Appendix:
Statistical Significance Methods

In Chapter II, I used various methods to compute the significance of the statistically derived quantities. In this Appendix I describe the methods used to derive these significance statistics. For the computation of the significance of the composite anomaly winds and the significance of the correlation coefficients between SOI and number of events, I used standard statistical tests that assume our populations are normally distributed. To test the significance of the monthly distribution of the WWEs I used a bootstrap method.

A.1. Composite wind anomaly significance test

I use a Students-t test to estimate the statistical significance of the composite results, so I require the standard deviation of the zonal and meridional wind anomalies that make up each composite. In particular, I needed to know these standard deviations for every 12 hour period during the 19 days of the composite . These standard deviations were computed as follows:

$$\vec{\sigma}(x,y,t_n) = \left\{ \frac{\sum_{i=1}^{N} \left[\vec{U}(x,y,t_n) - \vec{u}(x,y,\tau_i + t_n) \right]^2}{N-1} \right\}^{\frac{1}{2}}$$

where $\boldsymbol{\sigma}$ is the standard deviation vector, **U**, **u**, x, y, t_n, and $\{\tau_i\}$ and N are the same as in Section II.2, and I define a vector power to be:

$$\vec{\xi} \equiv \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \Rightarrow \vec{\zeta}^{\alpha} \equiv \begin{pmatrix} \xi_1^{\alpha} \\ \xi_2^{\alpha} \end{pmatrix}$$

I performed a Student's-t test (Bickel and Docksun, 1977, p.210-215) on each component of our composite vector wind anomaly field, to determine whether the true mean was distinguishable from zero, to the 99% confidence limit. I use a double sided significance test, since I are interested in wind components that are statistically significant, regardless of the sign. I set the number of degrees of freedom in our test to be the number of individual events that went into generating the composite minus 1, using previous notation: N-1. According to the Student's-t test, our averages were significant to the 99% level if the following inequality was satisfied:

$$\left|\frac{\vec{U}(x,y,t_n)\cdot\hat{e}_j}{\vec{\sigma}(x,y,t_n)\cdot\hat{e}_j}\right| > \frac{t_{99.5\%,N}}{\sqrt{N-1}}$$

Where **U**,**σ**, x, y, t_n and N are as before; $\mathbf{e}_j$ is the unit vector in the zonal (j=1) or meridional (j=2) direction; $t_{99\%,N}$ is the double sided Student's-t coefficient at the 99% significance level for N degrees of freedom; and indicates the vector dot product operation. The values for $t_{99\%,N}$ are tabulated in most statistics or mathematical table handbooks, e.g. Speigel (1994), Zwillinger (1996).

A.2. Correlation coefficient significance test

I performed a Student's-t (Bickel and Docksun, 1977, p.220) test on our sample correlation coefficients, r, to evaluate whether the true correlation coefficient, r, approximated r was distinguishable from zero, to a prescribed significance level, l. I use a double sided significance test because I are interested in significant correlations regardless of the sign of the correlation. The following inequality had to be satisfied in this test:

$$r^2 > \frac{t_{l,N-2}^2}{N-2+t_{l,N-2}^2}$$

where $t_{l,N-2}$ is the double sided Student's-t coefficient for l significance level and for N-2 degrees of freedom. Since in our case N = 10, the threshold for 95% significance is |r|>0.63, and the threshold for 99% significance is |r|>0.77, for the zero lag correlation tests; while for the one year lagged case N=9, so the threshold for 90% significance is |r|>0.58, and for 95% significance |r|>0.67. The values for $t_{l,N-2}$ are tabulated in most statistics or mathematical table handbooks, e.g. Speigel (1994), Zwillinger (1996).

A.3. Monthly distribution significance test

I performed a bootstrap test (Efron and Tibshirani, 1991) on the monthly distribution data to find the probability, P, the "on" season I observed for each event had of occurring randomly. I then took the significance level of the seasonal distribution to be 100∞(1-P) %. The probability was computed using a bootstrap method, with 100,000 bootstrap samples.

The procedure to determine the probability of N_{seas} events being distributed over a continuous L_{seas} month period on a random distribution of N events is as follows. I sample 100,000 times; at each sampling I randomly distribute the N events for the particular event type, into twelve equally likely months. I then test whether there exists a continuous period of L_{seas} months for which the total number of events randomly placed is equal to or greater to N_{seas}. I count the number of times the test turns out true in our bootstrap procedure, and define the bootstrap probability to be P = E/B; where E is the number of positive tests, and B is the number of bootstrap samples. The bootstrap probability converged for B > 1,000, but I took B = 100,000 for good measure.

A.4. Event sequencing significance test

I performed a bootstrap test (Efron and Tibshirani, 1991) to find the probability, P, the sequencing pattern I observed for each event pair had of occurring randomly. I then took the significance level of the event sequencing to be 100∞(1-P) %. The probability was computed using a bootstrap method, with 10,000 bootstrap samples.

I determined the probability of distributing N_1 center days of type 1 and N_2 center days of type 2 on a ten year time axis semi-randomly, so that no two center days of the same type were within seven days of each other, and having M or more ordered pairs of events (n_1,n_2) within three days of each other in the following manner. I sample 10,000 times, at each sampling step semi-randomly distributing N_1 and N_2 events as just described. I then test to see how many ordered pairs (n_1,n_2) are within 3 days of each other. I count the number of times the test turns out true in our bootstrap procedure, and define the bootstrap probability to be $P = E/B$; where E is the number of positive tests, and B is the number of bootstrap samples. The bootstrap probability converged for $B > 1{,}000$, but I took $B = 10{,}000$ for good measure.