

The Mansurov effect: Statistical significance and the role of autocorrelation

Jone Edvartsen^{1,*}, Ville Maliniemi¹, Hilde Nesse Tyssøy¹, Timo Asikainen², and Spencer Hatch¹

¹ Birkeland Center for Space Science, Department of Physics and Technology, University of Bergen, 5007 Bergen, Norway

² Space Physics and Astronomy Research Unit, University of Oulu, 90570 Oulu, Finland

Received 22 November 2021 / Accepted 17 March 2022

Abstract—The Mansurov effect is related to the interplanetary magnetic field (IMF) and its ability to modulate the global electric circuit, which is further hypothesized to impact the polar troposphere through cloud generation processes. We investigate the connection between IMF B_y -component and polar surface pressure by using daily ERA5 reanalysis for geopotential height since 1980. Previous studies produce a 27-day cyclic response during solar cycle 23 which appears to be significant according to conventional statistical tests. However, we show here that when statistical tests appropriate for strongly autocorrelated variables are applied, there is a fairly high probability of obtaining the cyclic response and associated correlation merely by chance. Our results also show that data from three other solar cycles produce similar cyclic responses as during solar cycle 23, but with seemingly random offset in respect to the timing of the signal. By generating random normally distributed noise with different levels of temporal autocorrelation and using the real IMF B_y -time series as forcing, we show that the methods applied to support the Mansurov hypothesis up to now are highly susceptible to random chance as cyclic patterns always arise as artifacts of the methods. The potential non-stationary behavior of the Mansurov effect makes it difficult to achieve solid statistical significance on decadal time scales. We suggest more research on, e.g., seasonal dependence of the Mansurov effect to understand better potential IMF effects in the atmosphere.

Keywords: solar-climate link / significance testing / Monte-Carlo / false-detection-rate / periodic forcing

1 Introduction

First proposed in 1974, the Mansurov effect is based on the correlation between daily polar surface pressure and the B_y -component of the interplanetary magnetic field (IMF). A significant correlation has been shown in multiple studies (Mansurov et al., 1974; Burns et al., 2008; Lam et al., 2013, 2014). Evidence of significant ionospheric perturbations related to the same change in B_y also exists (Tinsley, 2000, 2008; Frank-Kamenetsky et al., 2001; Kabin et al., 2003; Pettigrew et al., 2010; Lam et al., 2013). A physical mechanism involving the Global Electric Circuit (GEC) modulating cloud generation processes has been suggested to link IMF B_y to the polar surface pressure (Lam & Tinsley, 2016). Studies have also focused on the internally generated vertical current density (J_z). The internally driven changes in J_z have been linked to changes in the polar pressure (Tinsley, 2008; Lam & Tinsley, 2016; Zhou et al., 2018), indicating that the IMF B_y , which also induces changes in J_z , could play an important role.

For the Mansurov effect, the theory predicts a positive and negative relation between the IMF B_y -component and the polar surface pressure/geopotential height in the southern and northern hemispheres, respectively (Burns et al., 2008). The impact on the microphysics of clouds is predicted to begin in less than a day. As this effect is small, it is expected to take days for the accumulative effect to change cloud radiative forcing, leading to pressure changes related to the Mansurov effect (Frederick et al., 2019; Tinsley et al., 2020). The effect has been found to be first detectable in the lower troposphere (Lam et al., 2014). Mansurov et al. (1974) found correlations between IMF B_y and surface pressure in the time period around 1956–1964 (approximately solar cycle 19). Three individual periods (1964–1974, 1995–2005, and 2006–2015) have been found to show the associated pressure anomalies in both hemispheres (Mansurov et al., 1974; Page, 1989; Zhou et al., 2018). However, the statistical significance is only calculated through t -test or as one standard deviation of the mean. Most other publications on the effect focus on the period of solar cycle 23 (Burns et al., 2008; Lam et al., 2013, 2014, 2018; Zhou et al., 2018). This time interval produces statistical significance in both hemispheres when assessed by

*Corresponding author: jone.edvartsen@uib.no

the t -test. Burns et al. (2008) (hereafter B2008) thoroughly investigate the 1995–2005 period.

The IMF B_y has a 27-day periodicity associated with the solar rotation period (e.g., Gonzalez & Gonzalez, 1987). B2008 found a 27-day periodic pressure response in both hemispheres when regressing polar pressure to the IMF B_y for the period 1995–2005. This periodic response was attributed as evidence for a physical link between the IMF B_y and the polar pressure. In the southern hemisphere (SH), statistical significance calculated through the t -test showed this periodic response to be significant for the given period, while no significance was found for the northern hemisphere (NH). However, it was noted that while statistical significance was not achieved in the NH, the appearance of a 27-day periodic pressure response serves as evidence of the Mansurov effect. Tinsley et al. (2020) found a 27-day periodic response when correlating the IMF B_y to optical thickness of the overhead stratus-type clouds, which was put forward as evidence of the pathway of the Mansurov effect. In addition, Lam et al. (2018) correlated the IMF B_y with atmospheric temperature for 1999–2002. The significance is calculated without taking into account the temporal autocorrelation but nonetheless shows a significant temperature perturbation at near-surface atmospheric levels. In the paper, it is also noted that the troposphere shows no significant temperature perturbation. However, a 27-day cycle in the temperature response at this level (and all lower atmospheric levels) is used as evidence for a physical link to the IMF B_y .

Two different analysis methods are typically used to demonstrate this effect. The first is the superposed epoch method (Mansurov et al., 1974; Lam et al., 2013, 2014). The pressure/geopotential height on days with strong positive B_y deflections are binned, where the pressure/geopotential height on the days with strong negative B_y deflections are binned and subtracted from the first bin. This can be represented by the formula $\Delta p = B_y(+)-B_y(-)$. The day of the largest deflections is marked as the key date, while different lead–lags are calculated with respect to the key date (similar to time-lagged cross-correlation). The second method is lead–lag regression plots (B2008). Here, the average pressure/geopotential height is calculated in five B_y bins (<-3 , -3 to -1 , -1 to 1 , 1 to 3 , >3 (nT)), and the slope of the regression line between the averaged B_y bins and the corresponding average pressure/geopotential height (regressing 5 data points) is calculated and plotted for chosen daily leads and lags (also similar to time-lagged cross-correlation). We emphasize that both methods yield approximately the same results, as the slope of the regression line strongly depends on the pressure/geopotential height in the lowest and highest B_y bins.

This paper revisits the Mansurov hypothesis and previously applied methods with a more rigorous estimate of the statistical significance. Emphasis is also put on time periods other than solar cycle 23 (1995–2005). In addition, we examine the lead–lag regression method with the help of Monte Carlo simulations and randomly generated normally distributed temporally uncorrelated (white) noise and autocorrelated (red) noise. The aim is to demonstrate the need for appropriate significance tests, as well as the risk of misinterpreting a response from strongly periodic forcing. The implication of these findings goes beyond the current study as it will apply to all periodic forcing with an autocorrelated response variable.

2 Data and method

2.1 Solar wind (B_y) data

We use hourly averaged IMF B_y (GSM) values obtained from the National Space Science Data Center (NSSDC) OMNI-Web database (<http://omniweb.gsfc.nasa.gov>) for the interval 1980–2016. IMF B_y daily averages are calculated when at least 1 hourly value is available.

2.2 Pressure/geopotential height data

For the atmospheric data, we use the European Center for Medium-Range Weather Forecast Re-Analysis (ERA5) (<https://cds.climate.copernicus.eu>). As well as being constructed by numerical simulations and models, ERA-5, and all other reanalysis data, uses large amounts of observational values to set the frame. Effectively, the numerical simulations and models work to interpolate the gaps between these observations. Thus, reanalysis data does not have the same accuracy as purely observational data at every grid point. However, it provides a physically justified estimate in these grid points where observations are not available. It is noted that reanalysis data have previously been applied to support the Mansurov effect, particularly ERA5 (Zhou et al., 2018) and NCEP/NCAR (Lam et al., 2013, 2014, 2018; Freeman & Lam, 2019). Mooney et al. (2011) have compared NCEP/NCAR reanalysis data with earlier ERA reanalysis versions, as well as observational data, finding good agreements between all.

We obtain the daily averaged geopotential heights at the 700 hPa (SH) and 1000 hPa (NH) level poleward of 70° in geomagnetic coordinates (mlat), covering the time period 1980–2016. The 700 hPa level is chosen for the SH as it represents the surface level in the Antarctic, while 1000 hPa represents the surface level in the NH. Geomagnetic coordinates are used as the perturbation of IMF B_y in the ionosphere is centered around the geomagnetic pole. For comparison, B2008 used surface pressure measurements obtained for 11 Antarctic sites from the NNDC (NOAA [National Oceanic and Atmospheric Administration] National Data Centers), selecting values within 90 min of 12 UT. An analog to the quantity Δp (pressure anomalies) that B2008 calculated, a variation value ΔZ_g (geopotential height anomalies) is obtained for the geopotential height by subtracting a running mean of ± 15 days from the daily value in order to remove seasonal variability. It is noted that ΔZ_g is averaged over 70 – 90° mlat.

Figure 1 shows the temporal autocorrelation in ΔZ_g (geopotential height anomalies) for the period 1980–2016 in the SH. Positive auto-correlation occurs until day 5. A similar autocorrelation is also found for the period 1995–2005, as well as for ΔZ_g in the NH.

2.3 False detection rate method

For rigorous statistical testing of our results, we use the False Detection Rate (FDR) method. It was developed by Benjamini & Hochberg (1995) and later applied to atmospheric data by Wilks (2016). The main goal of the method is to account for the expected proportion of falsely rejected hypotheses when dealing with multiple null hypotheses scenarios. Statistically speaking, a result obtaining a p -value of 0.05

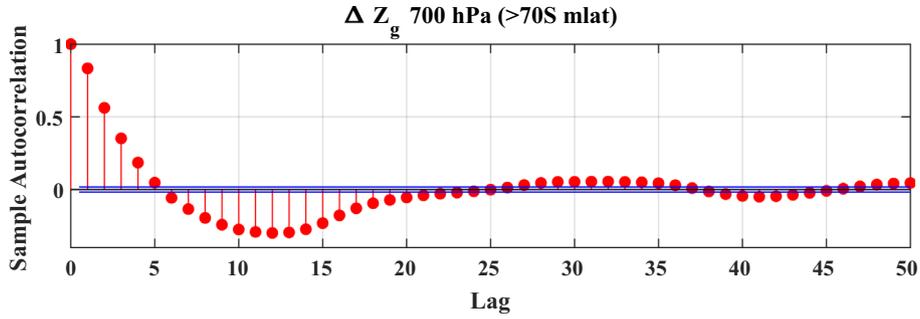


Fig. 1. Temporal autocorrelation of ΔZ_g over the period 1980–2016. Positive auto-correlation occurs until day 5. The blue lines show the 95% confidence bounds of the autocorrelation function.

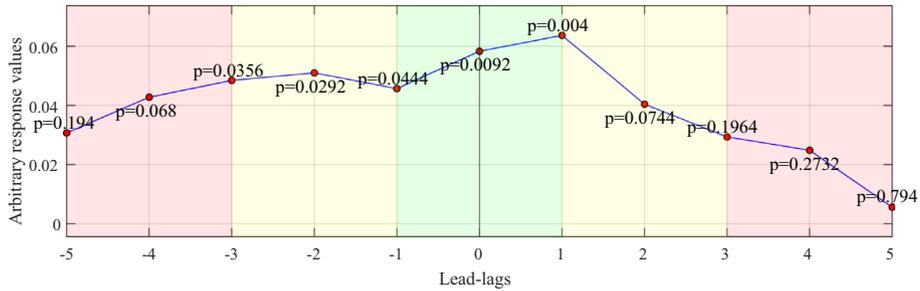


Fig. 2. Arbitrary response values on a temporal lead–lag x -axis (e.g. days). Every data point has also appointed a p -value. Different lead–lag intervals are shaded in different colors.

Table 1. FDR based sorting of p -values for the whole interval in ascending order.

n	1	2	3	4	5	6	7	8	9	10	11 = N
p -values	0.004	0.009	0.029	0.035	0.044	0.068	0.074	0.194	0.196	0.273	0.794

implies a 5% probability of that specific result being caused by chance. With an increasing number of null hypotheses (e.g., map plot with multiple grids or a temporal plot showing consecutive days after the onset of a forcing), this 5% probability ultimately leads to an increasing number of falsely rejected null hypotheses.

In FDR, it is stated that if the global null hypothesis cannot be rejected, one cannot conclude that any of the individual tests constitute rejection of the null hypothesis. The method is applied by calculating the p -values for each individual data point. These p -values are then sorted in ascending order, matching the set $i = 1, \dots, N$, where N represents the total number of individual tests. The new global p -value, p_{FDR} ;

$$p_{\text{FDR}} = \max[p(i) : p(i) \leq (i/N)\alpha_{\text{FDR}}], \quad i = 1, \dots, N \quad (1)$$

is then calculated with $\alpha_{\text{FDR}} = 0.05$, corresponding to significance at the 95% level (Wilks, 2016).

Figure 2 illustrates how FDR is used and calculated for a superposed epoch analysis on a daily scale represented by lead–lags. Also included are p -values obtained for each lead–lag. In this example, there is an arbitrary forcing that is nonzero and starts to increase at day -5 , reaching a maximum at day 0, before it slowly decreases to zero at day $+5$. We also assume that the arbitrary forcing has an impact on the arbitrary response as long as it is nonzero. As the forcing is nonzero through the whole interval, we can also assume that every individual lead–lag has the same null hypothesis and that we are dealing

with a multiple hypotheses situation for lead–lags -5 to $+5$. According to FDR, we first have to sort the p -values for the whole interval in ascending order (see Table 1).

Then, we have to apply Equation (1) iteratively until we reach the maximum p -value satisfying the criteria:

$$P(n)_{\text{ascending}} \leq n \times \frac{0.05}{N}, \quad (2)$$

$$0.004 \leq 1 \times \frac{0.05}{11} = \text{True} \quad (3)$$

$$0.009 \leq 2 \times \frac{0.05}{11} = \text{True} \quad (4)$$

$$0.029 \leq 3 \times \frac{0.05}{11} = \text{False} \quad (5)$$

⋮

Everyother is also False.

As the $p = 0.009$ is the maximum value satisfying the criteria, this becomes the global p -value (p_{FDR}) and defines the limit for the individual p -values to be regraded as significant at the 0.05 level after one has accounted for the false detection rate. In our example, this means that when the signal is looked at as a set of multiple equivalent null hypotheses, statistical

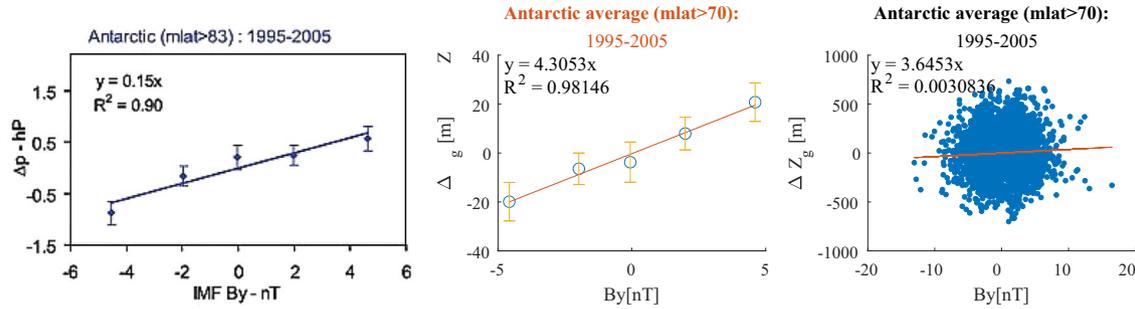


Fig. 3. *Left panel:* A copy of the upper panel of Figure 1 in B2008. It represents linear regression of Δp after the original measurement from three Antarctic stations at $m\text{lat} > 83^\circ \text{ S}$ was grouped according to the IMF B_y . *Middle panel:* Reproduction of the linear regression method using ΔZ_g at $\sim m\text{lat} > 70^\circ \text{ S}$. Error bars are plus/minus one standard-error-in-the-mean. *Right panel:* Scatter plot and linear regression for the ΔZ_g data without the initial five-bin grouping. The upper panel of Figure 1 in B2008 is reproduced with permission from John Wiley and Sons.

significance is found at lead-lag 0 and +1. As we know the onset and offset of the forcing, this could be interpreted as lead-lag 0 and +1 being the only days where it is possible to distinguish a signal from the background noise in the data.

For this method to be correctly applied, it is important that the definition of equivalent null hypotheses is correct. For instance, assuming only three consecutive days around day zero (-1 to $+1$) to have equivalent null hypotheses, and performing the FDR method, would result in all of them satisfying the criteria ($p_{\text{FDR}} = 0.044$). This would yield one more significant data point than what was acquired when the full interval -5 to $+5$ was grouped as a whole through the FDR method. Because of this, we will be testing different intervals when estimating the significance using the FDR method in lead-lag correlation plots in the following section. Multiple hypothesis testing situations can also be dealt with other methods than FDR, e.g., calculating a field significance or effective spatial degrees of freedom (Bretherton et al., 1999). While the FDR method is not yet well known in the atmospheric or space science communities, it offers a simple but superb way to deal with multiple hypothesis testing scenarios (Wilks, 2016).

3 Analyses and results

3.1 Regression results for the time period 1995–2005

Based on observations from the 11 Antarctic stations, B2008 calculated the average Δp (surface pressure) values at each site within five separate IMF B_y bins: < -3 , -3 to -1 , -1 to 1 , 1 to 3 , and > 3 nT. Linear regression was then applied to the average value of Δp within these five intervals. The result for $> 83^\circ \text{ S}$ $m\text{lat}$, corresponding to the upper panel of Figure 1 in B2008, is shown in the left panel in Figure 3. The same procedure is done for ΔZ_g (equivalent to surface pressure), seen in the middle panel in Figure 3. Also included is a linear regression without the initial binning and averaging, as seen in the right panel in Figure 3. Note that the regression coefficients are similar with or without performing the initial binning, while the correlation coefficient (R^2) differs substantially.

From the regression coefficient produced by these five data bins, lead-lag variations are calculated by B2008, as seen in the left panel of Figure 4. A clear 27-day cycle is seen for both data

sets, with the peak pressure value lagging the driver by -2 days. The significance has been estimated by Student’s t -test, with the uncertainty illustrated by the cross at the keydate. Figures 3 and 4 indicate that ΔZ_g yields a similar response as Δp in B2008. Furthermore, note that the normal regression without the initial grouping gives similar lead-lag regression coefficients.

When applying the t -test, a highly significant pattern is observed, as shown in the right panel of Figure 4. However, the lead-lag analysis is strongly affected by the temporal autocorrelation in the ΔZ_g time series (Fig. 1). Instead of a t -test, we perform a Monte Carlo (MC) simulation to estimate the significance of the regression coefficients. For every iteration of the MC-simulation, phase randomization is applied to the ΔZ_g data series. In essence, phase randomization scrambles the harmonic phases of the series. This results in a physically unrelated data series but preserves the autocorrelation function of ΔZ_g , which gives the phase randomized series the same number of independent data points as ΔZ_g . This process ensures that the MC simulation can perform the null hypothesis test on statistically suitable material (Theiler & Prichard, 1996; Thejll et al., 2003). Before the B_y series is regressed onto the phase randomized ΔZ_g for every lead-lag, both data sets are standardized by subtracting their means and dividing by their standard deviations. This will ensure that the regression slope equals the linear correlation coefficient (Rodgers & Nicewander, 1988). The same standardization is also performed on the actual response (ΔZ_g) (transforming the regression slopes to correlation coefficients) before the actual result is compared to the distribution of correlation coefficients obtained from the MC simulation in each lead-lag. The fraction of correlation coefficients from the MC simulation with higher values than the actual response will represent the p -value.

Figure 5 shows the results after 3000 iterations of the MC simulation. The green shaded area shows the interval corresponding to 95% of the values from all iterations. The red shaded area shows above(below) the 97.5% (2.5%) percentile, corresponding to a p -value smaller or equal to 0.05 (both tails of the distribution). As can be seen, the significance is reduced compared to what is obtained by the t -test. Also, the peak around day 0 is only found significant at the 95% level for two data points, occurring at day -2 and -1 . However, multiple points with 95% significance are obtained at the peaks around -27 and $+27$ days, along with the minimum around

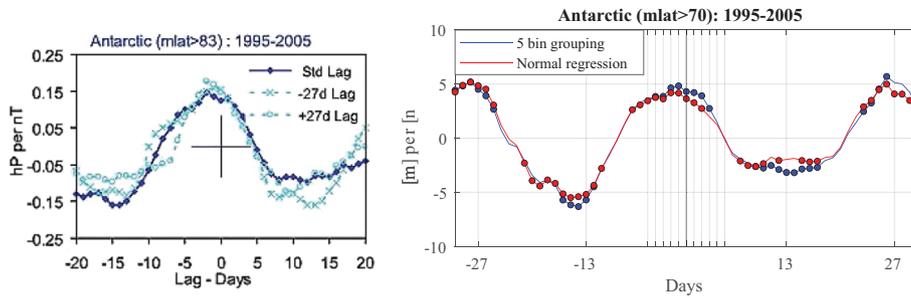


Fig. 4. *Left panel:* A copy of the upper panel of Figure 2 in B2008. The figure illustrates calculated regression coefficients showing lead–lag variations of Δp at $mLat > 83^\circ S$. It shows three cycles of IMF B_y , where the dark blue line represents the regression coefficients without any lag, while x and o cyan lines represent a -27 and $+27$ day lag between IMF B_y and Δp data series. All maxima in Δp are seen to occur -2 days before the peak in the IMF driver, which occurs at day 0. *Right panel:* Lead–lag variations of ΔZ_g at $mLat > 70^\circ S$. The blue line is the calculated regression coefficients showing lead–lags when the five bin method by B2008 is used. The red line is the regression coefficients showing lead–lag variations when regression is done without the initial grouping. Negative days (leads) represent ΔZ_g occurring before the B_y component, and positive days B_y occurring before ΔZ_g . Dots indicate significance at the 95% level for the regression coefficients calculated by Student’s t -test. The upper panel of Figure 2 in B2008 is reproduced with permission from John Wiley and Sons.

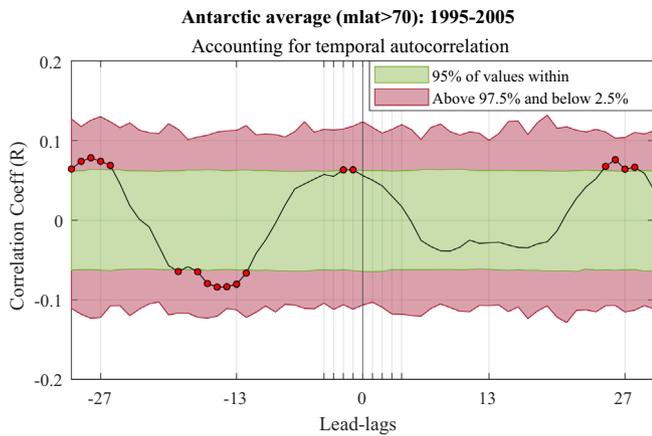


Fig. 5. The significance level for the lead–lag correlation coefficients after 3000 MC-iterations for the period 1995–2005. The red area equates to a p -value of 0.05. The green region shows where 95% of all values land for every lead–lag after 3000 iterations. Note that the significant data points (dark red circles) represent individual hypothesis tests before false detection rate method is applied.

-13 days. For day -2 the correlation coefficient is equal to 0.064: for days -15 , -27 , and $+27$, it is approximately 0.08. This implies that B_y can explain less than one percent of the pressure variability ($R^2 < 0.01$).

B2008 refers to the apparent periodic response in Figure 5 as support for B_y forcing. Furthermore, B2008 results, shown in Figure 3, include 95 tests of individual null hypotheses (one for each lead–lag regression), while 55 are included in our replication given in Figure 4. In both our and B2008 results, we have the strange phenomena of the peak pressure response occurring before the peak forcing. We also obtained higher correlation coefficients at day -27 and $+27$, which are days where the forcing is actually weaker than at day 0. Together with the B_y being continuous, a reasonable assumption is that the forcing always has an impact through this period and would render all null hypotheses in the interval -27 to $+27$ ($N = 55$) equivalent. Another assumption can be derived from the fact that as the

IMF B_y has a 27-day periodicity, one can assume that the forcing is mostly positive for the interval -13 to $+13$ ($N = 27$); this also takes into account a longer time delay for the response to occur. The last suggestion would be to only look at the interval -2 to $+2$ ($N = 5$), as this is when the proposed forcing peaks. Here we also capture the two significant data points after the MC-simulation at lead–lag -2 and -1 . According to theory, it takes days before the accumulative effect on cloud properties leads to pressure changes (Frederick et al., 2019; Tinsley et al., 2020). Hence, a reasonable window would also be from day 0 and some days onwards. However, no significant (after MC) pressure peak occurs from day 0 and onwards. As of this, doing the FDR for lead–lag 0 and some days onward makes no sense.

When the FDR method is applied, no significance is obtained at the 95% level for any lead–lag in the period 1995–2005 for any of the suggested intervals. This means that the response as a whole cannot be assumed to be statistically significant. However, one must note that if only a single lead or lag (e.g., leads -2 or -1) is presented, the significance at the 95% level is justified (see Eq. (1)). However, from a physical perspective, it is hard to justify the response occurring 1 or 2 days (or more than 12 days) before the forcing instead of at day 0 or after.

Figure 6 shows the same procedure for the period 1999–2002 previously investigated by e.g., Burns et al. (2008) and Lam et al. (2013, 2014). After 3000 MC iterations, only 1 significant data point remains close to day 0 in the SH (top left panel), and 2 remain in the NH (top right panel). However, the application of FDR shows that no leads or lags that by themselves are above the 95% significance level constitute evidence in favor of rejecting the global null hypothesis in any of the hemispheres (bottom panels). This is true whether we calculate p_{FDR} for lead–lag intervals -27 to $+27$ ($N = 55$), -13 to $+13$ ($N = 27$) or even for -2 to $+2$ ($N = 5$) ($+2$ to $+6$ ($N = 5$) for the SH). Although the correlation coefficients for this period are not inconsistent with a physical effect, as the peak ΔZ_g anomaly occurs after day 0 in both hemispheres, they are not significant in regards to the rejection of the global null hypothesis.

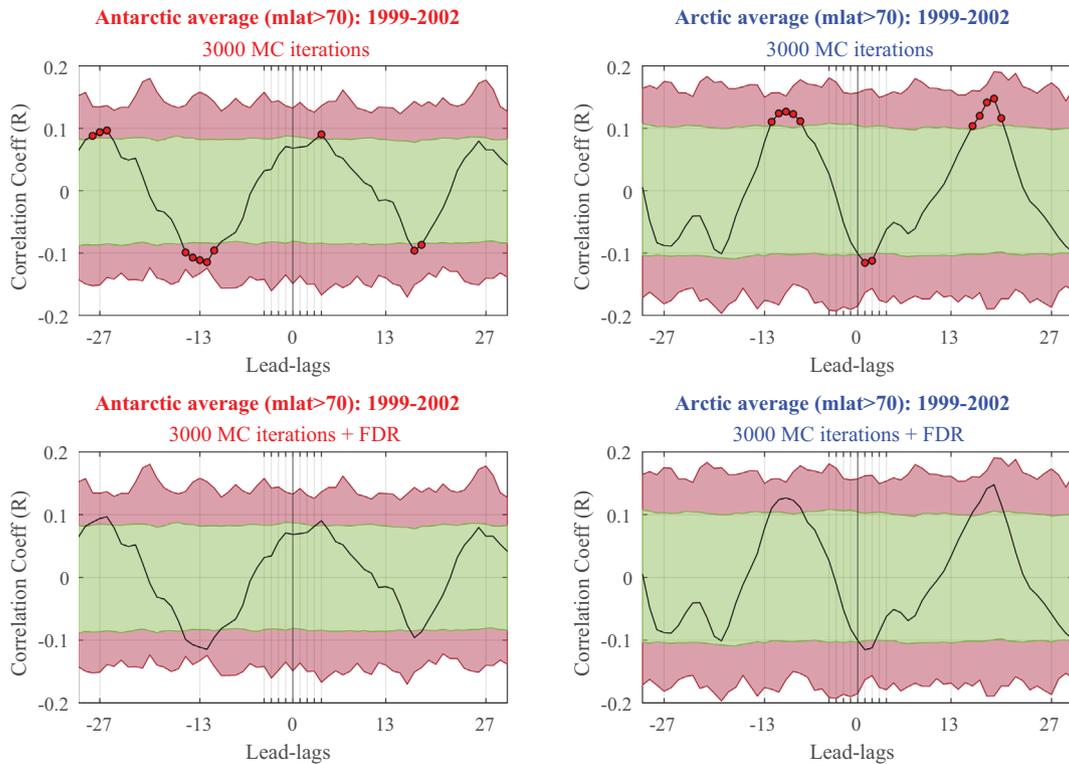


Fig. 6. *Left panels:* The significance level for the lead–lag correlation coefficients after 3000 MC-iterations for the period 1999–2002 in the SH. Dark red circles indicate 95% significance of the individual hypothesis tests (*top panel*). No significance is obtained after FDR. This is the case whether FDR is computed for the interval -27 to $+27$ ($N = 55$), -13 to $+13$ ($N = 27$) or $+2$ to $+6$ ($N = 5$) lead-lags (*bottom panel*). *Right panels:* Same procedure, only for the NH (*top panel*). No significance is obtained after FDR. This is the case whether FDR is computed for the interval -27 to $+27$ ($N = 55$), -13 to $+13$ ($N = 27$) or -2 to $+2$ ($N = 5$) lead-lags (*bottom panel*).

3.2 Other time periods

Figure 7 shows the correlation between ΔZ_g and B_y for the periods 1984–1994, 1995–2005, and 2006–2016 in both hemispheres (top panels). The bottom panels show the same, only for 4-year periods centered around four different solar maxima. Nearly all of the time periods in both hemispheres show cyclic responses exhibiting a periodicity of ~ 27 days. However, none of the time periods outside of solar cycle 23 (1995–2005 or 1999–2002) show responses supported by the theory (positive response in the SH and negative response in the NH at day zero or shortly after). Instead, the peaks occur seemingly at random but with an apparent periodicity of approximately 27 days.

3.3 Monte Carlo simulations with different levels of temporal autocorrelation

Figure 7 demonstrates that the periodic response in ΔZ_g of ~ 27 days is not unique to the 1995–2005 period, as it occurs in other time periods as well. Since the responses do not seem to have any relation to the forcing (day 0), the resulting cyclic response could be an artifact of the method itself, enhanced by the high temporal autocorrelation of the explanatory variable.

Figure 8 shows the power spectrum (left panel) and the autocorrelation function (right panel) of the IMF B_y over the time period 1995–2005. A strong 27-day solar rotation periodicity can be observed in both. When the regression coefficients

for lead–lag variations are calculated, one data set is moved with respect to the other, where the regression coefficient is calculated for each lag between the data sets. In essence, this can lead to the responses seen at day ± 27 days, being partially replications of the response seen at day 0, occurring as a consequence of the periodicity of the forcing. This is especially relevant if the response variable has a strong temporal autocorrelation.

To demonstrate this, we calculate three Monte Carlo simulations with varying levels of autocorrelation of the response variable. For all cases, the geopotential height data (ΔZ_g) is replaced by randomly generated normally distributed noise with the same length as the 1995–2005 period. For the first, second, and third cases, lag-1 autocorrelation is set to 0, 0.5, and 0.94, respectively. An autocorrelation of 0 represents a data set of normally distributed white noise, while the autocorrelation of 0.94 reflects the autocorrelation seen in the original geopotential height data series (not shown). The ± 15 -day moving average is further subtracted from the three random data series, analog to the calculation of ΔZ_g .

For all three cases, 1000 independent Monte Carlo iterations are run. For each run, we calculate the lead–lag correlation coefficients between the real B_y forcing in the period 1995–2005 and the randomly generated data series. Figure 9 summarizes the results. The first column represents the lead–lag correlation coefficients for all runs in the three cases. The lead–lag curves appear to be random. However, if each curve is shifted such that the maximum value occurring inside the range

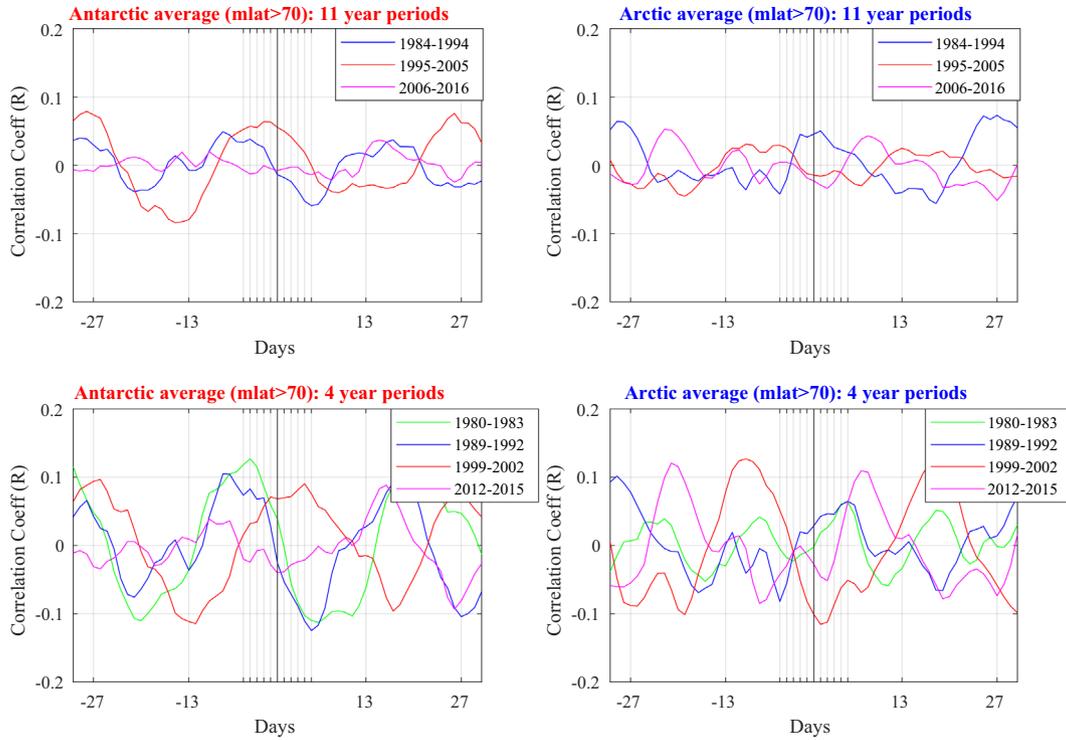


Fig. 7. Lead-lag correlation coefficients between ΔZ_g and B_y in both hemispheres for three 11-year periods spanning 1984–2016 (*top panels*) and four 4-year periods centered around solar maximum (*bottom panels*).

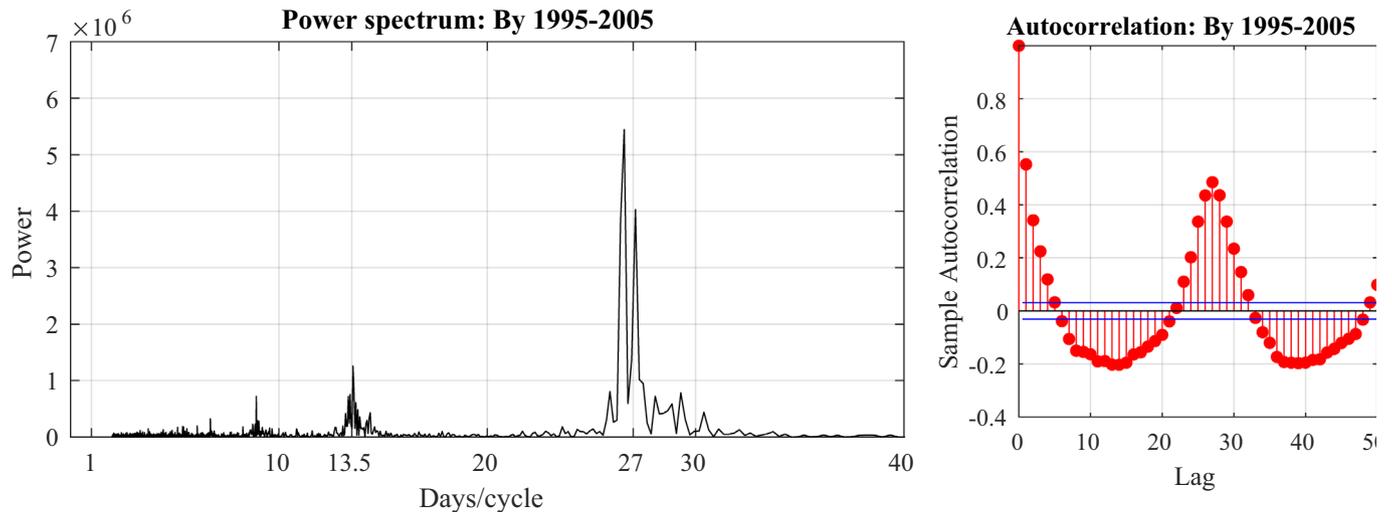


Fig. 8. *Left panel:* Power spectrum of the IMF B_y -index in the time period 1995–2005. *Right panel:* Autocorrelation function of the IMF B_y -index in the time period 1995–2005. The blue lines show the 95% confidence bounds of the autocorrelation function.

(−13, 13) days from day 0 is shifted to day 0, a pattern emerges. This is illustrated in the middle row of panels. When the responses are averaged over all independent simulations, as shown on the right, the resulting average lead-lag curve exhibits a periodicity equal to the periodicity of B_y . Furthermore, it is apparent that the higher the autocorrelation of the random data series at lag-1, the larger the amplitudes of the artificially created response. It is particularly interesting that the correlation

coefficients in [Figure 7](#) are comparable to the correlation coefficients resulting from the third artificial case (lag-1 autocorrelation = 0.94) in [Figure 9](#).

[Figure 9](#) clearly shows that the 27-day cyclic response in surface pressure to the B_y -component cannot be used as a strong argument supporting the Mansurov effect. Furthermore, it clearly demonstrates the necessity of using FDR or a similar method when estimating the significance of the response.

4013 data points (By data from the period 1995–2005)

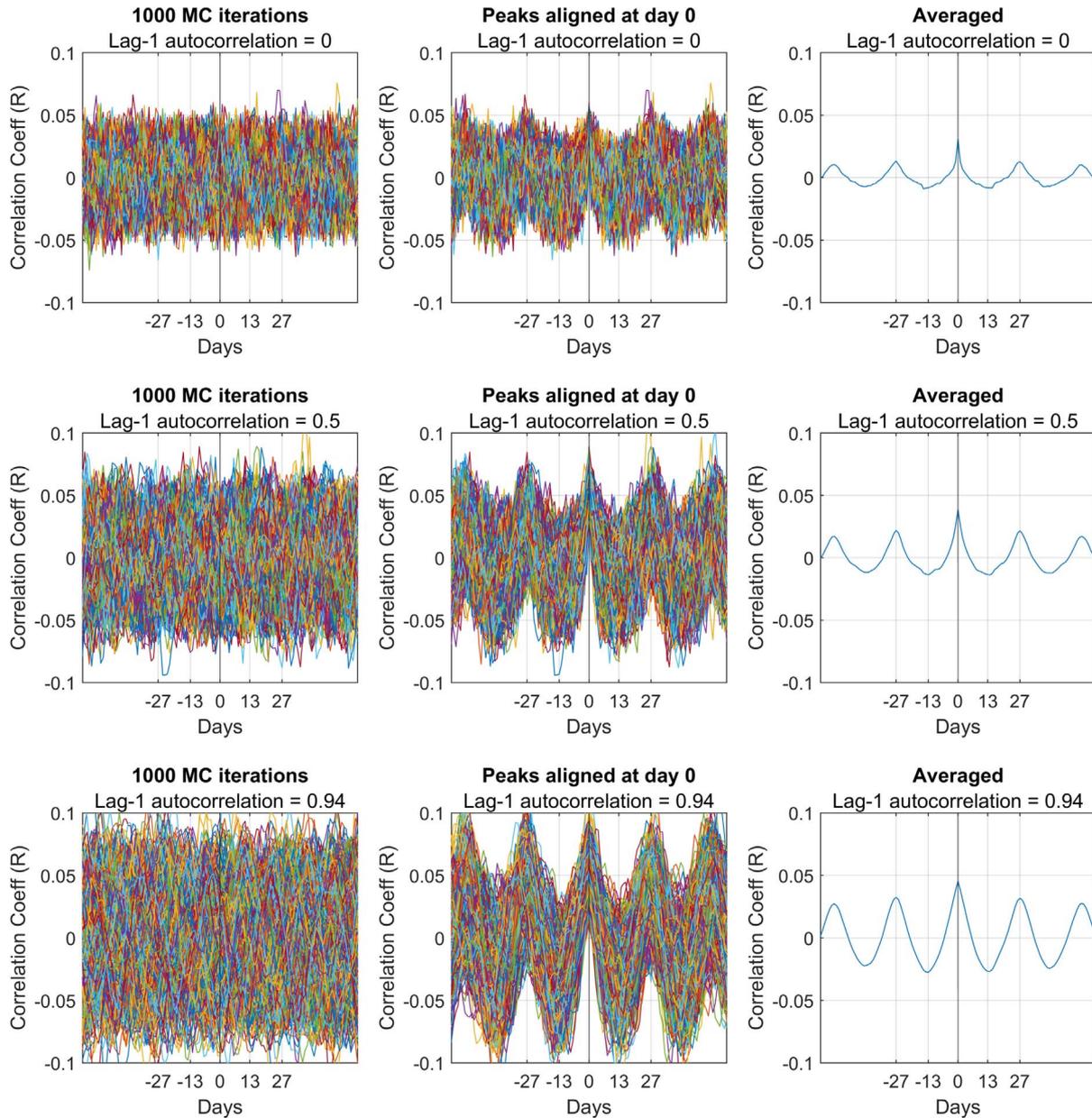


Fig. 9. *Left panels:* 1000 MC iterations where the correlation coefficients are calculated between the B_y data in the period 1995–2005 and normally distributed noise with three different lag-1 autocorrelation values (0, 0.5, 0.94) for every lead–lag between -60 and $+60$. *Middle panels:* All 1000 individual lead–lag plots aligned such that the maximum value within -13 to $+13$ is projected to day 0. *Right panels:* Averaged response of the middle panels.

4 Discussion

The aim of this paper is to demonstrate the need for appropriate significance tests, as well as the risk of misinterpreting a response from a strongly periodic forcing when studying the Mansurov effect (and also, more generally, any phenomena in cases of strong temporal autocorrelation). Figure 3 shows that similar values for the regression slopes are obtained with the five-bin grouping used by B2008 and the normal regression.

However, the explanatory power of the two models largely depends on whether or not the measurements are binned (with binning $R^2 = 0.99$, without binning $R^2 = 0.0033$). Further, both the five-bin grouping and the normal regression produce similar lead–lag plots, as illustrated by Figure 4. Therefore, it is clear that the five-bin grouping gives the impression of a significantly better fit than it is found in the original data.

The majority of the research articles on the Mansurov effect focus on solar cycle 23 (B2008; Lam et al., 2013, 2014, 2018;

Zhou et al., 2018). We showed, however, that simple t -tests are not sufficient to establish significance for the link between the IMF B_y and the geopotential height variability at the polar surface. By applying MC simulations to validate the null hypotheses in addition to the false detection rate method, we showed that neither the period 1995–2005 nor the solar maximum period 1999–2002 indicate a statistically significant response. This remains true as long as the response is analyzed with multiple leads and lags greater or equal to 5 days, as the individual p -values exceed the global p -value (Eq. (1)) even for -2 to $+2$ lead-lags in all cases for solar cycle 23. Nonetheless, if only a single lead or lag is presented, the significance at the 95% level obtained by the MC simulation alone would be justified. During the period 1995–2005, the points with high statistical significance at leads -2 or -1 are hard to justify on physical grounds, as the surface pressure effect occurs before the forcing. However, individual significant data points obtained in the SH (day +4) and NH (day +1 and +2) for the period 1999–2002 cannot be completely discarded from the viewpoint of a single null hypothesis, as the effect occurs after the forcing.

By similar methodology, we observe periodic geopotential height responses in both hemispheres in other time periods, but with varying offset in respect to the forcing, as illustrated by Figure 7. The geopotential height deflections are also fairly equal to the amplitudes seen for solar cycle 23. Hence, the cyclic responses seen in solar cycle 23 are not unique to this period.

B2008, Lam et al. (2018) and Tinsley et al. (2020) all use this 27-day periodicity in the results as evidence in favor of the Mansurov effect. By using MC simulations of randomly generated data series with different levels of lag-1 autocorrelation, we showed that plotting lead-lag regression coefficients for a highly periodic forcing produces periodic responses, even when no physical relationship is present (Fig. 9). The periodic response always mimics the periodicity of the variable used as the forcing. One can also observe how this cyclic response is enhanced by a higher autocorrelation of the response variable. From this perspective, the alignment of the period 1999–2002 with the theory could, in fact, be a coincidence (1995–2005 is also approximately aligned with the theory in the SH). This result extends beyond the Mansurov effect itself and is applicable in any case where the relationship between a periodic explaining variable and an autocorrelated response variable is examined on a temporal scale.

However, the effect could be nonstationary in relation to atmospheric variability and the solar phases. If so, time periods restricted by similar atmospheric and solar conditions would be expected to respond in a similar manner, while averages of large continuous time periods would smoothen out the effect, making it much harder to detect. Tinsley et al. (2020) found a higher correlation between cloud irradiance and changes in the vertical electric field related to B_y during local northern winter (Oct–Apr 2004–2015), then local summer months. However, no statistical assessment of the correlation coefficients in respect to the temporal autocorrelation was made. An equal probable explanation for the larger coefficients could be the higher atmospheric variability in winter compared to summer. This could lead to higher levels of noise in the results, which are artificially replicated into a periodic response via the method used, as our results show. In agreement with Tinsley et al. (2020) and Zhou

et al. (2018) also found results with local winter in both hemispheres producing the largest response between the vertical electric field and surface pressure. However, only the period 1998–2001 is analyzed, and the results lack proper statistical testing. Sorting according to non-stationary behavior is beyond the scope of this article but is a recommended pathway for further research on the Mansurov effect, as the articles discussed here are pointing to a potential seasonal variability. However, future studies need to take into account the autocorrelation of variables and multiple hypothesis testing scenarios when assessing the statistical significance of their results.

5 Conclusion

We revisited the previous evidence suggesting a significant link between the IMF B_y and the surface pressure/geopotential height variability. We showed that after the pressure/geopotential height and IMF B_y data were subjected to rigorous estimation of statistical significance, evidence for the Mansurov effect during solar cycle 23 was not found when considering the whole year without individual seasons/months. In addition, our analyses showed that other time periods (before and after solar cycle 23) produced cyclic responses with a similar magnitude but with random offset with respect to the IMF B_y forcing. We also provided evidence showing that high temporal autocorrelation of variables can explain the cyclic responses without the need for a physical connection between the variables. These results underline the importance of robust statistical methods, especially when analyzing periodic variables or data with high temporal autocorrelation.

For the Mansurov effect, our applied methods indicate that even if a connection between IMF B_y changes and cloud microphysics exists, this effect is not strong enough to produce significant correlations for a stationary signal in surface polar geopotential height/pressure over interannual to decadal time-scales. We encourage more research on the topic to assess the potential cause of non-stationary behavior and seasonal variability.

Acknowledgements. We thank the ECMWF (European Center for Medium Weather Forecast) for ERA5 data (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>) and the NASA Goddard Space Center for OMNIWeb database (<https://omniweb.gsfc.nasa.gov/>). All data used in this study are openly available. All codes and data required to reproduce the results of this study can be downloaded from Zenodo (<https://doi.org/10.5281/zenodo.5996692>). The research was funded by the Norwegian Research Council under contracts 223252/F50 (BCSS) and 300724 (EPIC). The editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Statist Soc: Ser B (Methodol)* **57**: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.

- Bretherton CS, Widmann M, Dymnikov VP, Wallace JM, Bladé I. 1999. The effective number of spatial degrees of freedom of a time-varying field. *J Climate* **12**(7): 1990–2009. [https://doi.org/10.1175/1520-0442\(1999\)012<1990:TENOSD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2).
- Burns GB, Tinsley BA, French WJR, Troshichev OA, Frank-Kamenetsky AV. 2008. Atmospheric circuit influences on ground-level pressure in the Antarctic and Arctic. *J Geophys Res* **113**: D15112. <https://doi.org/10.1029/2007JD009618>.
- Frank-Kamenetsky AV, Troshichev OA, Burns GB, Papitashvili VO. 2001. Variations of the atmospheric electric field in the near-pole region related to the interplanetary magnetic field. *J Geophys Res* **106**: 179–190. <https://doi.org/10.1029/2000JA900058>.
- Frederick JE, Tinsley BA, Zhou L. 2019. Relationships between the solar wind magnetic field and ground-level longwave irradiance at high northern latitudes. *J Atmos Sol-Terr Phys* **193**: 105063. <https://doi.org/10.1016/j.jastp.2019.105063>.
- Freeman MP, Lam MM. 2019. Regional, seasonal, and inter-annual variations of Antarctic and sub-Antarctic temperature anomalies related to the Mansurov effect. *Environ Res Commun* **1**: 111007. <https://doi.org/10.1088/2515-7620/ab4a84>.
- Gonzalez ALC, Gonzalez WD. 1987. Periodicities in the interplanetary magnetic field polarity. *J Geophys Res* **92**(A5): 4357–4375. <https://doi.org/10.1029/JA092iA05p04357>.
- Kabin K, Rankin R, Marchand R, Gombosi TI, Clauer CR, Ridley AJ, Papitashvili VO, DeZeeuw DL. 2003. Dynamic response of Earth's magnetosphere to B_y reversals. *J Geophys Res* **108**: 1–13. <https://doi.org/10.1029/2002JA009480>.
- Lam MM, Tinsley BA. 2016. Solar wind-atmospheric electricity cloud microphysics connections to weather and climate. *J Atmos Sol-Terr Phys* **149**: 277–290. ISSN: 1364-6826. <https://doi.org/10.1016/j.jastp.2015.10.019>.
- Lam MM, Chisham G, Freeman MP. 2013. The interplanetary magnetic field influences mid-latitude surface atmospheric pressure. *Environ Res Lett* **8**: 045001. <https://doi.org/10.1088/1748-9326/8/4/045001>.
- Lam MM, Chisham G, Freeman MP. 2014. Solar-wind-driven geopotential height anomalies originate in the Antarctic lower troposphere. *Geophys Res Lett* **41**: 6509–6514. <https://doi.org/10.1002/2014GL061421>.
- Lam MM, Freeman M, Chisham G. 2018. IMF-driven change to the Antarctic tropospheric temperature due to the global atmospheric electric circuit. *J Atmos Sol-Terr Phys* **180**: 148–152. <https://doi.org/10.1016/j.jastp.2017.08.027>.
- Mansurov SM, Mansurova LG, Mansurov GS, Mikhnevich VV, Visotsky AM. 1974. North-south asymmetry of geomagnetic and tropospheric events. *J Atmos Terr Phys* **36**(11): 1957–1962. [https://doi.org/10.1016/0021-9169\(74\)90182-2](https://doi.org/10.1016/0021-9169(74)90182-2).
- Mooney P, Mulligan F, Fealy R. 2011. Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *Int J Climatol* **31**: 545–557. <https://doi.org/10.1002/joc.2098>.
- Page DE. 1989. The interplanetary magnetic field and sea level polar atmospheric pressure. In: *Workshop on mechanisms for tropospheric effects of solar variability and the quasi-Biennial oscillation*, Avery SK, Tinsley BA (Eds.), University of Colorado, Boulder, CO, USA, 22 p.
- Pettigrew ED, Shepherd SG, Ruohoniemi JM. 2010. Climatological patterns of high-latitude convection in the Northern and Southern hemispheres: Dipole tilt dependencies and interhemispheric comparisons. *J Geophys Res* **115**: A07305. <https://doi.org/10.1029/2009JA014956>.
- Rodgers JL, Nicewander AW. 1988. Thirteen ways to look at the correlation coefficient. *Am Statist* **42**: 59–66. <https://doi.org/10.1080/00031305.1988.10475524>.
- Theiler J, Prichard D. 1996. Constrained-realization Monte-Carlo method for hypothesis testing. *Phys D* **94**: 221–235. [https://doi.org/10.1016/0167-2789\(96\)00050-4](https://doi.org/10.1016/0167-2789(96)00050-4).
- Thejll P, Christiansen B, Gleisner H. 2003. On correlations between the North Atlantic Oscillation, geopotential heights, and geomagnetic activity. *Geophys Res Lett* **30**: 1347. <https://doi.org/10.1029/2002GL016598>.
- Tinsley BA. 2000. Influence of solar wind on the global electric circuit, and inferred effects on cloud microphysics, temperature, and dynamics in the troposphere. *Space Sci Rev* **94**: 231–258. <https://doi.org/10.1023/A:1026775408875>.
- Tinsley BA. 2008. The global atmospheric electric circuit and its effect on cloud microphysics. *Rep Prog Phys* **71**: 66801–66831. <https://doi.org/10.1088/0034-4885/71/6/066801>.
- Tinsley BA, Zhou L, Wang L, Zhang L. 2020. Seasonal and solar wind sector duration influences on the correlation of high latitude clouds with ionospheric potential. *J Geophys Res: Atmos* **126**: e2020JD034201. <https://doi.org/10.1029/2020JD034201>.
- Wilks DS. 2016. “The stippling shows statistically significant grid points”: How research results are routinely overstated and over interpreted, and what to do about it. *Bull Am Meteorol Soc* **97**: 2263–2273. <https://doi.org/10.1175/BAMS-D-15-00267.1>.
- Zhou L, Tinsley BA, Wang L, Burns GB. 2018. The zonal mean and regional tropospheric pressure responses to changes in ionospheric potential. *J Atmos Sol-Terr Phys* **171**: 111–118. <https://doi.org/10.1016/j.jastp.2017.07.010>.

Cite this article as: Edvartsen J, Maliniemi V, Nesse Tyssøy H, Asikainen T & Hatch S 2022. The Mansurov effect: Statistical significance and the role of autocorrelation. *J. Space Weather Space Clim.* **12**, 11. <https://doi.org/10.1051/swsc/2022008>.