Daily and intraseasonal relationships between lightning and NO$_2$ over the Maritime Continent

(Supplementary material)

Katrina S. Virts$^1$, Joel A. Thornton$^1$, John M. Wallace$^1$, Michael L. Hutchins$^2$,

Robert H. Holzworth$^2$, and Abram R. Jacobson$^2$

$^1$Department of Atmospheric Sciences, University of Washington, Seattle, WA

$^2$Department of Earth and Space Sciences, University of Washington, Seattle, WA
1. **Statistical significance of correlations**

We elected to analyze GOME-2 NO\textsubscript{2} retrievals rather than SCIAMACHY retrievals in hopes that the larger sample size provided by GOME-2’s larger footprint would more than compensate for the larger r.m.s. error of the GOME-2 retrievals. Using 3.5 years of daily GOME-2 observations in 1281 1° × 1° grid boxes extending over the domain of the analysis provides over 10\textsuperscript{6} observations. The effective number of temporal degrees of freedom of the NO\textsubscript{2} observations for a given grid point, calculated using the formula of *Leith* [1973], is \( \geq 650 \).

In order to determine the statistical significance of composite correlation coefficients similar to the composite regression coefficients in Fig. 2, it is also necessary to consider the spatial dependence of the reference grid boxes. Using the eigenvalue method described by *Bretherton et al.* [1999], the effective number of spatial degrees of freedom of the 100 reference grid boxes used to construct Fig. 2 is \( \geq 50 \). Thus, the total effective number of degrees of freedom for composite correlation coefficients between daily NO\textsubscript{2} concentrations and lightning flashes in the 1° × 1° grid boxes that were used to construct the regression maps in Fig. 2 is \( \geq 30000 \). Based on the Student’s \( t \) test, and assuming a two-sided distribution, this implies that composite correlation coefficients of 0.02 are significant at the 99% level. The observed composite correlations are on the order of 0.05 to 0.07. The high level of significance of the correlations is evidenced by the smoothness of the patterns.

2. **Further specifics on the performance of WWLLN**

Daily WWLLN lightning observations summed over the Indonesian domain (10°S to 10°N; 90°E to 150°E) for the period of overlap with GOME-2 are shown in Fig. S1. The number of detected lightning strokes has increased as more stations have been added to the network (e.g.,
the average detection efficiency for all lightning strokes in the United States has increased from around 3% in 2005 to 11% in 2010 [Jacobson et al., 2006; Abarca et al., 2010]. Rodger et al. [2009] compared lightning stroke observations from WWLLN and the New Zealand Lightning Detection Network (NZLDN) from 2007 and found that WWLLN detected 10% or more of all strokes over about 25 kA. Given the reasonably close proximity of New Zealand to our study area, it is reasonable to take 10% as a first-order estimate of WWLLN’s detection efficiency over Indonesia during the period of overlap with GOME-2. Patterns very similar to those shown in Figs. 1, 4, and S5 are obtained when the analysis is performed on lightning data restricted to the period of overlap with GOME-2.

WWLLN’s stroke detection efficiency has been compared in detail to simultaneous observations by the far more sensitive Los Alamos Sferic Array [Smith et al., 2002]. The WWLLN detection efficiency was found in this manner to depend primarily on stroke current amplitude, and not directly on stroke type [Jacobson et al., 2006]. Thus, although not a “total lightning” detection system, WWLLN is biased toward high current alone, and has no overt selection for ground strokes over cloud strokes. To that extent, WWLLN might be expected to serve as a reasonable interim proxy for lightning NO$_x$ production, at least for the period before total-lightning monitors are operational. WWLLN data do not provide any indication of stroke altitude.

The median global very low frequency (VLF) stroke power measured by WWLLN is $3 \times 10^6$ W, more than three orders of magnitude smaller than that indicated by previous measurements, which have shown the power radiated by strokes to be near $10^{10}$ W [Krider and Guo, 1983]. However, this apparent discrepancy is due to the difference in methodology. Previous measurements were of broad band peak power taken in the near field (within 100 km of
the lightning stroke), whereas WWLLN measures the r.m.s. power in the 6-18 kHz band in the far field. With these factors accounted for, the median power is comparable to the previously reported value of $10^{10}$ W [Krider and Guo, 1983].

3. Sensitivity tests on daily lightning/NO\textsubscript{2} relationship and NO\textsubscript{2}/stroke estimate

Figure 2 in Virts et al. [2011] shows the results of a composite lag regression analysis of daily lightning frequency and tropospheric NO\textsubscript{2} column density fields over Indonesia (see text for details). The composite analysis was based on the reference boxes indicated in Fig. S2. To test the robustness of these results, we have conducted a series of sensitivity studies, varying the following aspects of the analysis (note that for each estimate of NO\textsubscript{2} production, a WWLLN detection efficiency of 10\% was assumed).

- **Cloud fraction threshold** The results shown in Virts et al. [2011] were obtained by analyzing only tropospheric NO\textsubscript{2} retrievals classified as “meaningful” by the GOME-2 quality control algorithms [Boersma et al., 2004]. To test the impact of cloudy observations, the composite lag regression analyses were repeated with NO\textsubscript{2} time series that incorporated only observations for which the FRESCO cloud fraction [Koelemeijer et al., 2011] was below 0.1 (Fig. S3, top row).

- **Sample size** Although it is clearly desirable to choose reference boxes for which there is sufficient lightning to provide a day-to-day signal, our selection of the 100 grid boxes with the highest annual-mean lightning frequencies is somewhat arbitrary. Lag regression analyses performed on reference boxes with lightning frequencies in the top 50 and 500 are shown in the second and third rows of Fig. S3.
• **Location of reference boxes** The reference boxes used in the analysis in Fig. 2 are located over land or along the coasts of the Indonesian islands (Fig. S2). To test whether our results are sensitive to possible changes in the NO$_2$ plume from surface sources along the islands, the lag regression analysis was repeated using a set of reference boxes located over water or over less polluted land areas, as indicated in Fig. S4. The associated composite NO$_2$ regression patterns are shown in the bottom row of Fig. S3.

Comparison of Fig. S3 and Fig. 2 in the text shows that the spatial pattern and temporal evolution of the NO$_2$ field are robust with respect to variations in the analysis protocol. The production efficiencies estimated on the basis of these various protocols were used to obtain the estimated range of 1.7 to $2.5 \times 10^{25}$ NO$_2$ molecules per stroke put forth in *Virts et al.* [2011].

4. **MJO correlation maps**

Figure S5 shows maps of correlation coefficients between the MJO index and clouds, lightning, and NO$_2$, analogous to the maps of regression coefficients shown in Fig. 4 in the text. It is evident that the MJO signal in lightning and NO$_2$ over the eastern Indian Ocean is strongest just to the west of Sumatra, within the region in which mesoscale circulations driven by land/sea contrasts and terrain play an important role in triggering convection. In the correlation maps a weak enhancement of lightning is also apparent over new Guinea in phases 1 and 2, which is mirrored in the NO$_2$ regression pattern.

5. **Alternative explanations for NO$_2$ patterns**
In this section we examine three factors that can influence tropospheric NO$_2$ column retrievals and discuss whether they could produce NO$_2$ patterns similar to those shown in *Virts et al.* [2011].

- **Tropopause height variations.** Tropospheric NO$_2$ vertical column densities are obtained from total atmospheric slant column densities by subtracting an assumed stratospheric NO$_2$ column and dividing by a tropospheric air mass factor [Boersma et al., 2004]. The MJO modulates cold-point tropopause height; however, these variations are on the planetary scale [e.g., Zhang, 2005; Virts and Wallace, 2010] and thus cannot explain the more localized NO$_2$ patterns in Figs. 4 and S5.

- **Transport of non-lightning NO$_2$.** NO$_2$ produced by surface sources can be transported by convection to the upper troposphere, where its lifetime is longer and where it is more readily visible to the satellite. Winds can also transport NO$_2$ horizontally, as seen in Fig. 2 in the text. We have demonstrated that the plume of enhanced NO$_2$ associated with a lightning maximum is present regardless of whether we use reference boxes over the ocean or over less polluted land areas (Fig. S4). In addition, the MJO NO$_2$ signature in Figs. 4 and S5 is strongest to the west of Sumatra, where the influence of surface sources of NO$_2$ is much lower (e.g., Fig. 1).

- **Cloud contamination.** The dissimilarities in the cloud and NO$_2$ patterns in Figs. 1 and 4, combined with the fact that the NO$_2$ patterns in Figs. 2 and 4 do not change when only NO$_2$ retrievals with cloud fractions below 0.1 are included in the analysis (see, e.g., Fig. S3), indicate that cloud contamination is not a significant issue for these analyses, though it may need to be taken into account in quantifying the lightning NO$_x$ source.
Thus, while each of these factors influences GOME-2’s tropospheric NO$_2$ retrievals, none can account for the NO$_2$ patterns associated with lightning shown in the text. Other factors that impact tropospheric NO$_2$ retrievals are discussed in Boersma et al. [2004].

**References**


Figure S1. Time series of daily lightning strikes detected by WWLLN over Indonesia.

Figure S2. Map of 100 grid boxes with highest annual-mean lightning frequency. These reference boxes were used to generate the composites in Fig. 2.
**Figure S3.** As in Fig. 2, but regressions were calculated using NO$_2$ time series based on only those observations with FRESCO cloud fractions less than 0.1 (top row), the indicated number of reference boxes (second and third rows), and the reference boxes shown in Fig. S4 (bottom row).
Figure S4. Annual-mean lightning frequency in selected grid boxes over water or less polluted land areas.

Figure S5. As in Fig. 4, but cloud, lightning, and NO\textsubscript{2} are correlated with MJO indices.