in Fig. 2 of this comment.

Forbush et al. (1983) provide a thorough and detailed outline of the appropriate use of superposed epoch analysis. Using Forbush et al. (1983)s method of calculating the F-statistic associated with a given data set, it should be possible to reduce the effect of noise or of longterm changes in the data. Svensmark et al. (2012a) would therefore not be limited to studying only the first five FD events from Table 1 of Svensmark et al. (2012a). Forbush et al. (1983) warn quite strongly against accepting an apparent signal without testing for quasi-persistency within the data. However, since we are dealing with five non-sequential epochs, the tests for quasi-persistency described in Forbush et al. (1983) cannot be carried out. If the calculated F-statistic is larger than F_α for the appropriate confidence level, the response is found to be significant. For I epochs of J days, there are (J - 1) and (I - 1)(J - 1) degrees of freedom in the system; so for a confidence level α = 0.05, we have 35 and 140 degrees of freedom, giving F_α = 1.5073. Autocorrelations were not accounted for within the data. If the degrees of freedom of the system were reduced based on the number of independent data points, the value of F_α would increase, making the rejection of the null hypothesis less likely. Table 1 gives the F-statistic for each data set, together with the

W statistic from the BrownForsythe test for homogeneity of variances and the probability P that this F-statistic occurred by chance. If $W > F_\alpha$, the variances of the different epochs are inhomogeneous; however, Forbush et al. (1982) point out that moderate departures from normally distributed data sets with homogeneous variances have a negligible effect on the results of the tests. If $P < 0.05$, the variations are found to be statistically significant. Three of the data sets are found to be significant from this analysis; effective emissivity, LCF and CCN. When FD #2 is excluded, LCF is no longer found to be significant. It is notable that ϵ and LCF have by far the most inhomogeneous variances between epochs. The time series for and CCN are shown in Figures 3(a) and 3(b). The time series of CCN shows both a very low and very high peak in quick succession.

This test you propose is to investigate the equality of variances in the investigated epochs. It is one of many tests which all come with their own advantages and disadvantages. We note that the methods put forward in Forbush et al (1983) does not appear to have become standard in epoch analysis - if a well documented standard existed it would indeed make things easier. The concern here is akin to that put forward in comment RC C247 (Kristoffer Rypdal) about increasing fluctuations. We reply to this in AC C591, by looking at extrema going in the opposite direction of that expected by the physics and find it not to be a problem. That your analysis results in a significant result for CCN should maybe be a cause for concern, since the CCN product is by far the noisiest.

When examining a longer time period, missing data meant that two FD events had to be omitted from the time series running over [-40, 40]. Figures 3(c) and 3(d) use FD # 2, 3, 5 and 6. We can see from Figure 3(d) that the CCN time series experiences a great deal of variance and is likely to be a fat-tailed distribution. This is supported by the fact that Svensmark et al.

(2012a)s σ value for CCN was sufficiently large that they did not observe significance in the response. If the distributions from which CCN are drawn are not normally distributed, but are instead highly fat-tailed, this test may incorrectly reject the null hypothesis.

Using the Principal Component Analysis (PCA), which combines all parameters, we have examined even longer time periods than [-40, 40] and still found the signal to be significant. The PCA takes advantage of all parameters at the same time and thus provides a strong piece of evidence for the correlation between GCR and cloud parameters.

The use of the superposed epoch method has improved upon the methods of Svensmark et al. (2012a), and found no evidence for any statistically significant response in optical thickness, liquid water path, or effective radius. The response in liquid water path is dominated by FD# 2, and is no longer significant after its removal. Although there appears to be a statistically significant response in CCN concentrations and effective emissivity, this may be due to the inhomogeneity of variances between epochs. It would be advisable to account for autocorrelations within the data and compare a larger selection of Forbush decreases over a longer time series before accepting the significance of these signals as true. The analysis carried out by Svensmark et al. (2012a) is flawed for several reasons outlined in this and other comments on the manuscript, but especially due to the bias introduced to the system by neglecting to use the average over the whole sample. Therefore, in my opinion the manuscript does not have sufficient merit to proceed to ACP.

In this and previous repsonses we consider the comments put forward by all contributers to the discussion. We acknowledge some points, such as the time-dependent

standard deviation, and refute others. In addition we apply another method of analysis where the concerns expressed by the commenters are adressed - please see our final response. Taking the relevant factors into account we still find, with some modification, our original point - that there appears to be a significant response in cloud parameters to coronal mass ejections - to be valid.