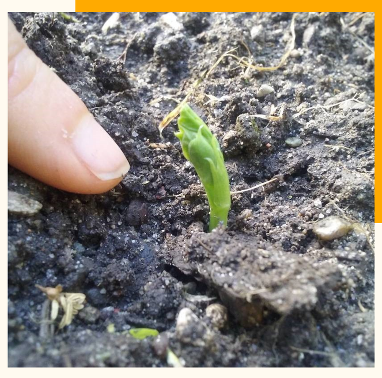
Minding the Gap:

Fertilizer Intensification as a Contributor to Yield Gaps



Background

Yield gaps, the difference between potential and actual yield in agriculture fields, are increasingly discussed as a critical issue in the global food system.

Hypothetical maximum agricultural yields are constrained by solar irradiation, water, atmospheric carbon, temperature, and nutrient availability[1]. The nutrients that most often constrain plant growth are nitrogen (N) and phosphorus (P). It is therefore tempting to draw the conclusion that the more N a farmer adds to soil, the closer she/he will get to obtaining maximum possible yields. However, application of nitrogen fertilizer causes NO_2 gas to form at the soil surface, which in turn catalyzes the formation of ozone[2]. As a strong oxidizing agent, ozone causes chlorosis and necrosis in plants, which reduce photosynthetic activity and therefore yields[3]. Here I examine whether there are areas of cropland in the coterminous US in which excessive nitrogen application has decreased yields, rather than increasing them.

Methodology

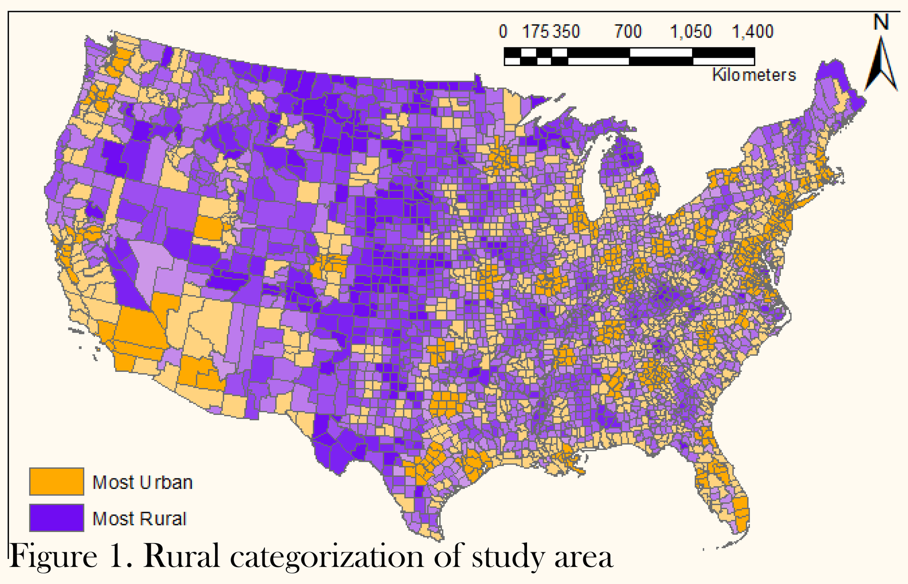


Figure 1. Rural categorization of study area

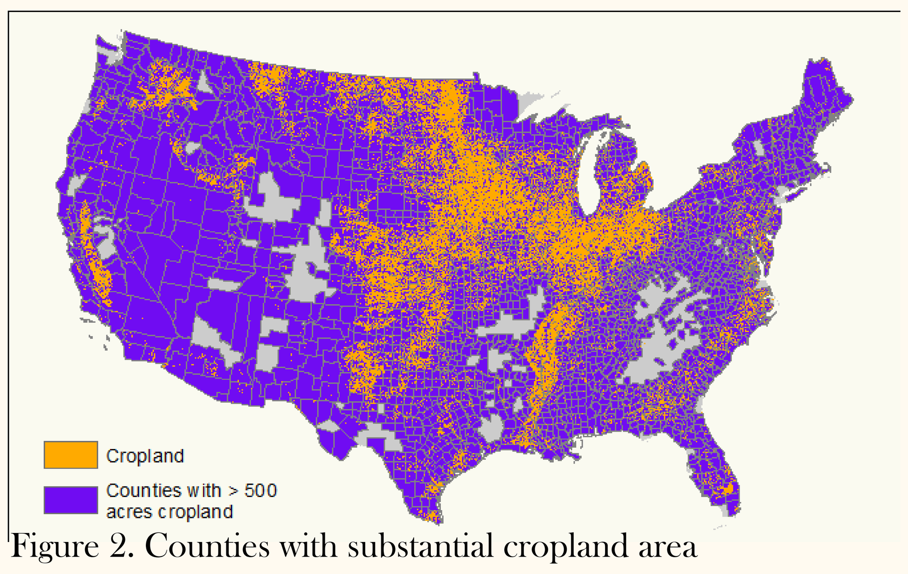


Figure 2. Counties with substantial cropland area

There are several factors that must be considered when modeling crop yields as a function of N. First, urban activities create large amounts of ground-level NO_2 (which can diffuse onto cropland), and may need to be considered separately from rural counties[4] (Figure 1). Second, the smallest granularity for which farmland N application data are available is at the county level[5]. All subsequent variables must therefore also be defined at the county level for analysis. Third, because we are concerned with crop yields, we want to include only cropland[6] in the analysis. Fourth, net primary productivity (NPP; a net measure of how much carbon plants fix per unit area, and a proxy for yields) data are variable throughout the year. I therefore use June 2005 values[7], a month/year that is available in all other datasets of interest[4-9].

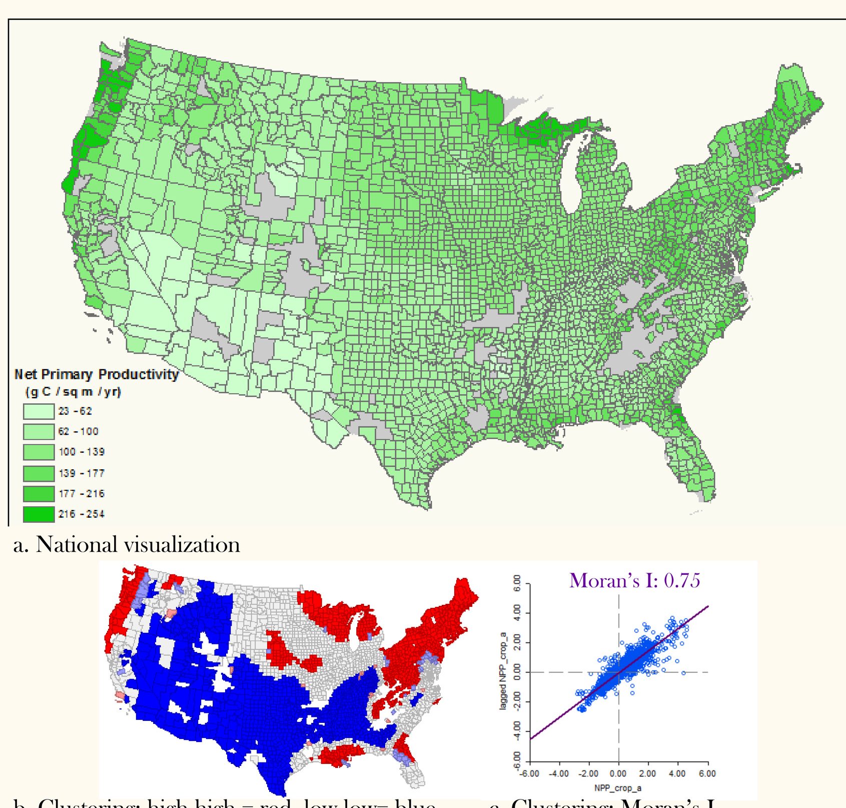
All datasets used in spatial regressions (Figures 3-6) show a variable's average value on cropland in a given county. I determined these averages by clipping the raster for each variable with a cropland mask, and using zonal statistics to find each county's cropland average. In addition, counties are only used if they have significant cropland area (>500 acres; Figure 2).

I then use a spatial error regression to model yield as a function of N application, using a maximum distance threshold weights matrix that gives each county at least one neighbor (147km). In addition to N, NPP is likely a function of atmospheric carbon, solar irradiation, and temperature, among other variables. As atmospheric C will remain roughly constant across the US, it is not included in analysis. I use latitude as a proxy for solar radiation, and average maximum daily temperature[8] as a proxy for temperature. I also allow N-temperature interactions, as temperature could conceivably affect how much N is appropriate to use on cropland (increased N volatility and plant growth impact NO_2 formation and plant uptake). I include an N^2 term as well, to allow the relationship between N and NPP to be curved in nature.

I also model NO_2[9] as a function of N in order to determine whether the two are significantly correlated, thereby lending credence to the hypothesized mechanism. I include only a term for N, rural (binary), and temperature, as this relationship is less likely to be curved or be influenced by interactions.

There is significant spatial clustering in each variable used in the regression (Figures 3-6; Moran's I's range from 0.26-0.92). A classic OLS model also shows significant spatial autocorrelation in its residuals. It is therefore appropriate to use a spatial regression to model the data. Because neighboring counties' NPP should not have a causal influence on a focal county's NPP, spatial error is theoretically preferable to spatial lag. Spatial error is indicative of omitted covariates that are spatially correlated; in this case, water availability, crop type, agricultural practices, etc.

Cropland Averages

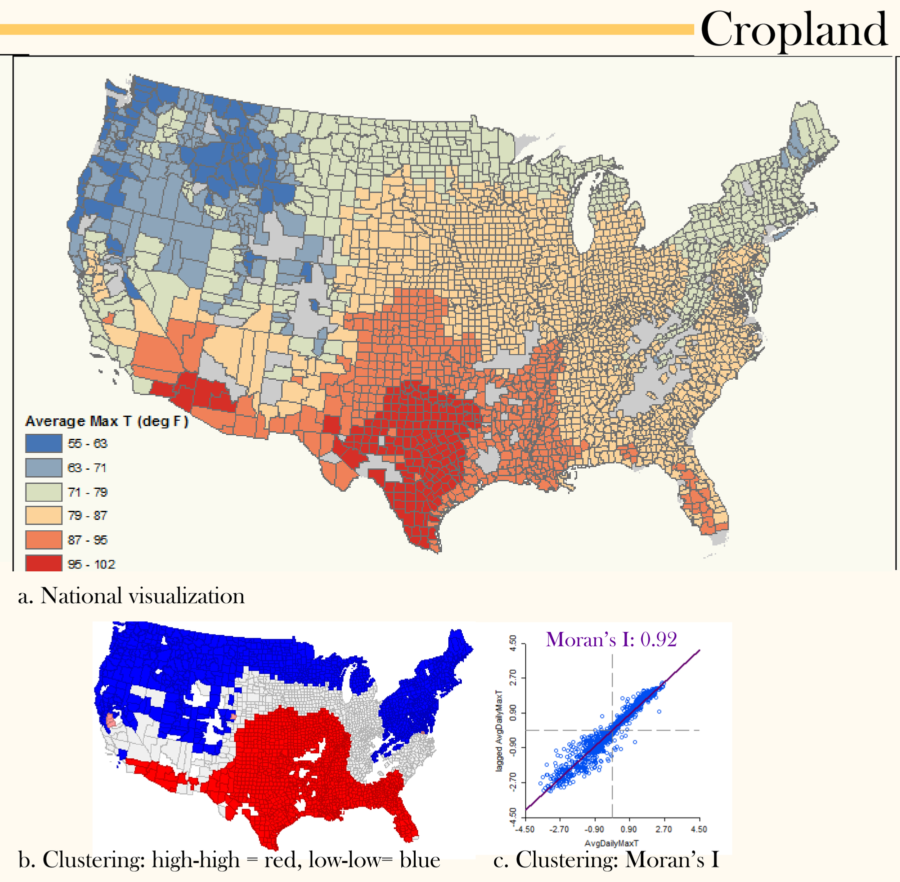


a. National visualization

b. Clustering: high-high = red, low-low= blue

c. Clustering: Moran's I

Figure 3. Net primary productivity

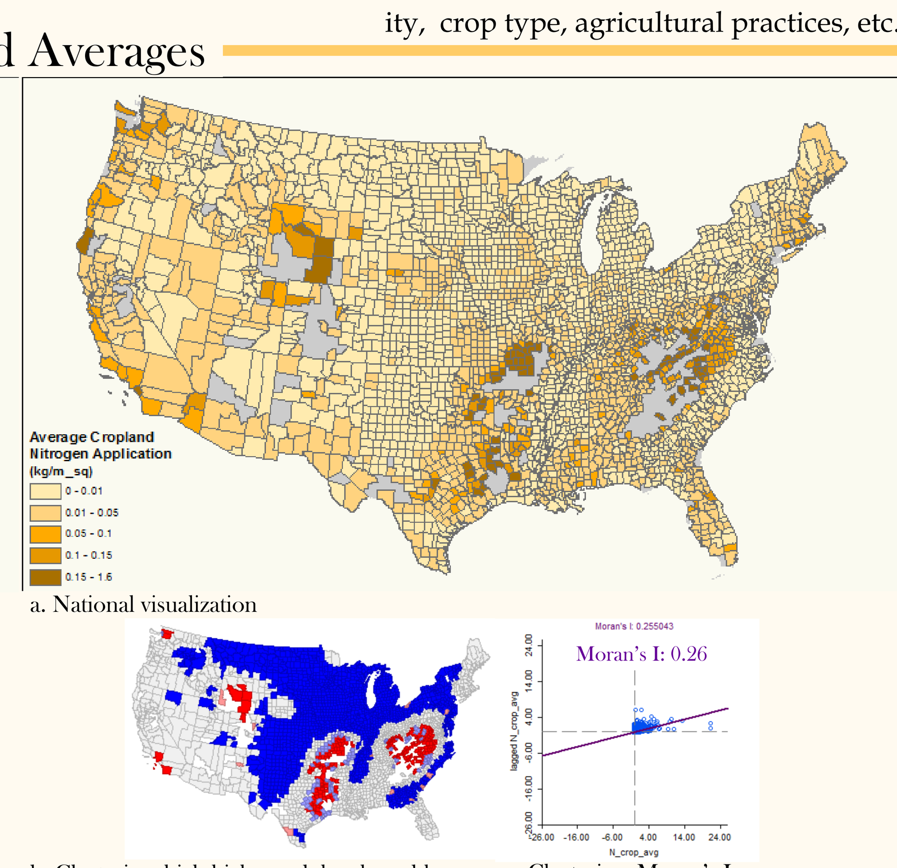


a. National visualization

b. Clustering: high-high = red, low-low= blue

c. Clustering: Moran's I

Figure 4. Average max temp, June '05

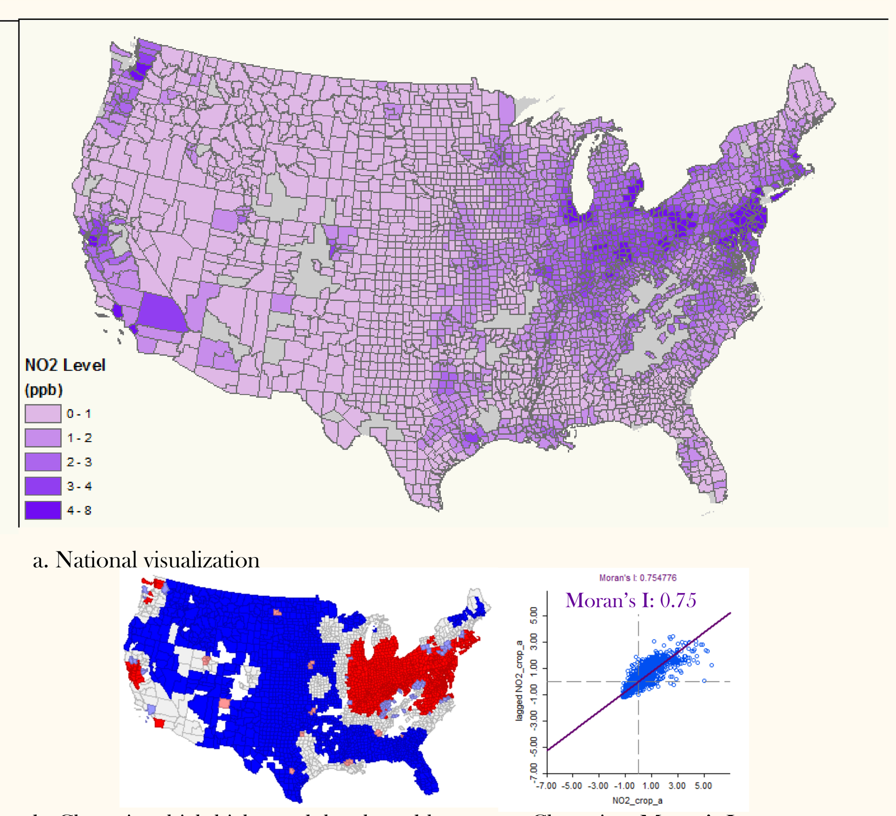


a. National visualization

b. Clustering: high-high = red, low-low= blue

c. Clustering: Moran's I

Figure 5. Farmland nitrogen application



a. National visualization

b. Clustering: high-high = red, low-low= blue

c. Clustering: Moran's I

Figure 6. Ground-level NO_2

Results

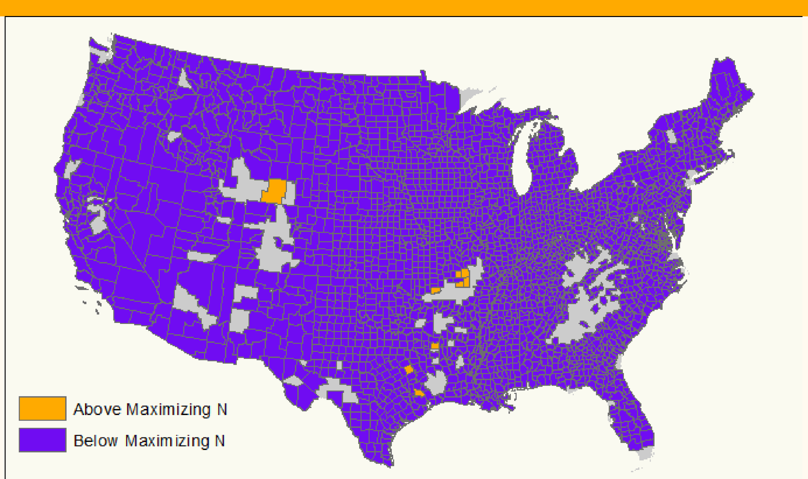


Figure 7. Counties with average overuse of N

Because the classic OLS regression of NPP yields a robust Lagrange Multiplier (LM) for lag that is insignificant (p=0.87), while the LM for error is highly significant (p<0.000), I use spatial error regressions for the remainder of the analysis. A spatial error regression of NPP on: N application, N application^2, rural (binary), latitude, average maximum June '05 temperature, cropland area, and the interaction between N and temperature, gives the results shown in Table 1, Regression A. Leaving non-N variables with insignificant coefficients out of the regression yields the following result:

$$NPP = 268 + 106(N) - 16.0(N^2) - 1.70(T) - 1.04(NT),$$

or Regression B in Table 1.

	Coefficients								R^2
	N application	N^2	Rural	Latitude	Temp.	N x T	Cropland Area	Lambda	
Spatial Error Regression A	105 (0.03)	-15.9 (0.02)	-0.309 (0.57)	4.78e-06 (0.6)	-1.71 (0.000)	-1.03 (0.06)	-2.69e-09 (0.000)	0.987 (0.000)	0.844
Spatial Error Regression B	106 (0.03)	**-16.0 (0.02)**	–	–	-1.70 (0.000)	-1.04 (0.06)	-2.7e-09 (0.000)	0.987 (0.000)	0.844

Table 1. Spatial error regressions of NPP on assorted variables and interactions. P-values are given in parentheses under coefficients.

Of particular note is the significant negative coefficient on N^2, which suggests that the relationship between N and NPP is concave down. This means we can find a value of N that will maximize NPP.

By taking the partial derivative of NPP with respect to N, we can find the value of N that maximizes NPP as follows:

$$(1)\quad \frac{\delta NPP}{\delta N} = 106 - 2*16N - 1.04T$$

$$(2)\quad 0 = 106 - 32N - 1.04T$$

$$(3)\quad N = \frac{106 - 1.04T}{32}$$

Figure 7 shows which counties in the US have an average farmland N application rate that is higher than this predicted maximum.

Finally, we see that county-level N application is a significant predictor of ground-level NO_2, the hypothesized mechanism behind the NPP-N relationship (Table 2).

	Coefficients				R^2
	N application	Rural	Temp.	Lambda	
Spatial Error Regression	0.357 (0.02)	-0.325 (0.000)	0.0432 (0.000)	0.974 (0.000)	0.832

Table 2. Spatial error regressions of NO_2 on N application, whether county is rural, and average max monthly temperature. P-values are given in parentheses under coefficients.

Spatial Questions:

1. After accounting for spatial variability in yields, do US counties with the highest rates of nitrogen application have lower net primary productivity (NPP, a measure of yield)? Is this relationship mediated by ground-level NO_2?

2. After building a spatial error regression that models NPP in terms of N fertilizer use, I can predict the N use in each county that would maximize NPP. Which counties have a N use above this level?

Tufts UNIVERSITY | Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Cartographer:

Yvonne Socolar

AFE Candidate 2018

Projection:

NAD_1983_2011_Contiguous_USA_Albers

Conclusions & Limitations

This analysis shows that there is a value of N that theoretically maximizes yields, with ground-level NO_2 formation as a likely mechanism. Further analysis should be done on a finer scale to determine what the optimum level of N application is for varying agricultural fields, potentially decreasing US N fertilizer use and its associated yield reductions, along with other environmental concerns.

Nevertheless, there are significant limitations to the present results. First, the large cell size of the NPP raster (12km) means that cropland values (which are measured on a 30m scale) may reflect surrounding vegetation, rather than crops, biasing NPP values upwards in counties with little or scattered cropland (assuming young crops will have lower NPP than perennial permanent vegetation). However, including cropland area in the regressions (as shown), should help to mediate this problem.

Second, all values are averaged across counties. Because there is likely significant variation across counties, these large spatial units will bias the analysis towards null results. There may well be smaller areas in which N is overapplied that are not captured here.

Third, each county is given the same weight, even though they vary significantly in cropland area. This means that areas of low cropland prevalence (which have overstated values of NPP) have a disproportionately high influence on results.

Finally, there are several variables omitted from this analysis that likely covary spatially with NPP, such as water availability, crop type, agricultural practices, etc. Though the spatial error regression should help to resolve these omissions, their potential interactions with N are not considered in the present analysis. The maximization therefore gives values of N_{max} that are not sufficiently variable by county.

Works Cited

1. Nonhebel, S. (1996). Effects of temperature rise and increase in CO2 concentration on simulated wheat yields in Europe. *Climatic Change*, *34*(1), 73-90.
2. Fields, Scott. "Global nitrogen: cycling out of control." *Environmental Health Perspectives* 112.10 (2004): A556.
3. Gill, Sarvajeet Singh, and Narendra Tuteja. "Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants." *Plant physiology and biochemistry* 48.12 (2010): 909-930.
4. "A County-Level Measure of Urban Influence" in Rural Development Perspective, Vol. 12 No. 2, February 1997. https://www.ers.usda.gov/data-products/urban-influence-codes.aspx
5. Jo Ann M. Gronberg and Norman E. Spahr. (2012). "County-Level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987-2006." U.S. Geological Survey. https://water.usgs.gov/GIS/metadata/usgswrd/XML/sir2012-5207_county_fertilizer.xml
6. Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011. "Completion of the 2006 National Land Cover Database for the Conterminous United States." *PE&RS*, Vol. 77(9):858-864. https://www.mrlc.gov/nlcd2006.php
7. NASA Earth Observations. (2017). "Net Primary Productivity (1 month - Terra/MODIS)." https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD17A2_M_PSN&year=2013
8. Center for Disease Control and Prevention. (2012). "North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011)." https://wonder.cdc.gov/NASA-NLDAS.html
9. NASA Socioeconomic Data and Applications Center. (2014). "Global 3-Year Running Mean Ground-Level NO2 Grids from GOME, SCIAMACHY and GOME-2, v1 (1996–2012)." http://sedac.ciesin.columbia.edu/data/set/sdei-global-3-year-running-mean-no2-gome-sciamachy-gome2/data-download