S1. Evaluation metrics

The evaluation is made using the metrics defined through the following equations,

$$mean = \bar{x} = \sum_{i=1}^{N} \frac{x^i}{N} \quad (1)$$

$$\sigma = \sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} |x - \bar{x}|^2} \quad (2)$$

$$\%bias = 100 * \frac{\bar{x}_{mod} - \bar{x}_{obs}}{\bar{x}_{obs}} \quad (3)$$

$$r = r(x, y) = r(y, x) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\overline{x_\iota - \bar{x}}}{\sigma_x}\right) \left(\frac{\overline{y_\iota - \bar{y}}}{\sigma_y}\right) \quad (4)$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(x_{mod}^i - x_{obs}^i)^2}{N}} \quad (5)$$

Where x^i is the observed (x_{obs}) or modelled (x_{mod}) values, σ is the standard deviation of obsevations (σ_{obs}) and model values (σ_{mod}), %bias is the model mean bias normalized by the observed mean, r is the Pearson correlation coefficient between observed and modelled values, and the RMSE is the root mean square error of the modelled values compared to the observed.

Table 2 in the main paper includes spatial mean of hourly statistics and spatial statistics of annual means:

- The spatial mean of hourly statistics is calculated as follows: The above metrics (eqns. (1)-(5)) are calculated for hourly near-surface O_3 concentrations at each of the measurement sites. These statistics are then averaged spatially over the measurement sites (as in eqn. (1)). This evaluates the temporal performance.
- The spatial statistics of hourly means are calculated as follows: The annual mean (2013) is calculated from the observed respective modelled hourly near-surface O_3 at each of the observation sites. These observed and modelled annual mean pairs are then used in the calculation of the above metrics (eqns. (1)-(5)). This evaluates the spatial performance.

S2. Evaluation of annual means

The 2dvar analysis significantly improves the correlation coefficient and RMSE at the observation sites of modelled annual mean near-surface O_3 as compared to the MFG simulation (Table S1): The average correlation coefficient, 0.46, in the MFG, is improved to 0.87 in the LONGTERM reanalysis, and reaches 0.99 in the ALL reanalysis. The RMSE is also improved in the ALL and LONGTERM reanalyzes. This is expected, since the ALL reanalysis is dependent on all the observations included in the evaluation, LONGTERM is dependent on part of the observations, whereas the MFG simulation is independent of the observations. It is striking that the mean bias is very low for all simulations, including the observation independent MFG. The MFG simulation underestimates the inter-annual variation, whereas the variations in the reanalyzes are similar to the variations in the measurements. The spatial statistics of the 2dvar analysis are similar to or better than the MFG simulation. The correlation coefficient of the multi-year means is poor for the MFG simulation and the spatial variation is underestimated, but both are strongly improved in the LONGTERM and ALL reanalyzes.

S3. Time series comparison of ALL and LONGTERM

To understand how the number of measurement sites included in the two assimilated data sets affects the time series for a larger spatial area, we compare the trends in annual mean and annual max (Fig. S2-S3) obtained with the two simulations. The annual values are averaged for three regions (North, Central and South, as illustrated in the main paper Fig. 3). The time series of ALL and LONGTERM diverge, especially in the later part of the period, which is due to an introduction of more measurement sites in the later part of the ALL simulation. Several of these sites experience strong night-time temperature inversions, which in turn result in very low night-time O_3 concentrations. For this reason the annual max does not diverge as much as the annual means. Thus the estimated trend differs for the two simulations, with the largest difference for annual mean in southern and central Sweden. To eliminate such impacts on the trend statistics, we will therefore focus on the LONGTERM simulation in the assessments of these metrics. In Fig. S3 we also include a comparison of the annual mean time series for observations and the MFG, LONGTERM and ALL simulations at each of the measurement sites. The trend figures illustrate the evaluation scores: good performance by MFG at many sites and improvements due to the variational analysis, with best performance compared to the observations by the ALL simulation. Further, it is clear that the time series of LONGTERM and ALL diverge at the measurement sites with fewer years of data, further strengthening our conclusion about using the LONGTERM data set for trend and extreme estimation.

S4. Figures and tables

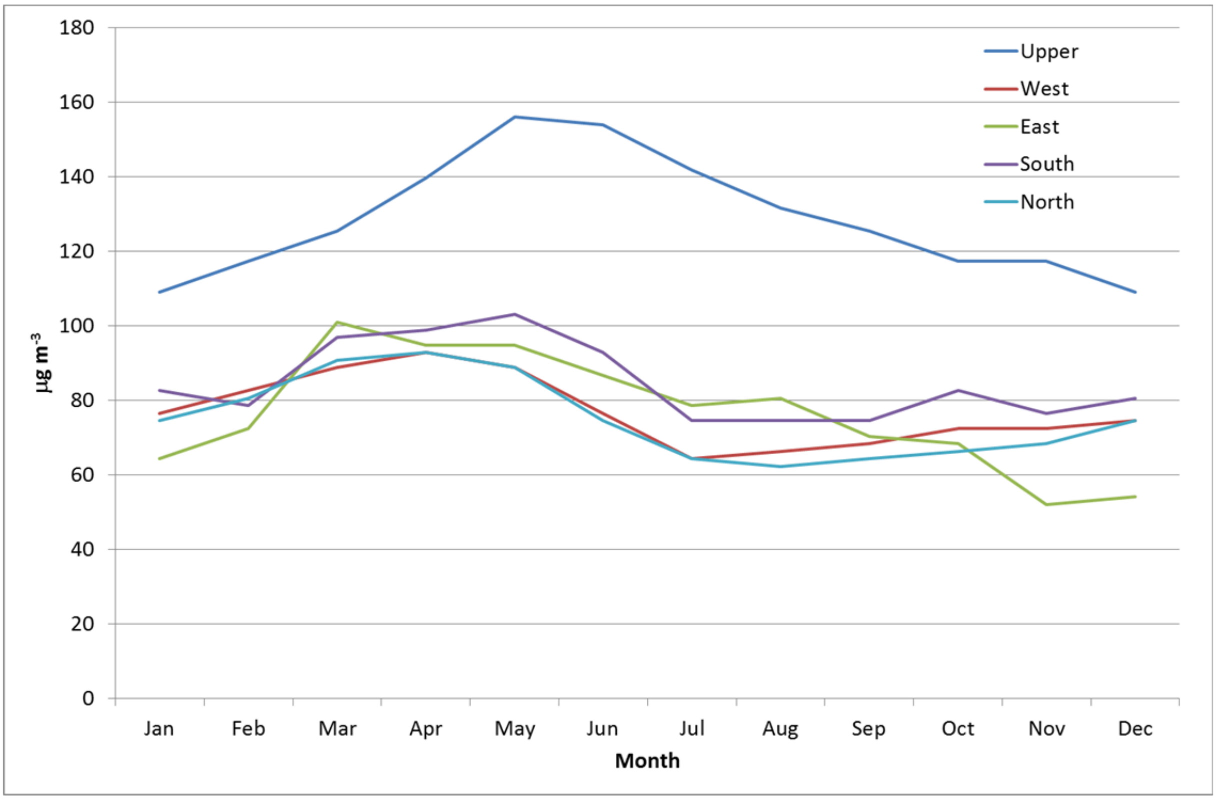

Figure S1. Seasonal variation in lateral and upper boundary conditions for ozone in the year 2011. The upper boundary is at approximately 5 km height. Unit: µg m^{-3}.

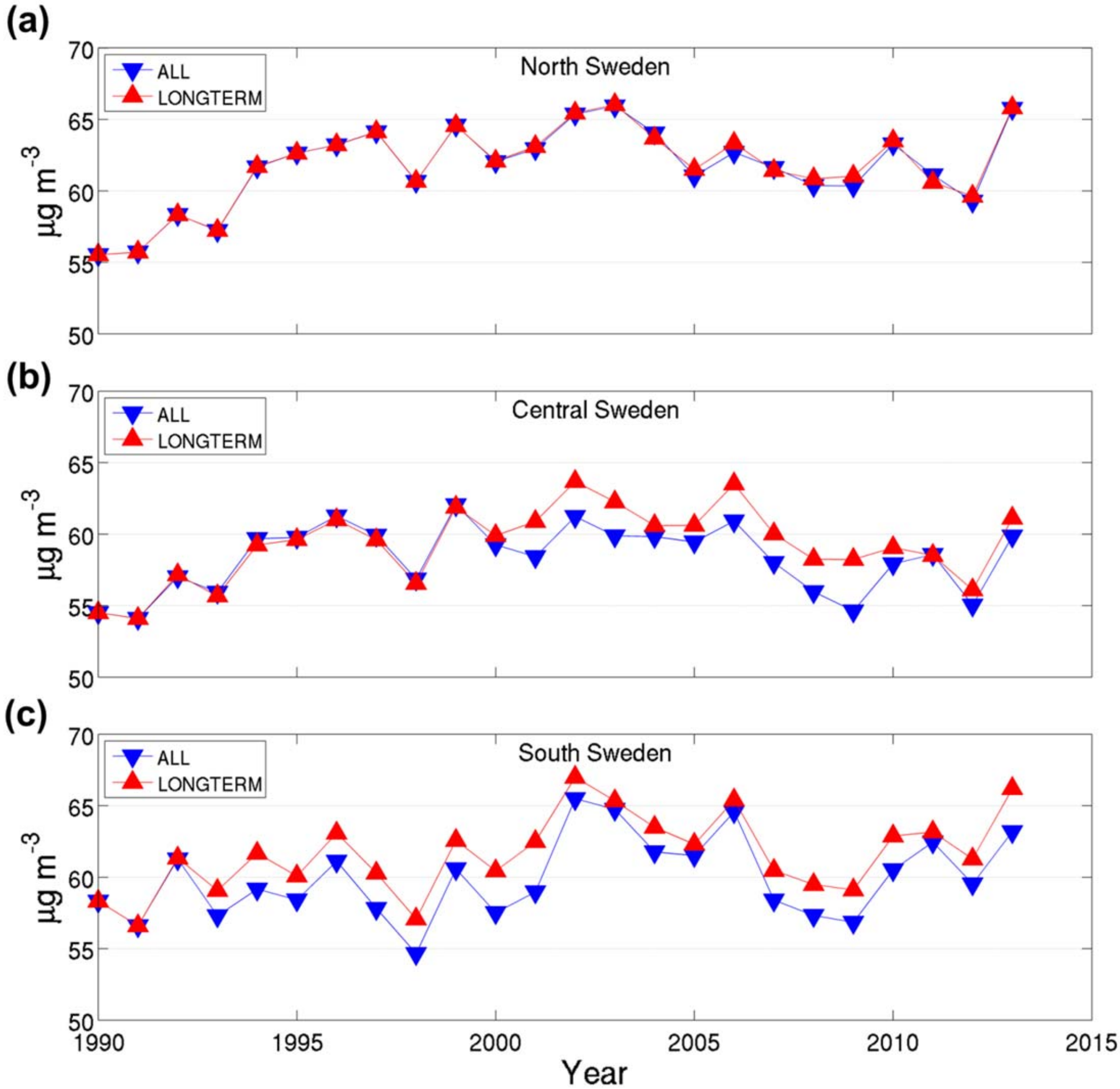

Figure S2. Time series of annual mean near-surface ozone concentrations averaged over three regions (North (a), Central (b) and South (c) Sweden, cf. main paper Fig. 3), for the two reanalyzes ALL (blue) and LONGTERM (red).

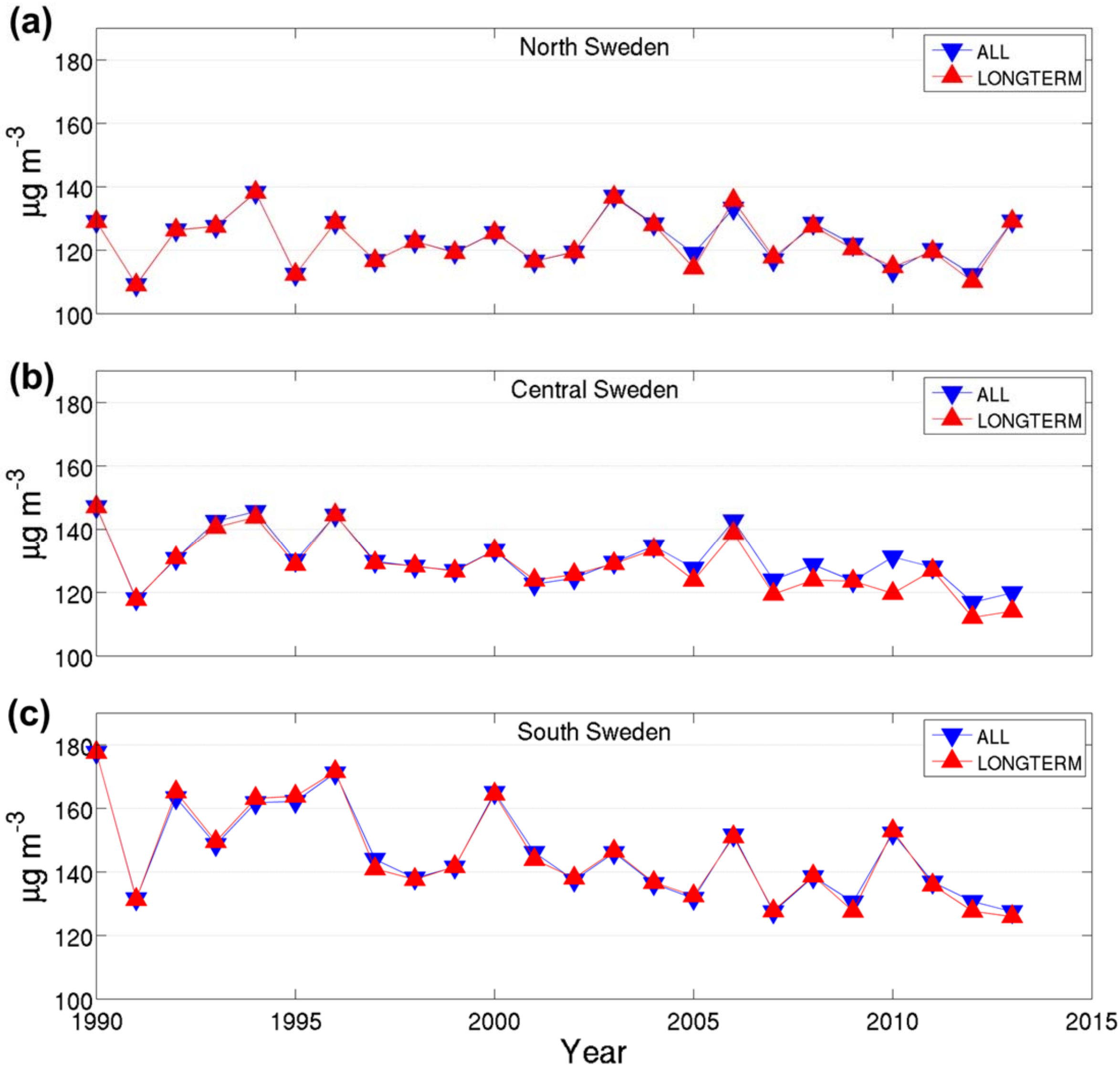

Figure S3. Time series of annual maximum of 1h mean near-surface ozone concentrations averaged over three regions (North (a), Central (b) and South (c) Sweden, cf. main paper Fig. 3), for the two reanalyzes ALL (blue) and LONGTERM (red).

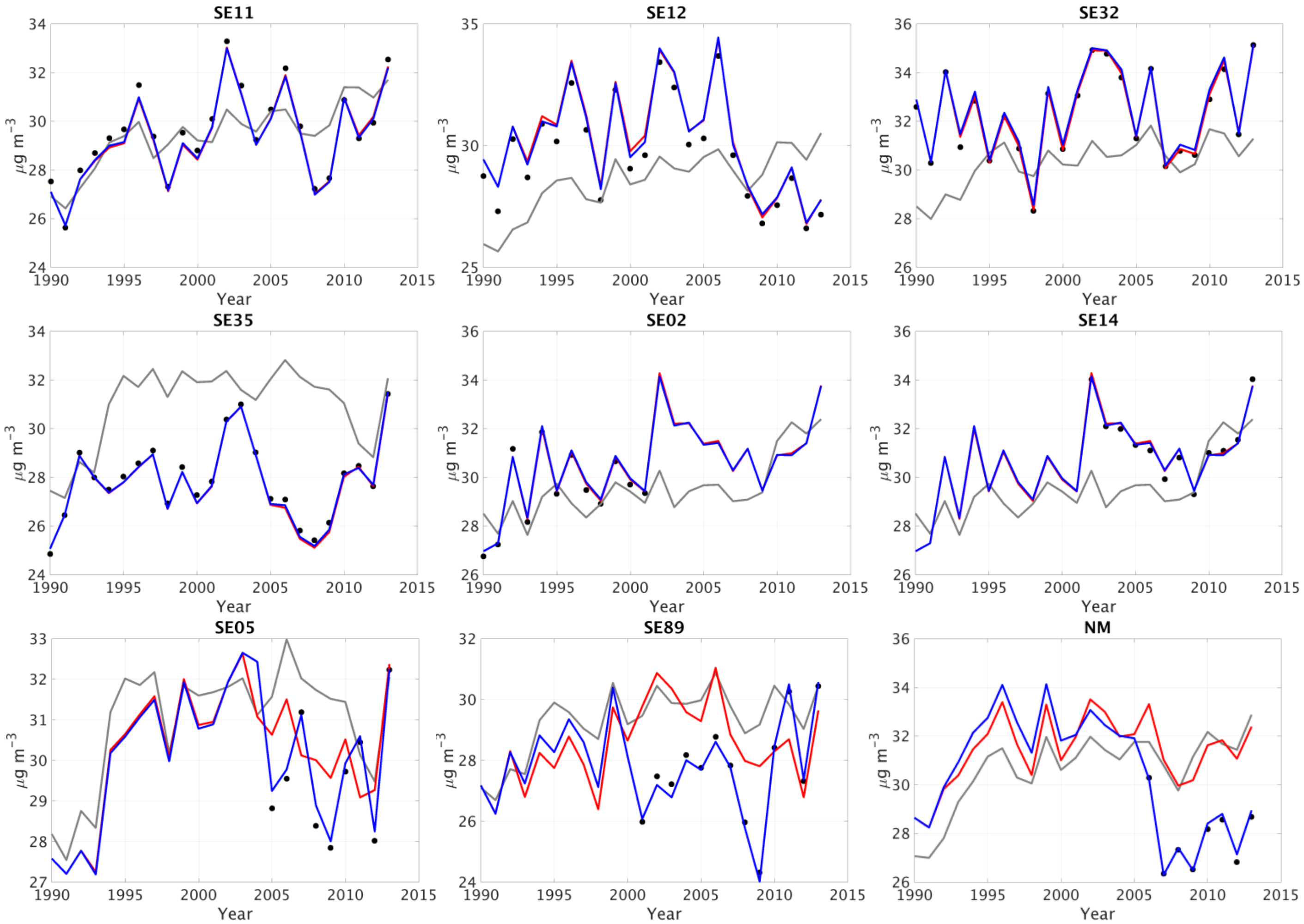

Figure S4. Time series of annual mean near-surface ozone at Swedish measurement sites with more than 1 year of measurement data. Observations (black circle), the “first guess” simulation MFG (grey line), the two reanalyzes LONGTERM (red) and ALL (blue).

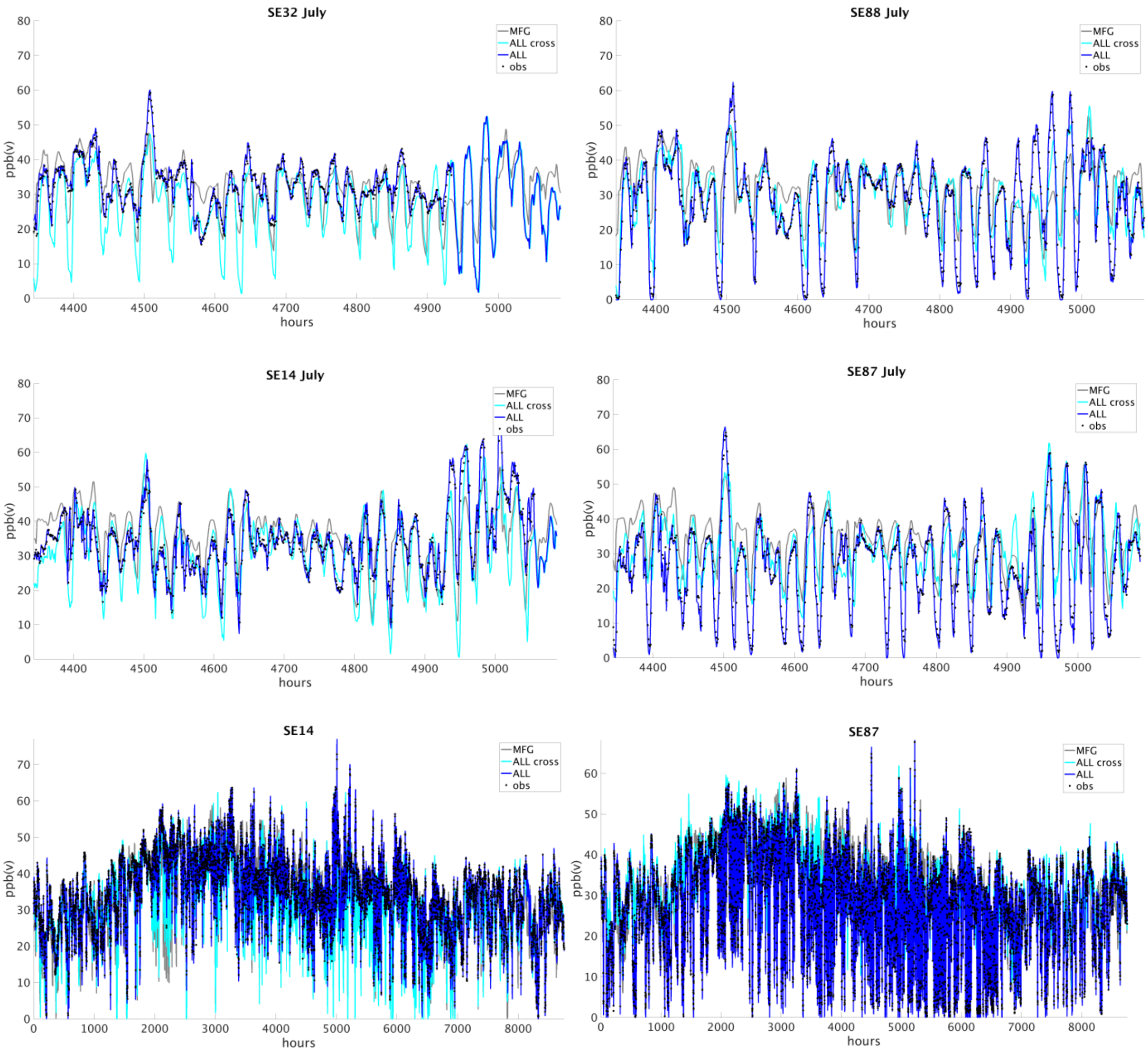

Figure S5. Hourly mean near-surface O_3 concentrations at selected sites including the MFG simulation (grey), ALL cross validation simulations (fair blue), the ALL reanalysis (dark blue) and observations (black circles). The bottom two panels show the full year 2013, the other zoom in on the month July 2013. Norra Kvill (SE32) and Råö (SE14) are less prone to be impacted by night-time inversions than Östad (SE87) and Asa (SE88).

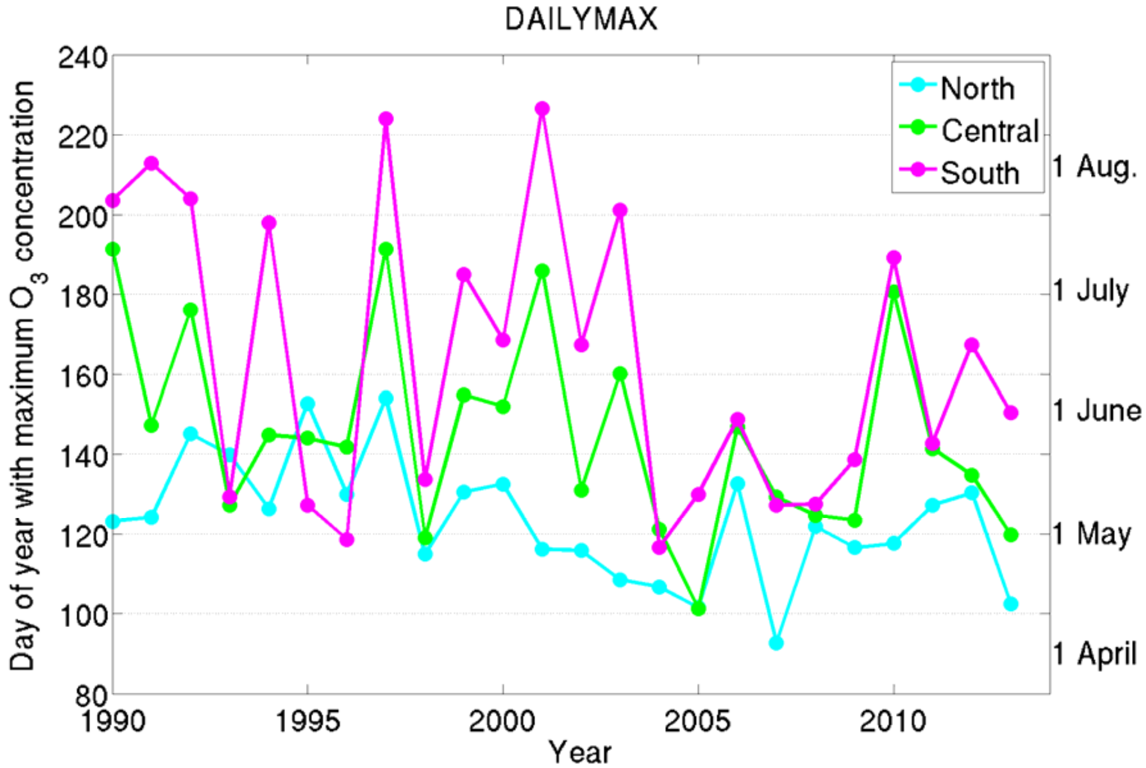

Figure S6. Day of the year (1=1 Jan, 32=1 Feb. etc.) when the daily maximum 1h mean near-surface ozone reaches its annual maxima. The values displayed are spatial averages over the 3 Swedish regions: North, Central and South. Results from the LONGTERM reanalysis.

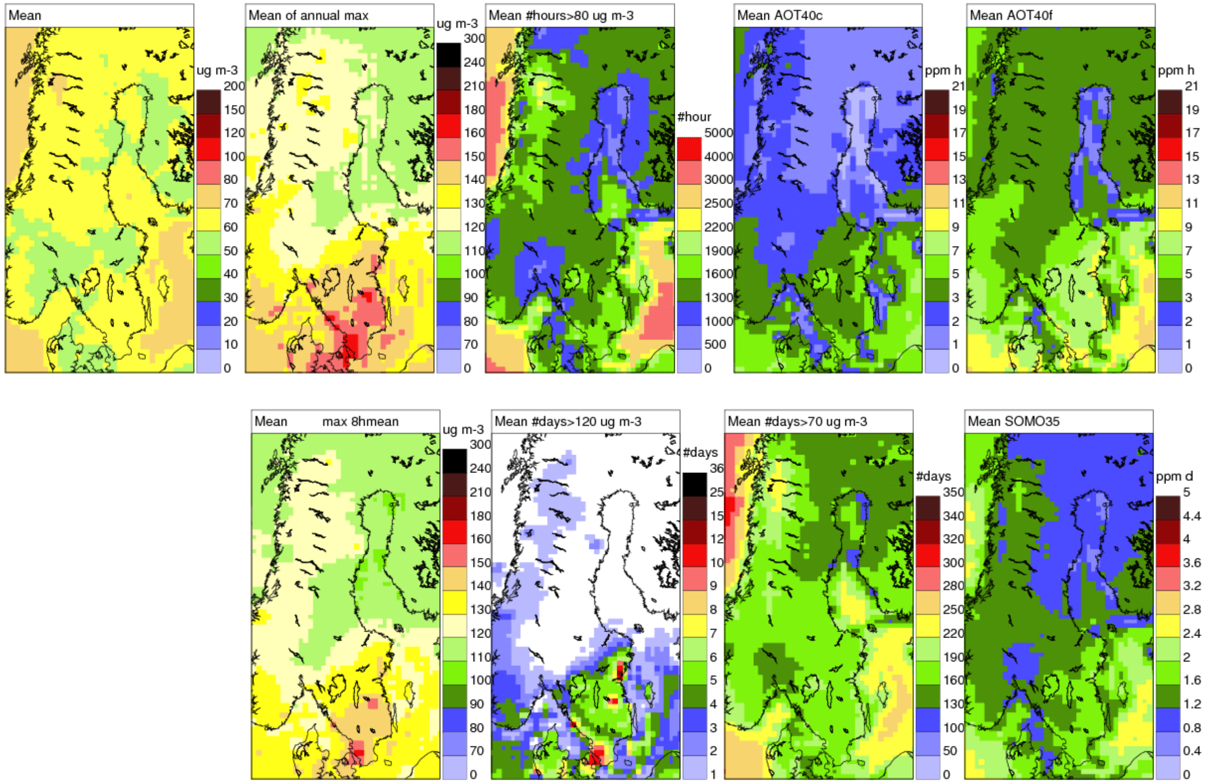

Figure S7. Period mean near-surface ozone during 1990-2013 from top left: Annual mean, annual max of 1hour mean (Max 1H), number of hours exceeding 80 µg m^{-3}, AOT40 in crop growing season (AOT40c; May-July), AOT40 in forest growing season (AOT40f; April-September), annual max of running 8hour mean, number of days with daily max of running 8hour mean exceeding 120 µg m^{-3} and 70 µg m^{-3}, and the health indicator SOMO35. Results from the LONGTERM reanalysis.

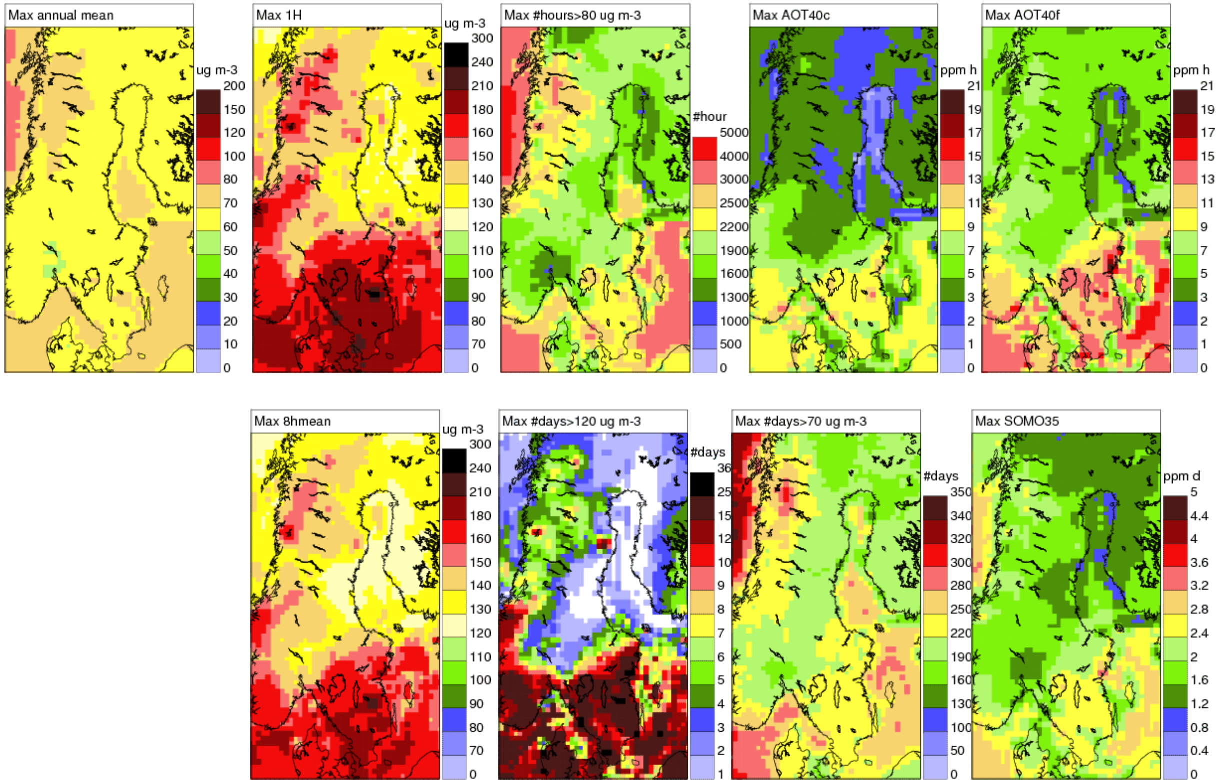

Figure S8. Period max value in near-surface ozone during 1990-2013 from top left: Annual mean, annual max of 1hour mean (Max 1H), number of hours exceeding 80 μg m^{-3}, AOT40 in crop growing season (AOT40c; May-July), AOT40 in forest growing season (AOT40f; April-September), annual max of running 8hour mean, number of days with daily max of running 8hour mean exceeding 120 μg m^{-3} and 70 μg m^{-3}, and the health indicator SOMO35. Results from the LONGTERM reanalysis.

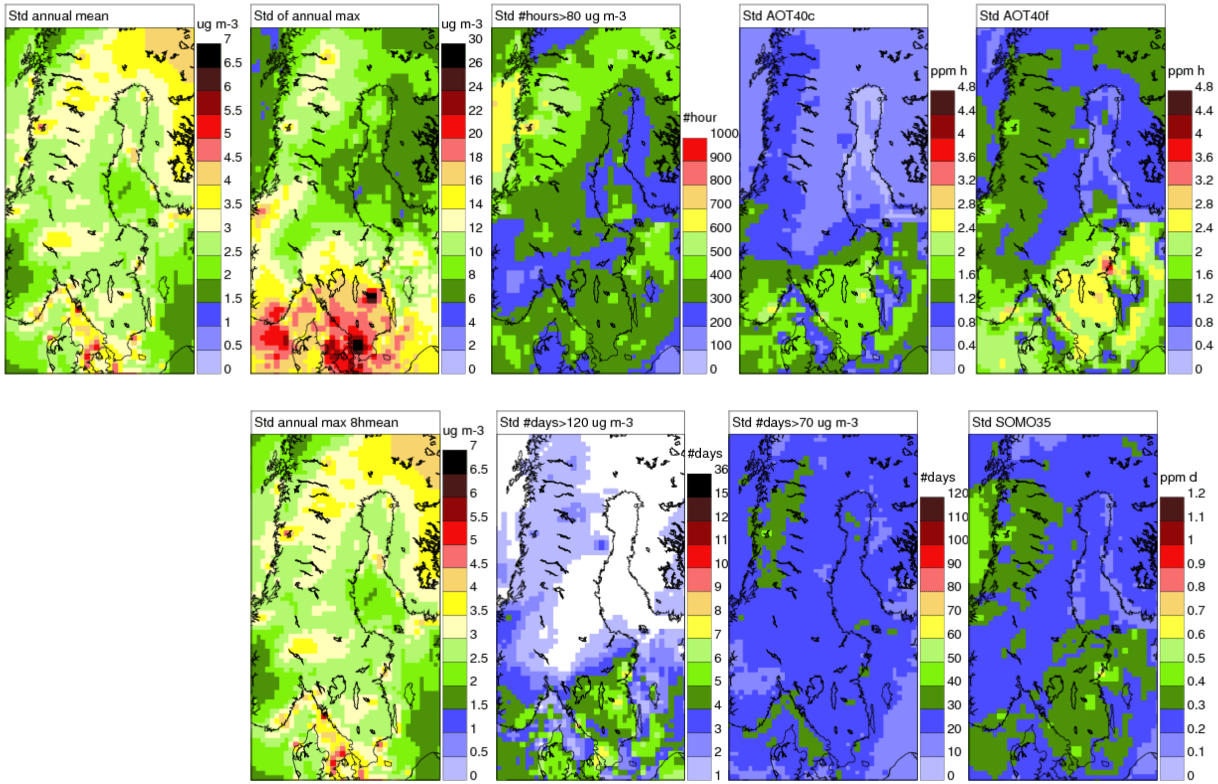

Figure S9. Period standard deviation in near-surface ozone during 1990-2013 from top left: Annual mean, annual max of 1hour mean (Max 1H), number of hours exceeding 80 μg m^{-3}, AOT40 in crop growing season (AOT40c; May-July), AOT40 in forest growing season (AOT40f; April-September), annual max of running 8hour mean, number of days with daily max of running 8hour mean exceeding 120 μg m^{-3} and 70 μg m^{-3}, and the health indicator SOMO35. Results from the LONGTERM reanalysis.

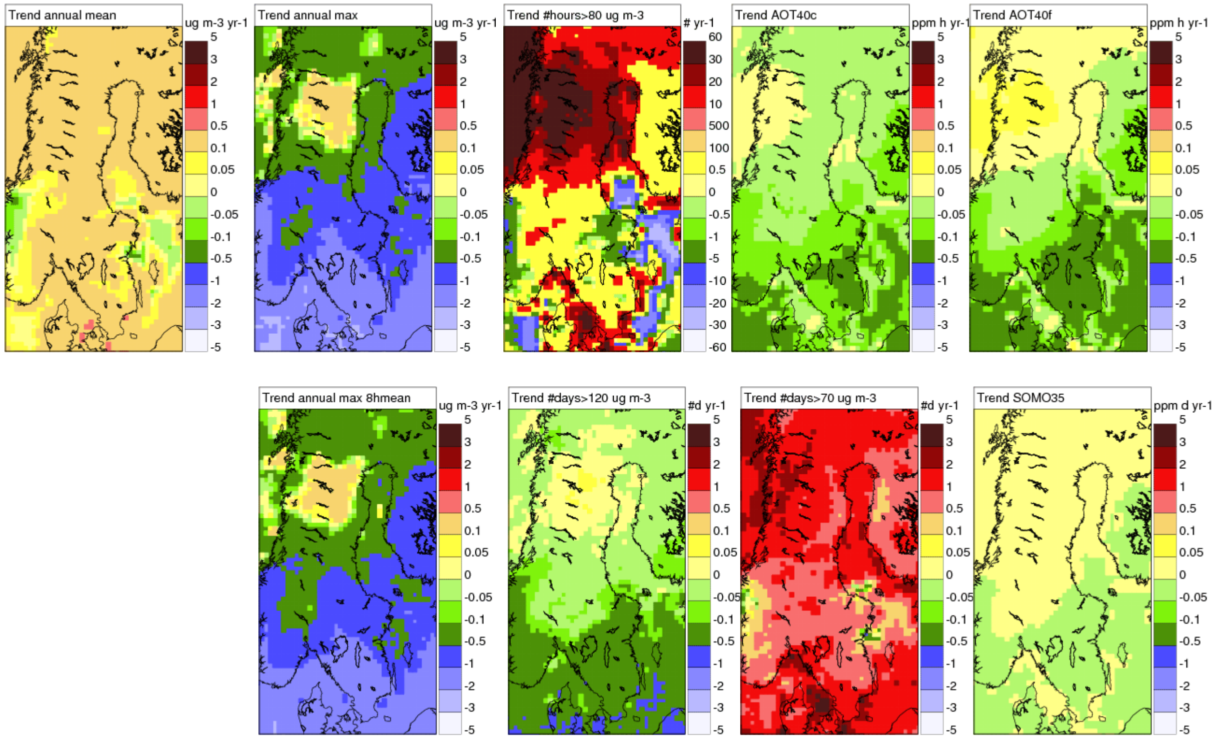

Figure S10. Period linear trend in near-surface ozone during 1990-2013 from top left: Annual mean, annual max of 1hour mean (Max 1H), number of hours exceeding 80 µg m^{-3}, AOT40 in crop growing season (AOT40c; May-July), AOT40 in forest growing season (AOT40f; April-September), annual max of running 8hour mean, number of days with daily max of running 8hour mean exceeding 120 µg m^{-3} and 70 µg m^{-3}, and the health indicator SOMO35. Results from the LONGTERM reanalysis.

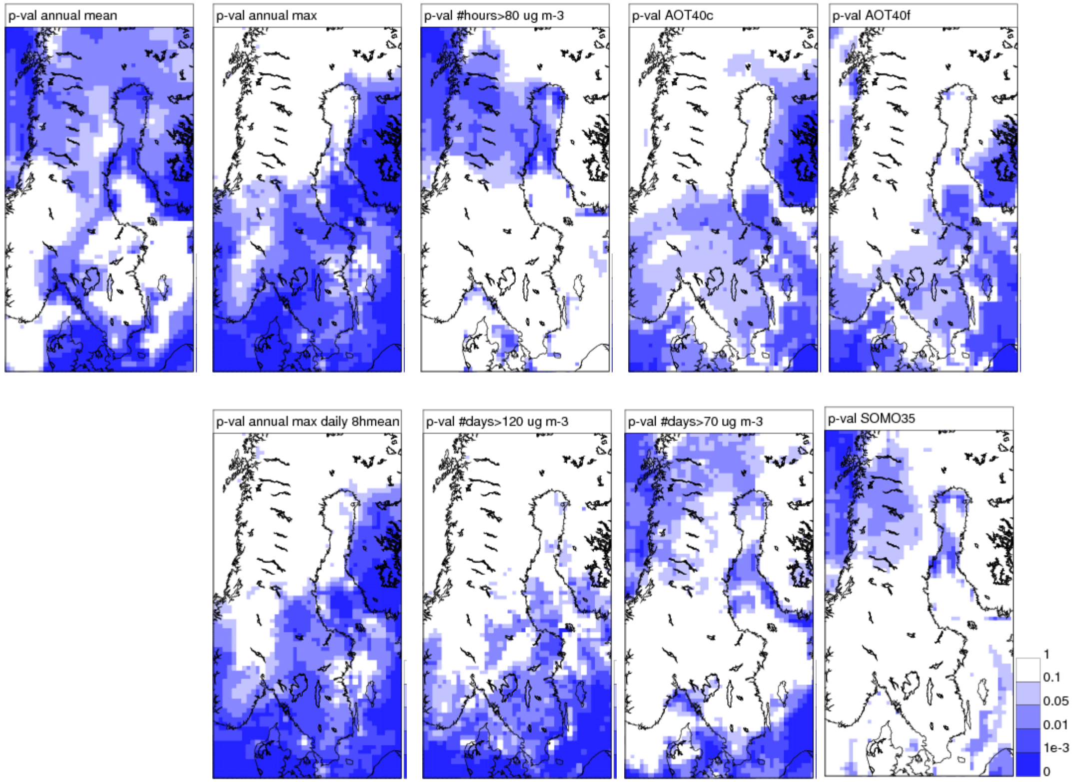

Figure S11. p-value in near-surface ozone linear trend over the period 1990-2013 from top left: Annual mean, annual max of 1hour mean (Max 1H), number of hours exceeding 80 µg m^{-3}, AOT40 in crop growing season (AOT40c; May-July), AOT40 in forest growing season (AOT40f; April-September), annual max of running 8hour mean, number of days with daily max of running 8hour mean exceeding 120 µg m^{-3} and 70 µg m^{-3}, and the health indicator SOMO35. p-values above 0.1 are non-significant (white). Results from the LONGTERM reanalysis.

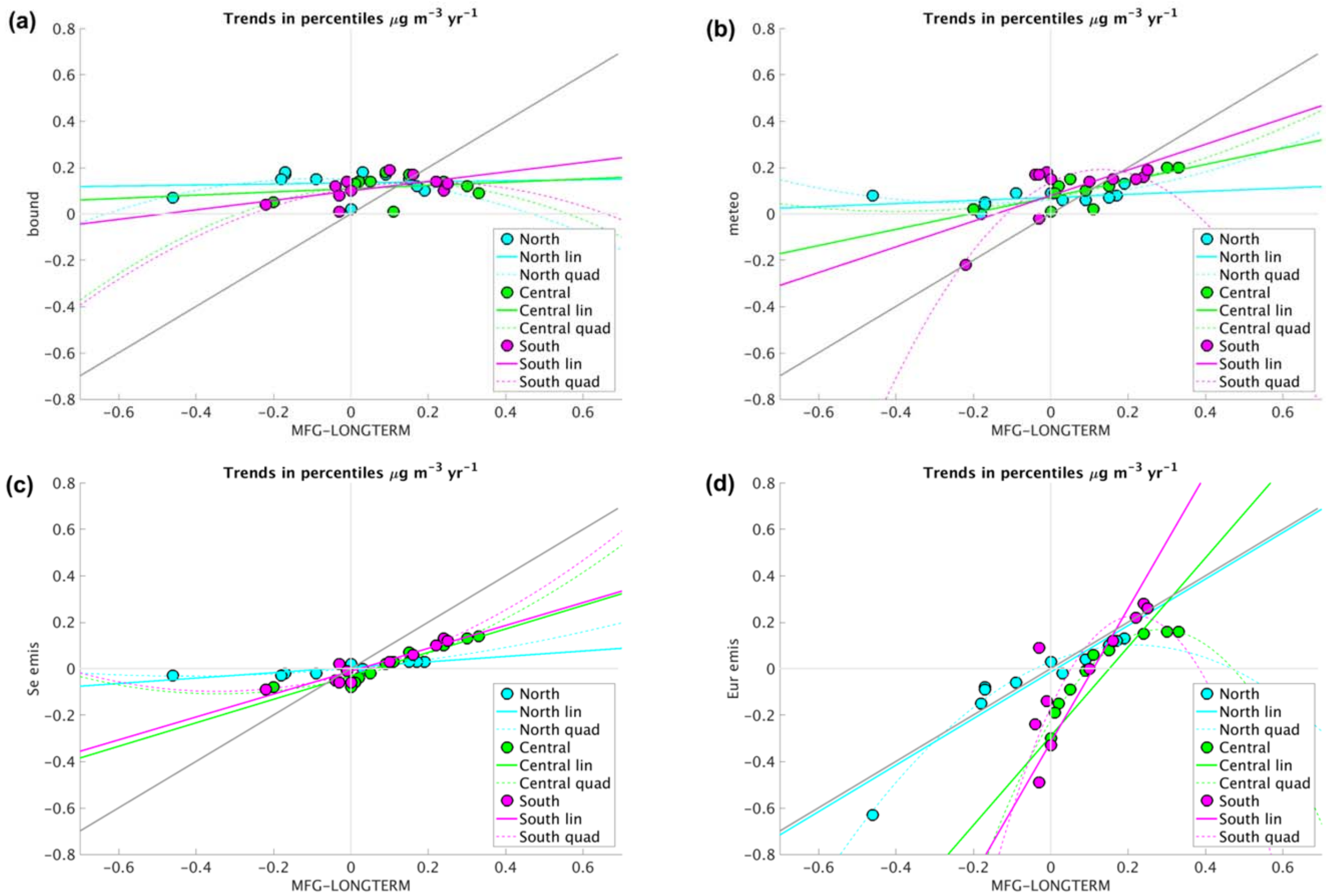

Figure S12. Trends in contributions (bound (a), meteo (b), Se emis (c), Eur emis(d)) versus the error in trend modelled by the CTM (difference between trends in the MFG reanalysis and the LONGTERM simulation) for the three regions (North (blue), Central (green) and South (magenta) Sweden, see Fig. 1 in the main paper). Circles represent different percentiles; solid line is the 1st degree and dashed line is the 2nd degree regression fit of all percentiles in the respective region.

Table S1. Evaluation of annual mean near-surface ozone concentration at Swedish measurement sites for the (observation independent) “first guess” (the MATCH base case simulation, MFG) and the two (observation dependent) reanalyzed data sets (ALL and LONGTERM) over the period 1990-2013. Mean value (mean), standard deviation (σ), mean bias normalized by the observed mean (%bias), Pearson correlation coefficients (r), root mean square error (RMSE) and mean number of years (#years) or measurement sites (#stns[1]). The top half of the table shows the mean over the 10 stations of the evaluation statistics at each measurement site (mean of yearly statistics). The bottom half of the table shows spatial evaluation statistics of the period (1990-2013) mean near-surface ozone concentration at the measurement sites (spatial statistics of multi-year means).

	mean of yearly statistics					
	mean (ppb(v))	**σ (ppb(v))**	**%bias (%)**	**r**	**RMSE (ppb(v))**	**#years**
Obs	29.8	1.7				17.3
MFG	30.0	1.1	0.5	0.46	2.65	17.3
LONGTERM	30.5	1.6	2.1	0.87	0.91	17.3
ALL	29.9	1.7	0.3	0.99	0.25	17.3
	spatial statistics of multi-year means					
	mean (ppb(v))	**σ (ppb(v))**	**%bias (%)**	**r**	**RMSE (ppb(v))**	**#stns**
Obs	29.8	1.4				10
MFG	30.0	0.9	0.5	-0.40	2.4	10
LONGTERM	30.5	1.3	2.1	0.78	1.3	10
ALL	29.9	1.5	0.3	0.99	0.2	10

[1] From the 13 Swedish measurement sites 12 were included in the evaluation, due to the requirement of a minimum of 6 years with more than 80% data coverage at observation sites. The station pair Rörvik and Råö, was considered as one site, thus the 10 sites in the spatial evaluation.

Table S2. Observed and modelled (MFG: MATCH modelled "first guess", cross: independent cross validation; ALL: observation dependent reanalysis) annual mean during 2013. Data sorted after the magnitude of the observational mean. Unit: ppb(v).

	obs	MFG	cross	ALL	
ALL					
RDB	27.0	27.1	24.4	24.6	Rödeby
SE12	27.2	30.5	31.4	27.4	Aspvreten
SE87	27.8	31.4	31.9	27.5	Östad
NM	28.7	32.8	33.1	28.9	Norr Malma
SE88	29.1	31.8	31.1	28.9	Asa försökspark
SE89	30.4	30.5	28.5	30.7	Grimsö
SE35	31.4	32.1	32.3	31.5	Vindeln
SE05	32.2	32.1	32.4	32.3	Bredkälen
SE11	32.5	32.2	32.0	32.7	Vavihill
SE14	34.0	32.4	31.2	33.8	Råö
SE32	35.1	31.3	29.1	35.6	Norra Kvill
SE13	35.1	28.8	29.5	35.2	Esrange
LONGTERM					
SE12	27.2	30.5	34.1	27.4	Aspvreten
SE35	31.4	32.1	32.3	31.4	Vindeln
SE11	32.5	32.2	34.1	32.6	Vavihill
SE14	34.0	32.4	33.1	33.8	Råö
SE32	35.1	31.3	30.0	35.4	Norra Kvill
SE13	35.1	28.8	29.5	35.2	Esrange

Table S3. Linear trend of percentiles in the 3 Swedish regions. Stars (*, **, and ***) indicate that the trend is significant ($p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, respectively). Unit: μg m^{-3} $year^{-1}$.

Percentile	**North**	**Central**	**South**
100^{th}	-0.14	-0.82**	-1.36***
98^{th}	+0.15	-0.26*	-0.40*
95^{th}	+0.24*	-0.10	-0.22
90^{th}	+0.27*	-0.02	-0.04
75^{th}	+0.21*	+0.07	+0.12
50^{th}	+0.16	+0.15	+0.20*
25^{th}	+0.17*	+0.24**	+0.30***
10^{th}	+0.17*	+0.28***	+0.39***
5^{th}	+0.17*	+0.27***	+0.43***
2^{nd}	+0.17*	+0.24***	+0.45***
0	+0.14*	+0.05*	+0.22***