Manuscript prepared for Atmos. Chem. Phys. Discuss.
with version 2.2 of the LaTeX class copernicus_discussions.cls.
Date: 27 March 2012

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

Eimear M. Dunne

University of Leeds

Correspondence to: Eimear Dunne
(E.Dunne@leeds.ac.uk)

Svensmark et al. (2012a) examine the apparent response of six MODIS data sets to short-term decreases in atmospheric ionisation known as Forbush decreases (FD). They concluded that they had observed statistically significant responses to FDs at the $2-3\sigma$ level in liquid cloud fraction (LCF), effective emissivity ϵ, aerosol optical thickness τ, and liquid water path (LWP), and at the $1-2\sigma$ level in the derived values of effective radius (R_{eff}) and column density of cloud condensation nuclei (CCN). Their online supplement, available at http://www.atmos-chem-phys-discuss.net/12/3595/2012/acpd-12-3595-2012-supplement.zip, allowed a further analysis of the data using the superposed epoch method. The fluctuations which they observed were not found to be statistically significant.

Statistics cannot prove any hypothesis; it can only provide a probability that a given hypothesis is or is not correct. If Svensmark et al. (2012a) want to support the theory that Forbush decreases have affected the MODIS data under examination, they should first provide a null hypothesis which can then be tested, and possibly rejected. In this case, the null hypothesis is that the observed fluctuations in the data are part of the natural variability expected from the underlying behaviour of the system.

By using the mean from days (-15,0) or (-15,-5) of the Forbush decrease instead of the mean from the whole period, Svensmark et al. (2012a) have introduced a bias to their analysis. The motivation for their choice of this mean was the assumption that there was a response to the Forbush decrease within the data, and the mean from the full period had effectively been contaminated by this response. Simply by making this assumption, Svensmark et al. (2012a) have rendered their whole analysis questionable.

When testing for significance, a confidence level should be chosen, and a confidence interval defined. Svensmark et al. (2012a) have used $[\bar{X}_{[-15,-5]} - 2\sigma, \bar{X}_{[-15,-5]} + 2\sigma]$ 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 $[\bar{X} - t_{\alpha/2,N}S, \bar{X} + t_{\alpha/2,N}S]$ 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, $\bar{X}$ is the sample mean and S is the sample standard deviation.

As has correctly been pointed out by Rypdal (2012), a certain number of data points are expected to fall outside of the confidence interval. In Svensmark et al. (2012b), it was stated that strong autocorrelations within the data would affect the proportion of data points which would be expected to fall outside of the confidence interval. However, the autocorrelations would also increase the magnitude of the confidence interval. When data are strongly autocorrelated, their values have less spread than would be expected from random fluctuations about a mean, leading to an underestimation of the sample variance. Because there are only N_{eff} independent data points in the sample, the sample standard deviation should be $S = \sum_i (X - Xi)^2 / N_{eff}$, which will be larger than the sample standard deviation calculated from N (Wilks, 1997). This will increase the confidence interval and make the rejection of the null hypothesis less likely.

Superposed epoch analysis can be used to detect signals in data where the underlying variability of the system can cause problems (Chree, 1912). This is done by defining a new set of data $X_{ij} = Y_{ij} - m_i j - c_i + \bar{Y}$ where $m_i j$ is the value of the linear fit through the ith epoch's data on the jth day of the epoch, c_i is the ith epoch's intercept, Y_{ij} is the data from the jth day of the ith epoch, and $\bar{Y}$ is the ensemble mean over all epochs. This new definition removes any linear trend from the data and ensures that all epochs have the same mean $\bar{Y}$, thus providing a more comparable picture of the underlying distribution while retaining all variation within each epoch.

Svensmark et al. (2012a) have uploaded the IDL code which they have used in their statistical analysis, and it has therefore been possible to engage in an in-depth statistical assessment of the data. Figure 1 shows the change in the apparent response in LCF as the calculation of the 95% confidence interval (light grey) is improved using the method outlined above. LCF was chosen as it had the strongest response of the six variables when analysed using the superposed epoch method (cf. 2). To accurately determine whether any response is statistically significant, it would be necessary to test the data for normality using a Shapiro-Wilk or Kolmogorov-Smirnov test. If the data are from a fat-tailed rather than a normal distribution, the confidence interval calculated from the sample will be underestimated.

Figure 1(a) shows the response of the LCF using Svensmark et al. (2012a)'s analysis method. 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. Figure 1(c) uses the confidence interval $[\bar{X} - t_{\alpha/2,N}S, \bar{X} + t_{\alpha/2,N}S]$, instead of $[\bar{X} - 2\sigma, \bar{X} + 2\sigma]$. 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. 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. Figure 1(f) accounts for autocorrelations within the data when calculating the confidence interval. The effective sample size for LCF is $N_{eff} = 7.61$ 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.

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.

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 long-term 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 Brown-Forsythe 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 ϵ, 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 ϵ 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.

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 and ϵ are drawn are not normally distributed, but are instead highly fat-tailed, this test may incorrectly reject the null hypothesis.

My forthcoming paper (Dunne et al., 2012, in preparation) examines the response of atmospheric aerosol concentrations and optical properties to a short-term decrease in the nucleation rate using the global aerosol microphysics model GLOMAP. If a short-term change in the ion-induced nucleation rate generates a response in cloud and aerosol properties, a global aerosol microphysics model is the ideal tool to quantify that response. We did not find any statistically significant response to a short-term change in the nucleation rate.

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.

Acknowledgments

I thank Ben Laken, Jasa Čalogović, Lindsay Lee, Anja Schmidt and Matt Woodhouse for helpful discussion. All data analysed in this comment came from the supplement to Svensmark et al.

(2012a) found at http://www.atmos-chem-phys-discuss.net/12/3595/2012/acpd-12-3595-2012-supplement.zip. The author acknowledges support from EC's Seventh Framework Programme under grant agreement no. 215072 (Marie Curie Initial Training Network "CLOUD-ITN").

References

Brown, M. B. and Forsythe, A. B. (1974), "Robust tests for the equality of variances", *Journal of the American Statistical Association* , Vol. 69, pp. 364–367.

Chree, C. (1912), "Some phenomena of sunspots and of terrestrial magnetism at Kew Observatory.", *Philosophical Transactions of the Royal Society of London Series A* , Vol. 212, pp. 75–116.

Dunne, E. M., Carslaw, K. S., Lee, L. A. and Reddington, C. L. (2012), "No effect of a short-term decrease in the nucleation rate on atmospheric aerosols", *Atmospheric Chemistry and Physics Discussions, in preparation*, Vol. 12

Forbush, S., Duggal, S., Pomerantz, M. and Tsao, C. (1982), "Random Fluctuations, Persistence, and Quasi-Persistence in Geophysical and Cosmical Periodicities: A Sequel", *Reviews of Geophysics and Space Physics* , Vol. 20, No. 4, pp. 971–976.

Forbush, S., Pomerantz, M., Duggal, S. and Tsao, C. (1983), "Statistical considerations in the analysis of solar oscillation data by the superposed epoch method", *Solar Physics* , Vol. 82, pp. 113–122.

Rypdal, K. (2012), "Interactive comment on *Effects of cosmic ray decreases on cloud microphysics* by J. Svensmark et al." , *Atmospheric Chemistry and Physics Discussions*, Vol. 12 , pp. C247–C255.

Svensmark, J., Enghoff, M. B. and Svensmark, H. (2012 a), "Effects of cosmic ray decreases on cloud microphysics", *Atmospheric Chemistry and Physics Discussions* , Vol. 12, pp. 3595–3617.

Svensmark, J., Enghoff, M. B. and Svensmark, H. (2012 b), "Interactive comment on *Effects of cosmic ray decreases on cloud microphysics* by J. Svensmark et al." , *Atmospheric Chemistry and Physics Discussions* , Vol. 12 , pp. C585–C590.

Wilks, D. S. (1997), "Resampling hypothesis tests for autocorrelated fields", *Journal of Climate* , Vol. 10, pp. 65–82.

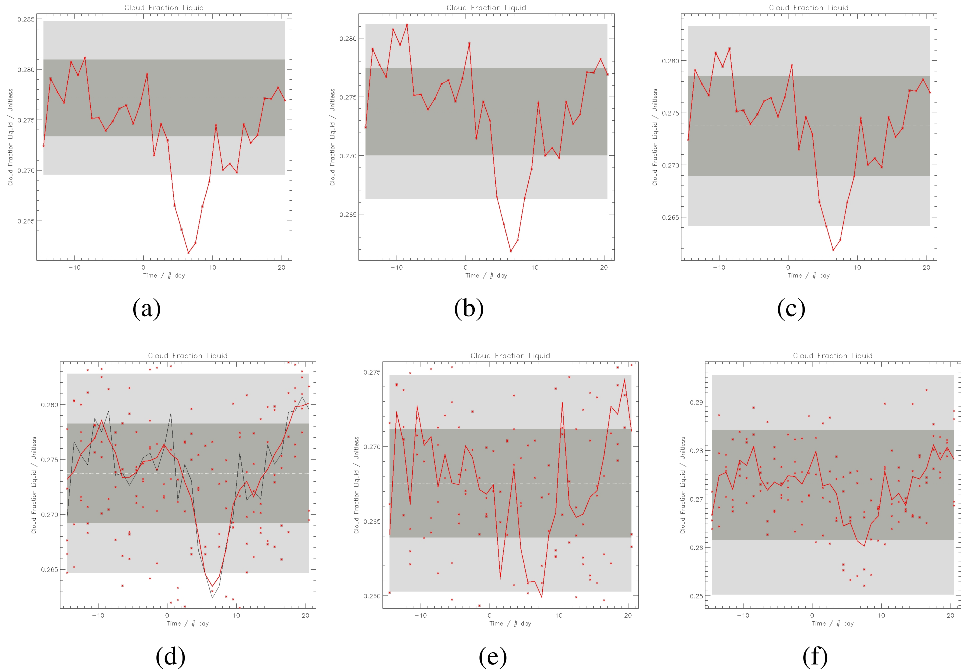
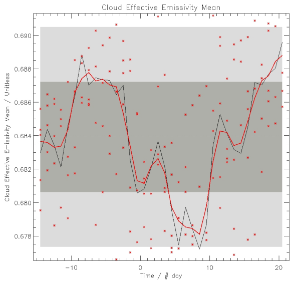
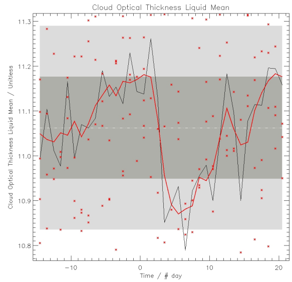


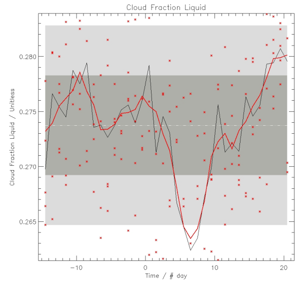
Fig. 1. Change in the apparent response of LCF with improved statistical analysis.



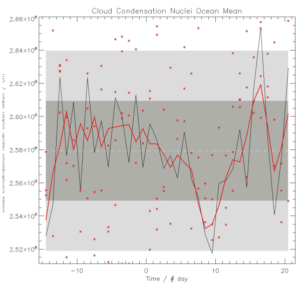
(a) ϵ



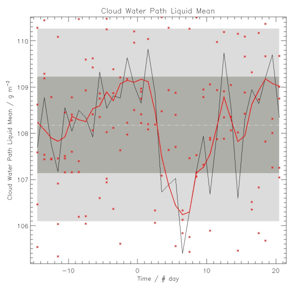
(b) τ



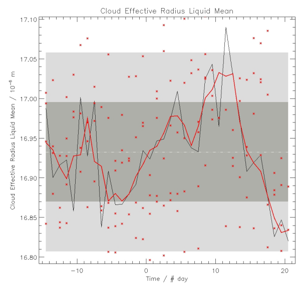
(c) LCF



(d) CCN

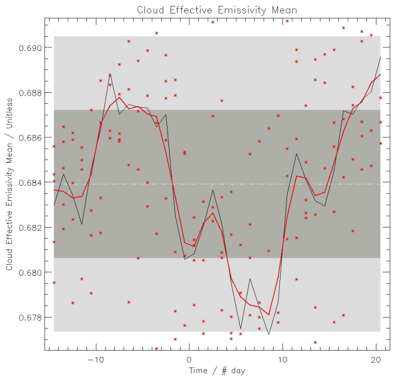


(e) LWP

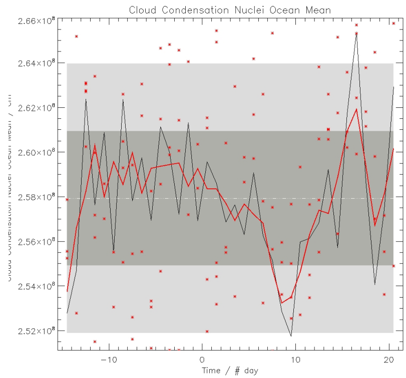


(f) R_{eff}

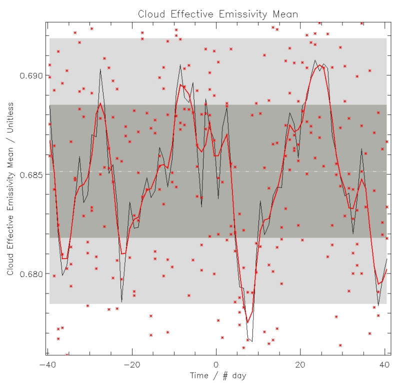
Fig. 2. Response in the MODIS data sets using superposed epoch analysis (equivalent of Figure 1 from Svensmark et al. (2012a).)



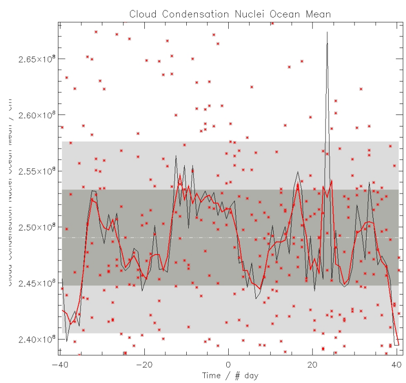
(a) ϵ on days (-15,20)



(b) CCN on days (-15,20)



(c) ϵ on days (-40,40)



(d) CCN on days (-40,40)

Fig. 3. Superposed epoch time series for effective emissivity ϵ and CN.

Table 1. The data sets found to be significant also had less homogeneous variances between epochs (larger W statistic). Based on this, it would be precipitate to accept this apparent signal without at least including additional Forbush decreases in the analysis and accounting for autocorrelations within the data.
* The residual variance is larger than the signal variance, and so instead 1/F was tested against F_{α}(140,35) = 1.6138, as was done by Forbush et al. (1983).
** LCF with FD #2 excluded was tested against $F_{\alpha}(35,105) = 1.5352$.

Variable	S_c^2	S_R^2	F	P	W
Effective Emmisivity	5.64×10^{-5}	2.22×10^{-5}	2.54	<0.0001	5.98
Optical Thickness*	0.067	0.069	1.02	0.49	2.28
Liquid Cloud Fraction	1.07×10^{-4}	6.17×10^{-5}	1.74	0.01	5.36
Liquid Cloud Fraction**	5.7×10^{-5}	4.7×10^{-5}	1.16	0.28	7.36
Column integrated CCN*	4.73×10^{13}	8.29×10^{13}	1.75	0.0277	3.12
Liquid Water Path	5.66	5.49	1.03	0.44	2.13
Effective Radius	0.021	0.018	1.14	0.29	3.80