Intraregional links between the trends in air pollutants observed at the EANET network sites in 2000-2014

Sergey A. Gromov (1), Alisa Trifonova-Yakovleva (1), and Sergey S. Gromov (1, 2)

(1) Institute of Global Climate and Ecology Roshydromet & RAS, Environmental Pollution Monitoring Division, Moscow, Russian Federation (sergey.gromov@igce.ru)
(2) Max Planck Institute for Chemistry, Mainz, Germany

Overview & Objectives

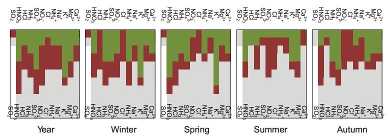
- Air pollution in East Asia
  - Substantial increase since mid-1990s due to rapidly growing anthropogenic emissions (fossil fuel use, energy prod.)
  - A key human health issue in China and North-East Asia
  - Many compounds (SO_x, NO_x, gas-phase & particulate)
  - Large region spans diverse atmospheric transport/chemistry regimes (dry/wet seasons, climatic zones)
- To monitor the state of air pollution in East Asia, we need
  - A wide observational network (EANET)
  - Consistent measurements over substantial time period (EANET)
  - A comprehensive analysis of the data (e.g. trends → this work)

EANET & Data

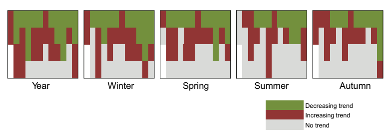
- Acid Deposition Monitoring Network in East Asia (EANET)
  - Active since 1998, operational since end-2000
  - Continuous monitoring of the air quality and precipitation
    - Dry/wet deposition measurements, precipitation chemistry
    - 45 remote and rural station over 13 countries in the region
    - Filter pack sampling / automated monitoring equipment
  - Long-term observations (>15 years) available, a lot of data
  - See more at http://www.eanet.asia
- For our comprehensive trend analysis we pick
  - 15 EANET stations in Japan, Korea, Russia, Mongolia
  - Gas-phase: SO_2, NH_3, HNO_3, HCl
  - Particulate: SO_4^{2-}, NO_3^-, Cl^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}

Recent changes in economic development tendencies and environmental protection policies in the East Asian countries raise hopes for improvement of regional air quality in this vast region populated by more than 3 billion people. To recognize anticipated changes in atmospheric pollutants levels, deposition rates and impact on the environment, the Acid Deposition Monitoring Network in East Asia (EANET, http://www.eanet.asia/) is regularly operating region-wide since 2000 in 13 countries. The network provides continuous monitoring data on the air quality and precipitation (including gas-phase and particulate chemistry) at 55 monitoring sites, including 20 remote and 14 rural sites. Observation of soil and inland water environments are performed at more than 30 monitoring sites [1].

- Fraction of remote stations showing statistically significant trends
  - 11 remote stations



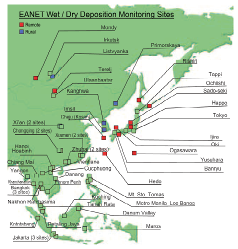
- Fraction of rural stations showing statistically significant trends
  - 4 rural stations



Distribution of the EANET stations considered in this study

Red (11) and blue (4) symbols denote the remote and rural EANET stations, respectively, which data we have considered in this study

Further network info:
http://www.eanet.asia/profile/index.html



In this study we focus on 1) the data quality assessment and preparation and 2) analysis of temporal trends of compositions observed at selected 26 non-urban EANET stations. Speciation includes gas-phase (SO_2, HNO_3, HCl, NH_3) and particulate matter (SO_4^{2-}, NO_3^-, Cl^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}) abundances analysed in samples collected using filterpack technique with sampling duration/frequency of one-two weeks. Data quality assessment (distribution test and manual inspection) allowed us to remove/repair random and operator errors. Wrong sample timing was found for 0.37% (severe) and 34% (mild inconsistency) of the total of 7630 samples regarded. Erroneous data flagging (*e.g.* missing or below the detection limit) was repaired for 9.3%, respectively. Some 1.8% of severely affected data were corrected (where possible) or removed. Thus refined 15-year dataset is made available for the scientific community. For convenience, we also provide data in netCDF format (per station or in an assembly).

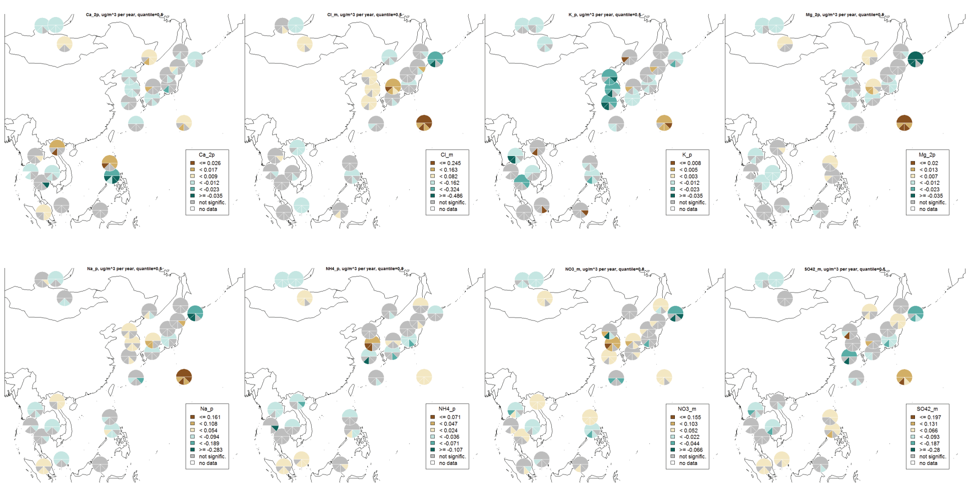
Comprehensive trend analysis?

- Substantial data analysis & preparation
  - "Brushing" to filter out processing / operator errors
  - Various methods to discern special events (outliers)
    - Distribution and monotonicity tests
    - Running average filtering using 3-σ & 4-σ rule, *etc.*
  - Correct treatment of the data below the detection limit
  - Longer sampling periods (integration)
- Consistent and robust statistical assessment
  - Use a suitable statistical apparatus (R software)
  - Estimate an ensemble of trends / bootstrap
    - All data → by season → by month
    - Trends on quantiles (highlight trends in emissions/sinks?)
  - Sensitivity tests
    - Quantile regression (sensitivity to extrema)
    - Breaking trends (future analyses)

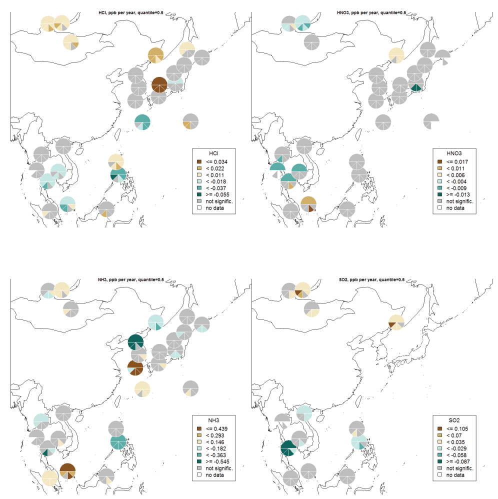
Results & Outlook

- Re-analysed, improved ("brushed") data
  - A data processing framework for other EANET data
  - Quality control for data users (*e.g.* modellers)
- Compendium of trends
  - "Carpets" (ensembles of trends)
  - Statistics (~50% of trends we estimate are significant)
  - Seasonal & spatial distributions
  - Special cases (*e.g.* significant seasonal trends "sum up" to a non-significant overall trend)
- Future analyses / intentions
  - Work on automated monitoring data (O_3, …)
  - Trend correlation
  - More sophisticated trend models (breaking, *etc.*)
  - Cluster analysis (stations with similar trends various sets of compounds)

Long-term (2000-2012) trends in PM particulate compounds ($ug/m^3/yr$) observed at selected EANET stations in 2000-2012

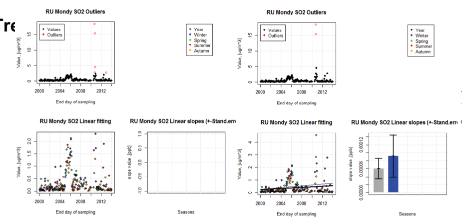


Long-term (2000-2012) trends in gas-phase compounds (ppb/yr) observed at selected EANET stations in 2000-2012

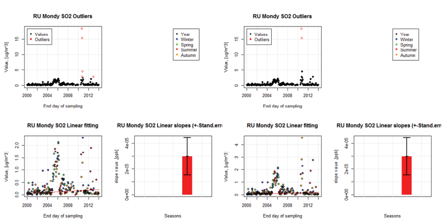


Figures on the left present the spatiotemporal distribution of the estimated trends. Circles denote particular station. The upper and four lower parts of each circle indicate the total and seasonal (DJF, MAM, JJA, SON, respectively) trends. Grey colour denote none/no significant trend estimated. No (sufficient) data for trends is denoted with blanks.

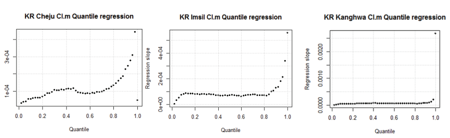
- Least squares regression
  - Small difference in number of outliers may result in large differences in resulting trends!



- Quantile regression
  - Robust, *i.e.* less sensitive to outliers



- Differences in quantile trends
  - Exhibit sensitivity of data to extrema / outliers
  - Highlight trends in the baseline data



- Exhibit sensitivity of data to extrema / outliers
- Highlight trends in the baseline data

3. REGRESSION QUANTILES

Our point of departure is an elementary definition of the sample quantiles which, by circumventing the usual reliance on an ordered set of sample observations, is readily extended to the linear model. As above, let $\{y_t: t = 1, \ldots, T\}$ be a random sample on a random variable Y having distribution function F. Then the θth sample quantile, $0 < \theta < 1$, may be defined as any solution to the minimization problem:

$$\min_{b \in \mathbb{R}} \left[\sum_{t \in \{t: y_t \geq b\}} \theta |y_t - b| + \sum_{t \in \{t: y_t < b\}} (1-\theta)|y_t - b| \right].$$

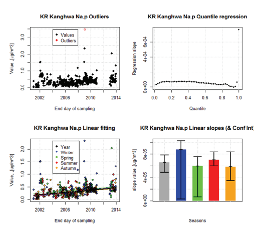
The case of the median ($\theta = 1/2$) is, of course, well known, but the general result has languished in the status of curiosum—appearing for example as an exercise in Ferguson [13].

Huber's [26] observation that outliers are difficult to identify in the regression context underlines the essential ambiguity involved in extending to the linear model the ordinary notions of sample quantiles based on an ordering of sample observations. A direct generalization of the minimization problem posed above resolves this ambiguity. Letting $\{x_t: t = 1, \ldots, T\}$ denote a sequence of (row) K-vectors of a known design matrix, suppose $\{y_t: t = 1, \ldots, T\}$ is a random sample on the regression process $u_t = y_t - x_t\beta$ having distribution function F. The *θth regression quantile*, $0 < \theta < 1$, is defined as any solution to the minimization problem:

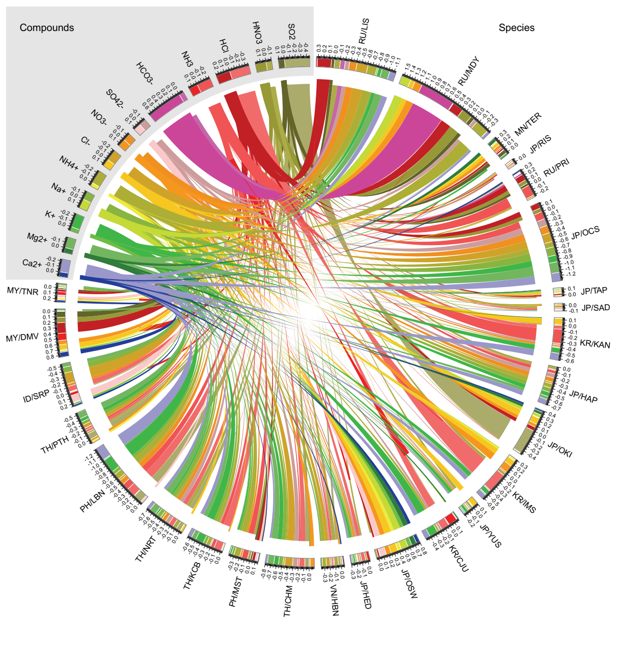
$$\min_{b \in \mathbb{R}^K} \left[\sum_{t \in \{t: y_t \geq x_t b\}} \theta |y_t - x_t b| + \sum_{t \in \{t: y_t < x_t b\}} (1-\theta)|y_t - x_t b| \right].$$

In the location model ($K = 1$, $x_t = 1$, for all t) the two minimization problems coincide. The least absolute error estimator is the regression median, i.e., the regression quantile for $\theta = 1/2$.

[Koenker & Bassett, 1978]



Overview of long-term trends in compounds observed at selected EANET stations over 2000-2012



We use CIRCOS software (www.circos.com) to give an overview of the entire trend dataset obtained.

Each ribbon connection in the circos diagram on the left denotes a statistically significant trend detected for a give compound (colour-coded, less saturated colours denote negative trends / top-left quadrant), observed at the given EANET station (e.g. RU/LIS for Listvyanka, note values are stacked / in right-bottom quadrants).

Our calculations indicate that about half of the median trends at EANET stations are significant, derived either for the entire observational period or for a given season, however not for the same species. The proportions of decreasing and increasing trends are comparable. The latter is the case for SO_2, HCl, Cl^-, NO_3 (except for Russia), while marked decrease in K^+ abundances is prevailing at all stations. Most unsystematic trends are seen for nitrogenated compounds, particularly HNO_3, which calls for deeper data quality analysis.

Links / Contact

Alisa Trifonova-Yakovleva
yakovleva.eanet@gmail.com

Sergey A. Gromov
sergey.gromov@igce.ru

Institute of Global Climate and Ecology (Roshydromet & RAS), Moscow (IGCE)
http://www.igce.ru

Acid Deposition Monitoring Network in East Asia (EANET)
http://www.eanet.asia

Asia Center for Air Pollution Research (ACAP)
http://www.acap.asia/index.html