



# Estimating sources of elemental and organic carbon and their temporal emission patterns using a least squares inverse model and hourly measurements from the St. Louis–Midwest supersite

B. de Foy<sup>1</sup>, Y. Y. Cui<sup>1,2,3</sup>, J. J. Schauer<sup>4</sup>, M. Janssen<sup>5</sup>, J. R. Turner<sup>6</sup>, and C. Wiedinmyer<sup>7</sup>

<sup>1</sup>Department of Earth and Atmospheric Sciences, Saint Louis University, St. Louis, MO, USA

<sup>2</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

<sup>3</sup>Chemical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration (NOAA), Boulder, CO, USA

<sup>4</sup>Civil and Environmental Engineering, University of Wisconsin, Madison, WI, USA

<sup>5</sup>Lake Michigan Air Directors Consortium (LADCO), Rosemont, IL, USA

<sup>6</sup>Energy, Environmental and Chemical Engineering Department, Washington University in St. Louis, St. Louis, MO, USA

<sup>7</sup>National Center for Atmospheric Research, Boulder, CO, USA

Correspondence to: B. de Foy (bdefoy@slu.edu)

Received: 26 March 2014 – Published in Atmos. Chem. Phys. Discuss.: 13 May 2014

Revised: 15 January 2015 – Accepted: 4 February 2015 – Published: 5 March 2015

**Abstract.** Emission inventories of elemental carbon (EC) and organic carbon (OC) contain large uncertainties both in their spatial and temporal distributions for different source types. An inverse model was used to evaluate EC and OC emissions based on 1 year of hourly measurements from the St. Louis–Midwest supersite. The input to the model consisted of continuous measurements of EC and OC obtained for 2002 using two semicontinuous analyzers. High resolution meteorological simulations were performed for the entire time period using the Weather Research and Forecasting Model (WRF). These were used to simulate hourly back trajectories at the measurement site using a Lagrangian model (FLEXPART-WRF). In combination, an Eulerian model (CAMx: The Comprehensive Air Quality Model with Extensions) was used to simulate the impacts at the measurement site using known emissions inventories for point and area sources from the Lake Michigan Directors Consortium (LADCO) as well as for open burning from the Fire Inventory from NCAR (FINN). By considering only passive transport of pollutants, the Bayesian inversion simplifies to a single least squares inversion. The inverse model combines forward Eulerian simulations with backward Lagrangian simulations to yield estimates of emissions from sources in current inventories as well as from emissions that

might be missing in the inventories. The CAMx impacts were disaggregated into separate time chunks in order to determine improved diurnal, weekday and monthly temporal patterns of emissions. Because EC is a primary species, the inverse model estimates can be interpreted directly as emissions. In contrast, OC is both a primary and a secondary species. As the inverse model does not differentiate between direct emissions and formation in the plume of those direct emissions, the estimates need to be interpreted as contributions to measured concentrations. Emissions of EC and OC in the St. Louis region from on-road, non-road, marine/aircraft/railroad (MAR), “other” and point sources were revised slightly downwards on average. In particular, both MAR and point sources had a more pronounced diurnal variation than in the inventory. The winter peak in “other” emissions was not corroborated by the inverse model. On-road emissions have a larger difference between weekday and weekends in the inverse estimates than in the inventory, and appear to be poorly simulated or characterized in the winter months. The model suggests that open burning emissions are significantly underestimated in the inventory. Finally, contributions of unknown sources seems to be from areas to the south of St. Louis and from afternoon and nighttime emissions.

## 1 Introduction

Within fine particulate matter ( $PM_{2.5}$ ), elemental carbon (EC) and organic carbon (OC) are thought to be some of the components most strongly associated with adverse health effects (Janssen et al., 2011; Bell et al., 2009; Rohr and Wyzga, 2012). In addition, black carbon (BC) has been identified as an important contributor to climate change (Bond et al., 2013; Ramanathan and Carmichael, 2008).

EC and OC are prevalent in the USA, with OC making up 20 to 40 % of  $PM_{2.5}$  in the upper Midwest and EC making up 5 to 15 % in urban areas and 3 to 5 % in rural areas (Hand et al., 2012). Levels of OC are more regionally homogeneous whereas levels of EC vary more between urban areas, while overall trends of total carbon have been decreasing nationwide (Hand et al., 2013). These observations are consistent with observations of the dynamic formation of organic aerosols leading to regional OC levels (Jimenez et al., 2009; Robinson et al., 2007).

In the Midwest, Lewandowski et al. (2008) found a strong seasonal signal in secondary OC that was associated with biogenic emissions. This production of secondary organic aerosol is insufficiently captured by current models leading to large underestimations of OC (Spak and Holloway, 2009). Napelenok et al. (2014) used source-specific tracers to identify the origin of particulate carbon as well as weaknesses in models and emissions inventories. This highlighted improvements required for secondary organic aerosol formation as well as uncertainties in mobile sources and forest fires.

Snyder et al. (2010) analyzed semicontinuous and daily averaged EC and OC measurements in the Midwest, finding that sites with similar concentrations of EC and OC could nonetheless be impacted by very different source types. This leads to the risk of misattributing impacts from distinct sources due to compensating errors in models, as also described in Napelenok et al. (2014). In Milwaukee, de Foy et al. (2012a) found that EC levels were predominantly due to mobile sources, although simulations suggested that 10 % could be due to shipping emissions from the Port of Milwaukee. While EC levels are clearly associated with mobile sources, the ratios of EC to other pollutants can vary between cities and there is therefore a clear need to improve monitoring of EC and OC in order to improve emissions inventories (Reche et al., 2011). This is illustrated by Gentner et al. (2012) who evaluate the different contributions of gasoline and diesel vehicles to secondary organic aerosol concentrations.

The present study is based on continuous, hourly measurements of EC and OC made during 2002 at the St. Louis–Midwest supersite (Bae et al., 2004b). Bae et al. (2004a) analyzed the temporal profiles of EC, OC and the EC to OC ratio. Both EC and OC have minimum concentrations from February to May. OC has maximum concentrations during the summer whereas EC has maximum concentrations during the fall. EC was found to vary by day of week with

a mid-week maximum and a minimum on Sundays. Furthermore, EC had peak concentrations in the morning and early evening. In contrast, OC does not vary by day of week and has a different diurnal pattern than EC, with lower concentrations during the early afternoon. Analysis of the EC to OC ratio suggest that the morning peak in EC is related to traffic emissions, but that the evening peak may be due to meteorological factors. Measurements of water-soluble OC (Sullivan et al., 2004) suggested that a significant fraction of the OC is from secondary organic aerosol formation, in agreement with the different temporal profiles of OC and EC at the measurement site. Sheesley et al. (2007) further analyzed the EC and OC data along with organic tracers. In addition to detecting impacts from point sources, they found differences in the temporal profiles in St. Louis with those of southern California. Bae et al. (2006) used 24 h averaged data for source attribution of OC. This identified a significant component of OC due to wood smoke and to high-emitting smoker vehicles. Jaekels et al. (2007) used positive matrix factorization to identify contributions to OC concentrations. They likewise found a strong component of wood combustion and secondary organic aerosol. In addition, they identified a mobile factor which has a strong monthly variation.

Cluster analysis of  $PM_{2.5}$  composition has shown that St. Louis has similar aerosol composition as other industrial midwest cities such as Chicago, Detroit and Cleveland (Austin et al., 2013). The St. Louis–Midwest supersite is impacted by metal processing point sources to the southwest, as shown using wind roses and conditional probability functions by Amato and Hopke (2012); Wang et al. (2011); Lee et al. (2006); Lee and Hopke (2006). These studies also confirmed the regional nature of OC, with source regions broadly from the southeast and southwest quadrants. EC has a similar signature, although it is less homogeneous and points to more source directions. Lee and Hopke (2006) used the Potential Source Contribution Function method based on back trajectories to show that sulfate levels at the site were impacted by the Ohio River valley, while nitrate levels were associated with transport from the west and northwest.

In this paper, we study the same year-long hourly time series of EC and OC measured at the St. Louis–Midwest supersite. We seek to obtain improved estimates of the diurnal and monthly emission profiles of specific types of sources by combining forward simulations of EC and OC concentrations from emissions inventories with the measurements using an inverse model. This is carried out for five different source categories as well as for emissions from open burning. In addition, the inverse model uses gridded back trajectories to identify regions that may be missing sources in the inventory. As discussed above, EC is not formed in the atmosphere, but rather emissions are transported until they are removed by deposition such that they can be simulated as passive tracers. In contrast, OC is both emitted and produced in the atmosphere. Our model is focused on transport; consequently, the results for EC can be straightforwardly com-

pared to emission inventories. For OC however, the model does not distinguish between primary OC that is emitted by a source and secondary OC that is created in the plume of that same source. The results are therefore best interpreted in terms of impacts at the measurement site rather than emissions at the source location.

## 2 Methods

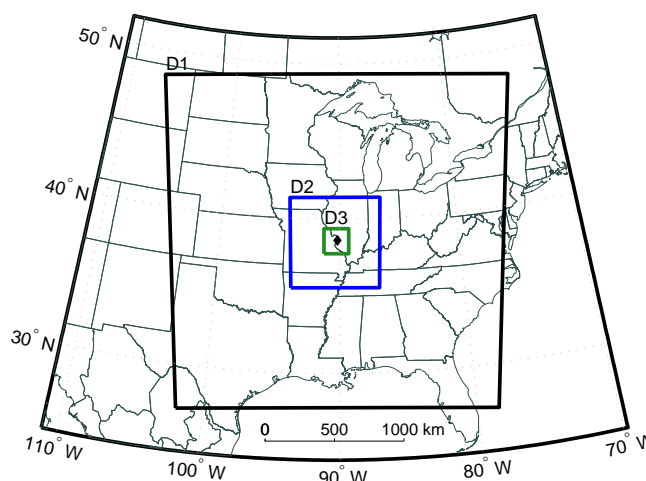
### 2.1 Measurements

The measurement site is the St. Louis–Midwest supersite which was funded by the United States Environmental Protection Agency (EPA). It is located in East St. Louis, approximately 3 km east of the central business district of St. Louis, on the other side of the Mississippi river in a low-density, mixed-use neighborhood impacted by industrial point sources nearby. Elemental carbon (EC) and organic carbon (OC) concentrations were measured using two Sunset Laboratory semicontinuous ECOC field analyzers. By having two instruments operating in tandem, it was possible to obtain continuous hourly measurements with one instrument in the collection phase while the other instrument was in the analysis phase. The measurements were validated against 24 h samples and are described in detail in Bae et al. (2004b). This study is based on 7091 valid data points measured for the duration of 2002. Fig. 1 shows the location of the measurement site.

Hourly meteorological observations were obtained from Lambert–St. Louis International Airport (KSTL) and St. Louis Downtown Airport (KCPS) in Cahokia, IL from the Integrated Surface Hourly Data available from the National Climatic Data Center. KSTL is across the Mississippi river 24 km northwest of the measurement site, and KCPS is 5 km south of the measurement site on the same side of the river. Meteorological data were also available at the supersite. These data were in agreement with the KCPS data, but the latter was more complete and was therefore selected for the analysis.

### 2.2 Emissions inventory

The Lake Michigan Air Directors Consortium (LADCO) emissions inventory for 2007 for the Midwest was used as a prior for the inverse model (LADCO, 2011). It is calculated on a 12 km grid with diurnal and monthly profiles and emissions separated by source category for on-road, non-road, marine/aircraft/railroad (MAR), “other”, biogenic and point sources. Point source emissions were specified using 2007 CEMS (continuous emission monitoring systems) data with updated temporal profiles to include adjustments for weekend/weekday emissions while still providing a solid platform for future projections (Edick and Janssen, 2006). Mobile emissions were estimated using the MOVES2010a model (EPA, 2012). Non-road emissions were updated to



**Figure 1.** Domains used for the WRF simulations: large (D1, 27 km resolution), Regional (D2, 9 km resolution) and Local (D3, 3 km resolution). CAMx simulations are performed on the Regional and Local domains, except for open burning which are performed on the Large and the Regional domains. The diamond shows the location of the St. Louis–Midwest supersite.

reflect higher agricultural equipment emissions during the spring and fall season rather than the default of a single summer maximum based on midwest crop calendars and tilling, planting, pesticide application and harvesting cycles (Thesing et al., 2004). For EC and OC, “other” sources consist mainly of residential wood and waste combustion with smaller contributions from unpaved roads, food preparation and construction.

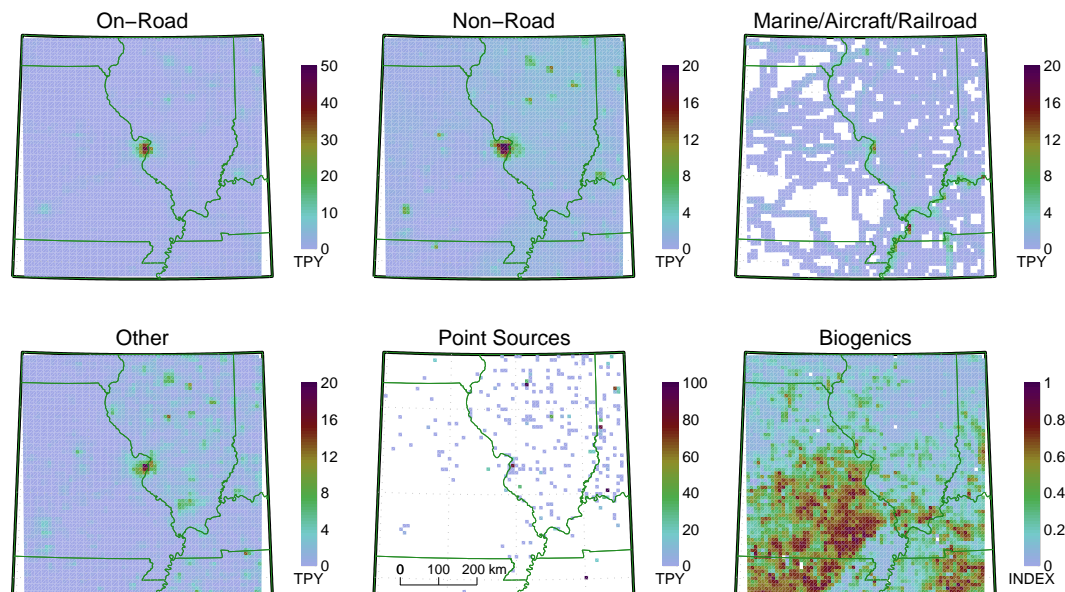
Figure 2 shows the emissions for EC in metric tonnes per year (tpy). OC emissions have similar patterns with the following average OC to EC ratios: 0.62 for on-road, 0.64 for non-road, 0.49 for MAR, 6.7 for “other” and 2.5 for point sources. Table 1 presents the emission totals for the Regional domain shown in Fig. 1.

Biogenic emissions in LADCO (2011) were calculated using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.03a (Guenther et al., 2006). As an example, Fig. 2 shows the spatial map of the biogenic emissions of condensable gases, category “CG5” in non-dimensional units. These will be used as a tracer of biogenic emissions in the Eulerian simulations, and the concentrations will be normalized before being included in the inversion algorithm. As will be discussed in Sect. 3.2, the inverse results therefore do not represent an estimate of actual biogenic emissions, but rather an estimate of the fraction of OC that could be ascribed to aerosol formation due to these emissions.

In order to have an additional comparison to the LADCO prior emissions and the inverse model results, the 2008 National Emissions Inventory (NEI) version 3 was obtained from the US Environmental Protection Agency. EC and OC

**Table 1.** Emission totals for EC and OC for the Regional domain around St. Louis by source category for the National Emissions Inventory (NEI), the LADCO inventory and the least squares inverse model. Note that OC Inverse totals combine primary emissions and secondary formation. (MAR = marine, aircraft and railroad.)

Source type	Elemental carbon (tpy)			Organic carbon (tpy)		
	NEI	LADCO	Inverse	NEI	LADCO	Inverse
On-road	2910	4268	2060	1663	2648	2495
Non-road	5896	5818	4729	2237	3740	5037
MAR	1803	2278	1652	411	1126	1069
“Other”	3217	4312	2248	24 799	28 907	26 399
Point sources	1724	1572	1331	2061	3892	2751



**Figure 2.** Elemental carbon emissions by source type from the LADCO inventory for the Regional domain in metric tonnes per year, and biogenic tracer emissions in non-dimensional units.

emissions were available in speciated files for  $\text{PM}_{2.5}$ . The on-road emissions in the NEI were calculated using the MOVES2010b model (EPA, 2012). The data were provided as annual totals by Federal Information Processing Standards (FIPS) codes. These were mapped to the Regional model grid in order to compare NEI emissions with the emissions in the LADCO prior and with the inverse model posterior.

EC and OC have experienced a downward trend in the US, with around 1 to 2 % decreases per year (Hand et al., 2013). This means that emissions calculated based on 2002 measurements could be expected to be 5 to 10 % higher than an emissions inventory for 2007. Although emission inventories existed for 2002, it was felt that the considerable improvements and developments that went into the LADCO 2007 inventory meant that this would be a better choice for the prior, and that consequently the 2008 NEI was the most appropriate comparison point to the prior. Nonetheless, the temporal discrepancy should be borne in mind when interpreting the results.

EC and OC emissions from open burning were calculated using the Fire Inventory from NCAR (FINN) version 1 (Wiedinmyer et al., 2011). FINN calculates daily emissions from fires identified by fire counts from the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) fire and thermal anomalies data provided from the official NASA MCD14ML product, Collection 5, version 1 (Giglio et al., 2003). Land cover and vegetation density needed to calculate the emissions were determined with the MODIS Land Cover Type product (Friedl et al., 2010) and the MODIS Vegetation Continuous Fields product (Collection 3 for 2001) (Hansen et al., 2003, 2005; Carroll et al., 2011), and fuel loadings from Hoelzemann et al. (2004) and Akagi et al. (2011). Ecosystem-specific emission factors for EC and OC emissions were compiled from existing literature (Table 1, Wiedinmyer et al., 2011). Ratios of OC to EC emission factors range from 4.8 for fires in croplands, to 39 for fires in boreal forests. Daily emission totals were distributed evenly

throughout the day as input to CAMx (The Comprehensive Air Quality Model with Extensions) simulations.

In FINN, open burning includes the fires which are detected by Terra MODIS. These are a combination of forest fires, prescribed burns and larger agricultural fires, with a minimum burn area of 1 km<sup>2</sup>. Hawbaker et al. (2008) analyzed the detection rate of MODIS compared to a set of reference fires. The rates were high when both Terra and Aqua were used, but dropped to 60 % in the Great Plains and 39 % in the eastern US when only Terra was used. Because we only have Terra data for 2002, this is an added source of uncertainty in the emission estimates.

Figure 3 shows the total gridded open burning emissions for 2002 on the Large model domain, and Table 2 shows the total emissions by sector. The inverse model calculated posterior emissions independently for the following six geographical sectors: local emissions within 100 km of the measurement site followed by the northeast, southeast, southwest, west and northwest as shown in Fig. 3. The largest emissions are in the southeast and southwest sector.

### 2.3 Numerical simulations

The meteorological simulations were performed with the Weather Research and Forecasting (WRF) model version 3.5.1 (Skamarock et al., 2005). The North American Regional Reanalysis (NARR) (Mesinger et al., 2006) was used for the initial and boundary conditions. The simulations used 3 domains with 27, 9 and 3 km horizontal resolution and 40 vertical levels. Figure 1 shows a map of the 3 domains, which will be referred to as the Large, the Regional and the Local domains.

The model was run with two-way nesting, with the Yonsei University (YSU) boundary layer scheme, the Kain–Fritsch convective parameterization, the NOAH land surface scheme, the WSM 3-class simple ice microphysics scheme, the Dudhia shortwave scheme and the Rapid Radiation Transfer Model longwave scheme. Individual simulations were performed lasting 162 h, of which the first 42 h were considered spin-up time and the remaining 5 days were used for analysis. The simulations are similar to those described in de Foy et al. (2014), where it was shown that the model accurately represents the statistical distribution of temperature, humidity, wind speed and wind direction at the surface (see Fig. 3 in de Foy et al., 2014).

Particle back trajectories were calculated from the super-site with FLEXPART (Stohl et al., 2005), using FLEXPART-WRF (Brioude et al., 2013b) for a duration of 4 days starting every hour of the year using the WRF simulated wind fields. 1000 particles were released per hour between 0 and 100 m above the ground and were allowed to disperse in three dimensions using the WRF mixing heights and surface friction velocity. The particles were treated as passive tracers with neither wet nor dry deposition. Sensitivity tests presented in de Foy et al. (2012b) found that 1000 particles were suffi-

cient to ensure that the results did not depend on the number of particles for inversions on a regional scale. The particle positions were converted to polar grids to provide a residence time analysis (RTA, Ashbaugh et al., 1985). This represents the amount of time that an air mass has spent in different grid cells before arriving at the measurement location and can be rescaled to yield the impact that a source in each grid cell would have at the receptor site (Seibert and Frank, 2004; Lin et al., 2003).

Concentration field analysis (CFA, Seibert et al., 1994; de Foy et al., 2009, 2007) was used as a preliminary method to evaluate possible source regions suggested by the residence time analysis and the hourly concentrations. Concentration field analysis is based on scaling the residence time analysis at each time step with the concentration at the measurement site. The sum over the entire measurement period is then normalized with the residence time analysis. This highlights air flow patterns that are associated with high receptor concentrations. As described in Sect. 3.1 below, standard CFA is sensitive to peak concentrations, and so we apply the method to an estimate of the column amount of pollutant. This “column CFA” is shown below to give a more reliable estimate of potential source regions than using CFA based on surface concentrations alone.

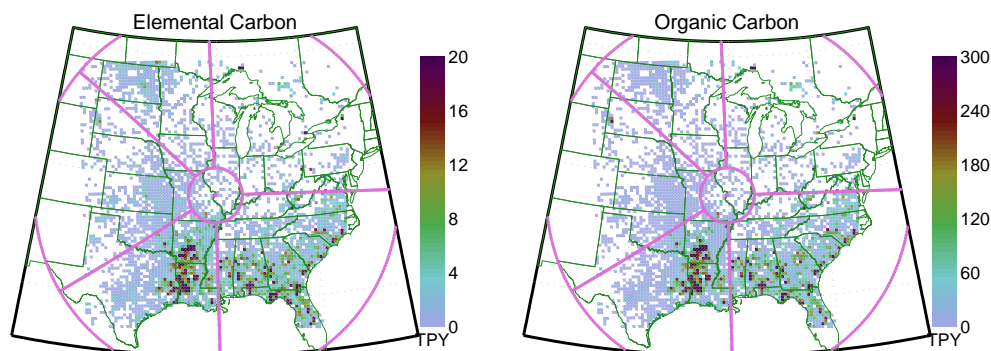
The Comprehensive Air Quality Model with Extensions (CAMx v6.00, ENVIRON, 2013), an Eulerian 3-D grid model, was used to obtain hourly concentrations of EC and OC at the measurement site based on the prior emissions inventory. Dry deposition was calculated using the Zhang et al. (2003) scheme, and wet deposition using the standard scheme in CAMx. For the LADCO inventory, CAMx was run with two nested domains: the Regional and the Local domains from the WRF simulations (shown in Fig. 1), whereas for open burning, CAMx was run with the Large and the Regional domains.

This study is focused on estimating source contributions from specific source groups based on atmospheric transport and therefore does not use the aerosol module in CAMx. Both EC and OC are simulated as passive tracers with wet and dry deposition. This is adequate for EC, and so the inverse model results can be straightforwardly compared to the emissions inventories. In contrast to EC, there is extensive formation of OC in the atmosphere which is not simulated in our model. This means that the inversion will not distinguish between primary and secondary OC, and that results are therefore best interpreted as impacts at the measurement site rather than as emissions at the source location. It also means that we are not able to evaluate the non-linear interactions of different plumes together.

### 2.4 Least squares inverse model

The least squares inverse model used in the present study was developed in (de Foy et al., 2012b, 2014), where it was used to evaluate emissions inventories of elemental and re-





**Figure 3.** Open burning emissions of EC and OC for the Large domain for 2002 using the FINN model, which include forest, prescribed and agricultural fires detected by Terra MODIS. Pink lines show the six sectors used in the inverse model, pink dot is the supersite.

**Table 2.** Emission totals for open burning by geographical sector relative to the measurement site for the FINN model and the least squares inverse model. Also shown are the ratios of the inverse emission estimates to the FINN prior estimates and the fraction of EC or OC at the measurement site that is estimated to be due to open burning. Note that OC Inverse totals combine primary emissions and secondary formation.

Sector	Elemental carbon				Organic carbon			
	FINN tpy	Inverse tpy	Ratio	Impact %	FINN tpy	Inverse tpy	Ratio	Impact %
Local – 100 km	173	4708	27.27	0.69	1508	21 696	14.39	0.47
Northeast	553	14 031	25.37	0.18	6430	130 196	20.25	0.22
Southeast	5088	180 228	35.42	1.36	66 187	1 541 067	23.28	2.53
Southwest	3714	11 749	3.16	0.86	48 028	103 156	2.15	1.63
West	558	592	1.06	0.82	4149	3611	0.87	1.00
Northwest	459	0	0.00	0.00	2899	0	0.00	0.00
Total	10 577	211 309	20.0	3.5	129 409	1 799 725	13.9	5.1

active mercury. The inverse model estimates emissions that contribute to measured concentrations at a receptor site. This is done by using both the passive transport from prior sources and the contribution of unknown sources using gridded back trajectories.

Inverse models based on back trajectories alone include Stohl et al. (2009); Brioude et al. (2011, 2013a). This work combines back trajectories with Eulerian simulations, and in this respect is similar to the methods presented in Rigby et al. (2011); Rödenbeck et al. (2009). The purpose of combining the Lagrangian and Eulerian simulations for Rigby et al. (2011); Rödenbeck et al. (2009) was to combine global transport of inert species with higher definition impacts from specific locations. In our case, the background levels of EC and OC are very low (see Fig. 4), and we expect minimal impacts from sources outside the study area. The purpose of combining Eulerian with Lagrangian simulations is therefore to estimate adjustments to known emission inventories with the Eulerian simulations, and to estimate impacts from unknown area sources in an overlapping domain with the Lagrangian simulations.

Hourly Eulerian simulations with CAMx were performed for the five different source groups in the LADCO inventory:

on-road, non-road, MAR, “other” and point sources. Because we are interested in evaluating the temporal profiles of the sources, we carry out separate simulations for emissions during different times of the day and different days of the week. The time slots were selected based on the diurnal profile used in the emissions inventory: 11:00 p.m. to 5:00 a.m., 5:00 to 8:00 a.m., 8:00 a.m. to 2:00 p.m., 2:00 to 6:00 p.m., and 6:00 to 11:00 p.m. Days of the week were split into a weekday group and a group containing Saturdays, Sundays and holidays. As an example, an hourly time series of concentrations was obtained from a CAMx simulation with on-road emissions only between 5:00 to 8:00 a.m. on weekdays. With 5 source groups, 5 time slots and 2 day types, this means that there were 50 CAMx simulations. We are also interested in the annual profile of the emissions, and so we divide the 50 resulting concentration time series into 12 months for a total of 600 input time series into the inverse model. With this method of resolving temporal profiles, individual time series are used for each temporal interval of interest. This is in contrast with Brunner et al. (2012) who use a Kalman filter to identify seasonal changes in emissions.

The open burning emissions are included in the inversion as six time series simulated by CAMx for the entire year

for the six geographic sectors shown in Fig. 3. We also include a CAMx time series representing impacts from biogenic emissions, as discussed in Sect. 3.2.

In addition to the forward Eulerian simulations, we perform backward Lagrangian simulations of particle back trajectories for each hour of the measurement campaign. These are mapped onto a polar grid surrounding the measurement site. The time series from each grid cell gives an estimate of the concentration at the measurement site that would be caused by a constant area emission in that cell. We divided these gridded time series into impacts due to weekdays and weekends, and also into four time slots during the day: 3:00 to 9:00 a.m., 9:00 a.m. to 3:00 p.m., 3:00 to 9:00 p.m., and 9:00 p.m. to 3:00 a.m. These were selected to capture the morning and afternoon rush hours in the middle of two of the slots, and to differentiate the daytime and nighttime emissions between those. The polar grid was chosen to have eighteen  $20^\circ$  segments, in 20 radial bands extending to 1000 km from the measurement site. There were therefore 360 time series from 8 time slots, for a total of 2880 time series to be used as input into the inverse model.

The inverse model derives a posterior estimate of emissions based on the Eulerian simulations that used the emissions inventory as a prior. In addition, the inverse model uses the Lagrangian simulations to derive an estimate of sources that may be missing from the inventory. This is done by using the polar grids of residence time analysis that represent the impact that an emission in a given grid cell would have at the measurement site. As all the known sources were already included in the CAMx simulations with the emissions inventory, we use a field of zero prior emissions for the polar grids from the Lagrangian simulations.

By limiting the input of the model to passive tracers and individual time series, we can use a least squares simplification developed in de Foy et al. (2012b) to the Bayesian formulation used in Stohl et al. (2009). This hybrid least squares method derives an estimate of the emissions vector  $\mathbf{x}$  that minimizes the cost function  $J$  given by the sum of the observation cost function and the emissions cost function:

$$J = \|(\mathbf{H}\mathbf{x}' - \mathbf{y}')\|_2 + \alpha^2 \|\mathbf{x}'\|_2. \quad (1)$$

Where  $\mathbf{x}' = \mathbf{x} - \mathbf{x}_o$  is the vector of emission corrections given prior emissions estimates  $\mathbf{x}_o$ . The individual entries in  $\mathbf{x}$  can take different forms: they can be actual emissions in units of mass per time, or they can be non-dimensional scaling factors.  $\mathbf{H}$  is the sensitivity matrix that converts emissions parameters  $\mathbf{x}$  into simulated concentrations. Vector  $\mathbf{y}' = \mathbf{y} - \mathbf{H}\mathbf{x}_o$  is the residual between the vector of concentration measurements  $\mathbf{y}$  and the time series produced by the prior emissions estimates  $\mathbf{H}\mathbf{x}_o$ .  $\alpha$  is the regularization parameter that balances the two parts of the cost function. In practice,  $\alpha$  can be replaced by a vector of parameters  $\mathbf{s}$  that scales each term in  $\mathbf{x}$  within the  $L_2$  norm. In this way, the method was shown to be equivalent to a Bayesian derivation when diagonal error covariance matrices are used (de Foy et al.,

2012b; Wunsch, 2006; Aster et al., 2012). In these cases, the regularization parameter is equal to the ratio of the uncertainty of the measurements to the uncertainty of the emissions parameter, as described in de Foy et al. (2012b).

The columns of  $\mathbf{H}$  contain the 606 input time series from the forward Eulerian simulations (in the same units as the measurements) as well as the 2880 time series from the back-trajectory grids (in units relating area emissions to measurement concentrations, see de Foy et al., 2012b), all of which are hourly time series for the whole of 2002. The rows of  $\mathbf{H}$  correspond to the impact of the different sources for each of the 7091 h with valid data, which are contained in vector  $\mathbf{y}$ . The vector  $\mathbf{x}$  contains  $(606 + 2880)$  entries which yield the posterior emissions estimate for the source groups and for the gridded area sources represented by the back trajectories. For the CAMx time series, the entries in  $\mathbf{x}$  are scaling factors on the LADCO emissions that went into the CAMx simulations. For the FLEXPART polar grids, the entries in  $\mathbf{x}$  represent emissions.

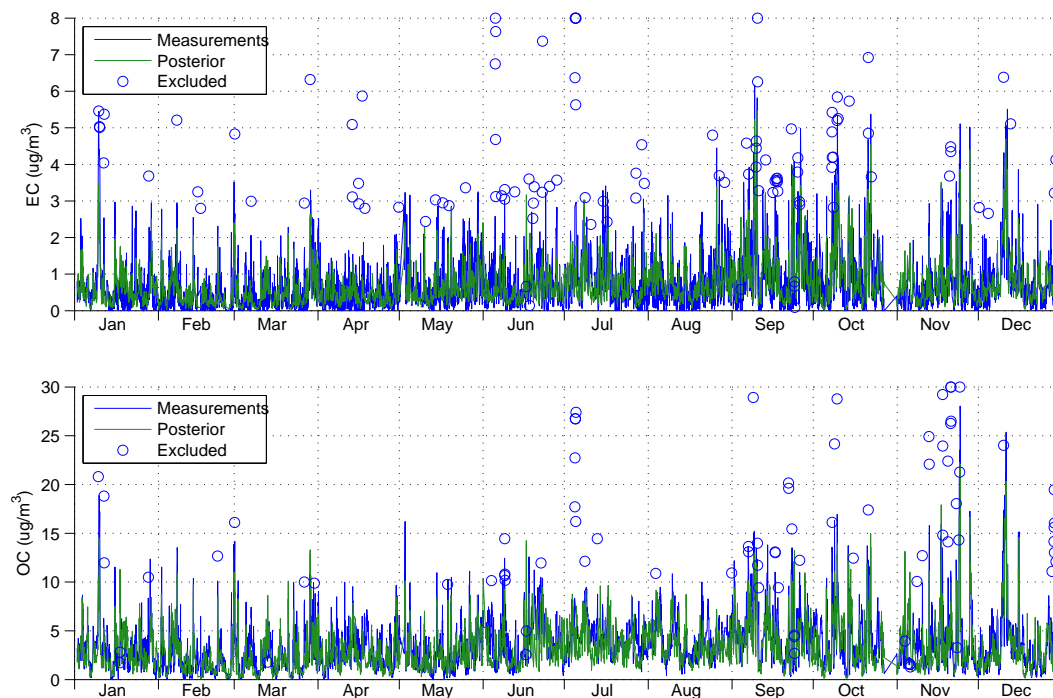
The system of equations can be solved with a single step of least squares using

$$J = \|\mathbf{s} \cdot (\mathbf{H}''\mathbf{x}' - \mathbf{y}'')\|_2, \quad (2)$$

where  $\mathbf{H}'' = (\mathbf{H}, \mathbf{I})$  and  $\mathbf{y}'' = (\mathbf{y}', \mathbf{x}_{\text{zero}})^T$  are the augmented versions of  $\mathbf{H}$  and  $\mathbf{y}'$ .  $\mathbf{I}$  is the identity matrix the size of  $\mathbf{x}$ , and  $\mathbf{x}_{\text{zero}}$  is a vector of zero values. Hence, the first part of  $\mathbf{H}''$  and of  $\mathbf{y}''$  correspond to the observation cost function and the second part to the emissions vector cost function.  $\mathbf{H}''$  has dimensions of  $(7091 + 3486)$  by  $(3486)$ , and  $\mathbf{y}''$  has dimensions of  $(7091 + 3486)$ . The vector  $\mathbf{s}$  contains scaling factors on the parts of the cost function: these are taken to be unit values for the observation cost function and contain the regularization parameter  $\alpha$  for the emissions cost function. A strength of the method is that boundaries can be straightforwardly applied to the vector  $\mathbf{x}'$  during the least squares solution to prevent nonphysical negative emissions.

An iteratively reweighted least squares (IRLS) scheme is used to reduce the sensitivity of the method to outliers in the data: after solving for  $\mathbf{x}$ , observation times that have a residual larger than 3 times the standard deviation of the residual values are removed from the analysis. This is performed iteratively to converge on a stable set of times to include in the inversion. EC and OC simulations were evaluated separately using the inverse model.

In a Bayesian framework, uncertainty estimates are required to obtain the error covariance matrices on the two parts of the cost function. In the absence of detailed prior information, Efron (2013) recommends using empirical Bayes methods where the prior information is obtained from the data set itself. If this is insufficient, then using frequentist methods is recommended as a check on the Bayesian simulations. In this context, the current method can be understood as a frequentist method where the inversion is performed multiple times using bootstrapping, and where the regularization parameter is obtained from the data itself.



**Figure 4.** Time series of elemental and organic carbon at the St. Louis–Midwest supersite for 2002. Measurements are shown in blue, circles show the data points excluded from the analysis by the iteratively reweighted least squares scheme. Green line shows the posterior time series, as produced by the least squares inverse model.

The inverse model therefore does not need prior error estimates, but rather relies on an optimization routine to determine the values of the regularization parameters in the vector  $s$  that minimize the total error following Henze et al. (2009). While in principle we can ascribe different values for each entry in the sensitivity matrix, we decided to use common values by source groups. The values of  $s$  were therefore determined separately for the emissions inventory sources, for the open burning sources and for the emissions based on back trajectories. The regularization parameter for the gridded emissions is scaled by the cell area to account for the increase in uncertainty with increasing distance from the measurement site. This yields values for the EC inversion of 0.025 for gridded emissions, 1 for emissions inventory sources and 0.03 for open burning. The corresponding parameters for OC are 0.015, 1, 0.25 and an additional parameter of 0.5 for the biogenic contribution. Taking the uncertainty of the measurements to be  $1 \mu\text{g m}^{-3}$ , this corresponds to an uncertainty of 100 % for the emissions inventory sources, and to an uncertainty factor for open burning of 33 for EC and 4 for OC.

Miller et al. (2014) review different methods to enforce positive emissions in the inversion, and show that some of these may bias the results. In the inverse model, the inversion is performed by the function `lsqin` in Matlab. This uses a trust-region reflective Newton method to solve the least squares problem and enforce positive constraints on the results. This does not prevent the model from estimating uncer-

tainties, as we derive a regularization parameter from the data and obtain the uncertainty estimates using bootstrapping.

We estimate uncertainties in the inverse model by two different methods. The first is to use expert judgment to determine an uncertainty on the measurements ( $y$ ) and on the model sensitivities ( $\mathbf{H}$ ) and to use Monte Carlo error propagation. We perform 100 realizations of the inversion with randomized scaling of the entries in  $y$  and  $\mathbf{H}$  in order to estimate the uncertainties in  $x$ . In practice, we assume that entries in  $y$  vary by plus or minus 20 % and those in  $\mathbf{H}$  by plus or minus 50 %.

An alternative method is to assume that by randomly sampling the data included in the inversion we are randomly sampling both the measurement errors and the simulation errors at the same time. This can be done with the bootstrap algorithm. Although measurement errors are assumed to be uncorrelated in time, meteorological events vary on the order of hours to days. In order to obtain samples that have different meteorological conditions, we perform block-bootstrapping with a block length of 24 h. We therefore perform 100 inversions with random selection with replacement of the days included in the analysis. In this way, the bootstrapping yields an estimate of the combined uncertainty due to measurement errors and due to transport modeling errors.

In outline, we first perform the optimization of the regularization parameters without bootstrapping for each set of sources in turn: for the RTA grids, for the LADCO emis-



sions, for the open burning emissions and for the biogenic tracer. This is repeated to make sure the values are stable. We then use the set of regularization parameters to obtain inverse results with the full data set, and for 100 realizations with block-bootstrapping.

### 3 Results and discussion

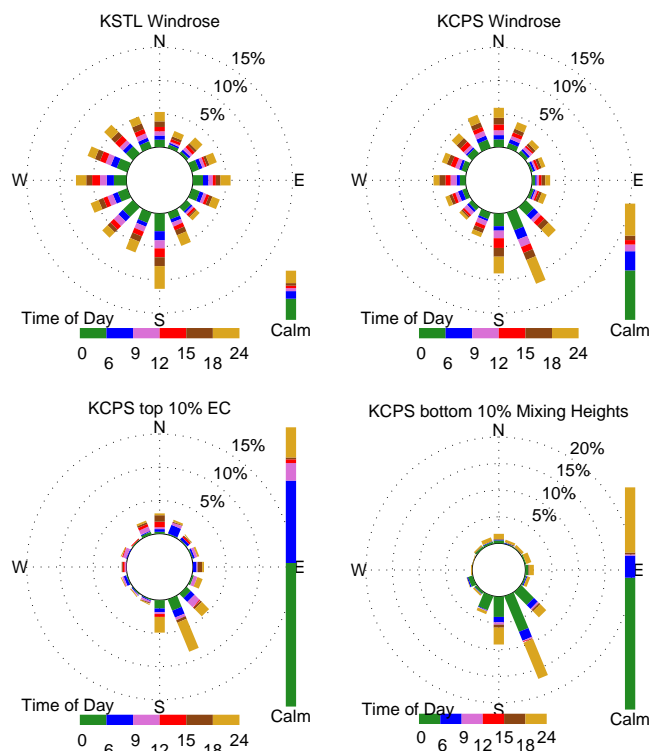
#### 3.1 Data analysis

Before presenting the results of the inverse model, this section presents the results of analyzing wind roses and back trajectories from the measurement site. Winds come from all directions at the Lambert–St. Louis International Airport with a predominance for westerly flow, as shown in the wind roses in Fig. 5. Nearer to the supersite at the Downtown St. Louis airport, however, there is a clear peak of southeasterly flow and a much larger proportion of calm hours (17% compared with 7% at KSTL). As a first cut analysis, Fig. 5 shows the wind rose for the hours in the top 10 percentile of EC concentrations. 54% of these have calm winds that occur between midnight and 9:00 a.m. As for the non-calm hours, they are most frequently from the southeast. This suggests that high EC concentrations are associated with calm conditions and hence with local sources. It also suggests that significant sources could be found southeast of the site, which is at odds with known inventories.

Figure 6 shows the probability density function for both the measurements and the simulations at KCPS. The distributions are very similar, and all variables passed the Kolmogorov–Smirnov test to much lower than the 1% significance level, showing that the model does not suffer from significant systematic biases. The auto-correlation times of the meteorological variables also shows that by using block-bootstrapping with 24 h intervals we will be sampling independent events.

We use residence time analysis to display the spatial pattern of wind transport to the measurement site over the course of 2002, see Fig. 7. On the Regional domain, this shows that air masses from all directions impact the site but that there is a predominant signature from the southwest, in agreement with the wind rose at KSTL. On the Local domain, we see again impacts from all directions, but in addition there is a very clear river valley effect. Simulated particles from the south follow the Mississippi river going north towards the measurement site.

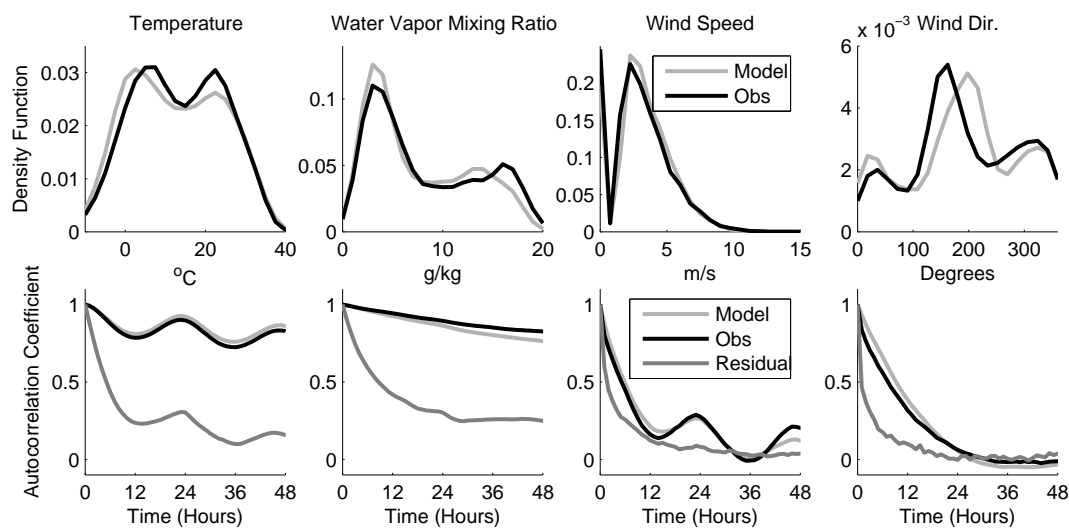
Concentration field analysis of EC and OC (Fig. 7) shows that peak concentrations are associated with transport from the southeast. This is in agreement with the pollution rose shown in Fig. 5 but is puzzling given that southern Illinois does not stand out as a large source region in Fig. 2. To resolve this conundrum, we consider the influence of mixing heights and stable atmospheric conditions at the supersite: the last rose in Fig. 5 shows the wind direction for hours with



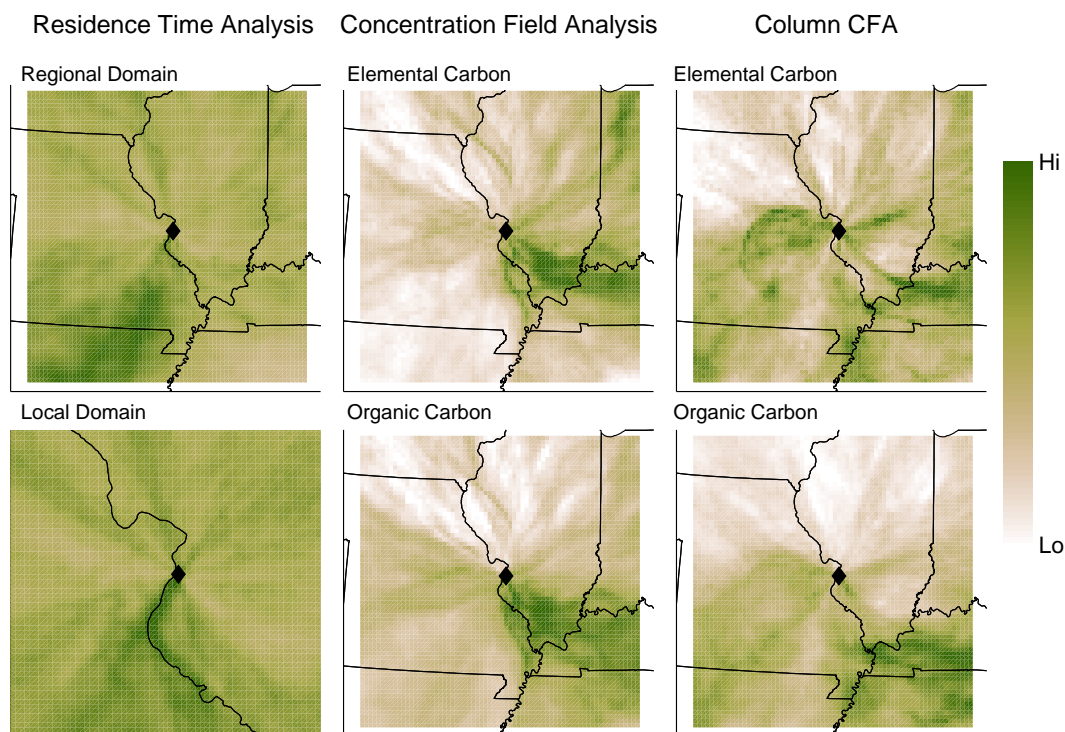
**Figure 5.** Top: wind roses for Lambert–St. Louis international airport (KSTL) and Downtown St. Louis airport (KCPS). Bottom: wind roses for hours in the top 10% of EC concentrations at the supersite using KCPS data, and bottom 10% of WRF mixing layer height. Color indicates time of day.

the lowest 10 percentile of mixing heights in the WRF simulations. This shows a picture similar to the EC pollution rose with nearly half of the hours experiencing calm winds, and the remaining having winds predominantly from the southeast. Snyder et al. (2009) found an episode where high levels of cadmium, antimony, barium and selenium were associated with a very clear southeast signature. This could be due to a power station 53 km away in that direction, although simulations with CAMx did not support such high impacts from this source. Further analysis found that peak EC concentrations are associated with these hours with very stable vertical mixing conditions which themselves are associated with weak southeasterly transport. They appear to be linked to occurrences of the low-level jet. This suggests that micrometeorology needs to be taken into account when analyzing high pollution events in St. Louis.

Wind rose analysis and CFA are sensitive to peak concentrations occurring during situations with very shallow boundary layers and so we need to expand the methods to be more sensitive to the amount of pollutant rather than to the peak concentration. This can be done by calculating a “column CFA”: CFA is carried out with an estimate of the total column of EC rather than with the surface concentration of EC. To do this, we assume that EC and OC are mainly in the plan-



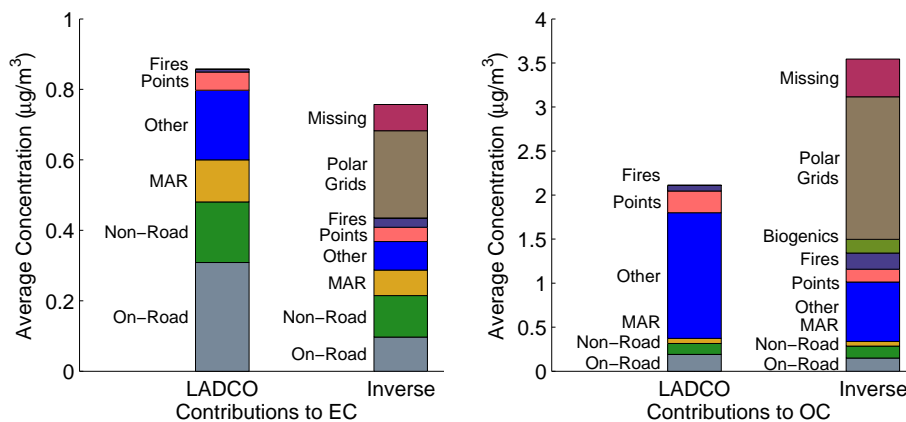
**Figure 6.** Top: probability density function of temperature, water vapor, wind speed and wind direction observations and simulations at KCPS. Bottom: autocorrelation coefficient of observations and simulations as well as of the residual between the two.



**Figure 7.** Left: residence time analysis of FLEXPART-WRF back trajectories using hourly releases during 2002 showing the origin of air masses arriving at the supersite (diamond). Center: concentration field analysis of EC and OC showing air mass transport associated with peak concentrations. Right: column concentration field analysis of EC and OC showing air mass transport associated with higher column amounts of EC and OC.

etary boundary layer and that concentrations are well mixed throughout. The column amount is obtained by multiplying the surface concentration by the height of the boundary layer. Since we do not have measurements of the mixing height, we use simulated values from the WRF model. The two graphs

on the right in Fig. 7 show the results of the Column CFA for EC and OC. For EC, we can now see a clear signature from the St. Louis metropolitan area as well as a smaller signature from the Illinois side of the urban zone. For OC, the St. Louis metropolitan area shows up but there is a stronger signal of



**Figure 8.** Contributions of different types of sources to the average concentration of EC and OC at the St. Louis–Midwest supersite using the LADCO inventory (prior) and the least squares inverse model (posterior).

general impacts from both the southeast and the southwest, which is consistent with regional atmospheric formation of OC compared with local transport of EC.

### 3.2 Inverse model results: time series and impacts

Figure 4 shows the EC and OC time series of the measurements and of the inverse model results. The time series from the inverse model are much improved compared with those simulated using the emissions prior, as shown by the statistical measures in Table 3. For the full time series, Pearson's correlation coefficient squared ( $r^2$ ) increases from around 0.1 to above 0.4. As described above, the inverse model uses Iteratively Reweighted Least Squares to reduce the impact of outliers on the results. The  $r^2$  statistics are also shown for this subset points, with an agreement of 0.53 for the EC inverse time series and of 0.56 for the OC time series.

The inverse model decomposes the measurement time series as the sum of the contributions from different source groups. If these are sufficiently well separated spatially and temporally it is possible to estimate the contribution of individual source groups to the average concentration at the site. In our current case, there is a certain level of overlap between the different source categories, as can be seen in Fig. 2. The closest time series are the impacts of on-road and those of “other” sources, with a correlation coefficient ( $r$ ) of 0.82, and with non-road sources with  $r$  of 0.65. By impacts, we mean the surface concentration of EC or OC at the measurement site that are due to transport of particular emissions to the site. The most distinct time series are the point sources, which have an  $r$  of 0.5 with MAR emissions but small or negative  $r$  with the other categories. In practice, block-bootstrapping was used to determine uncertainties in the inverse model results, and these were found to be robust, as will be discussed below.

Figure 8 shows the contributions of different source groups to average EC and OC concentrations at the measurement site

**Table 3.** Pearson's correlation coefficient squared for simulated time series of EC and OC for the complete time series as well as for the subset of points included in the inversion after the Iteratively Reweighted Least Squares (IRLS) procedure. The full inverse time series is the sum of the CAMx posterior and the impacts due to the gridded back trajectories.

$r^2$	Elemental carbon		Organic carbon	
	All points	IRLS points	All points	IRLS points
CAMx prior	0.11	0.19	0.07	0.10
CAMx posterior	0.28	0.37	0.26	0.29
Full inverse	0.42	0.53	0.47	0.56

for both the prior and the posterior emissions inventory. The prior inventory overestimates the average EC concentration at the measurement site by 13 % and suggests that on-road emissions account for 36 % of the pollutant load, non-road for 20 %, MAR for 13.9 %, “other” for 23 %, point sources for 6 % and open burning for 1 %. The posterior emissions underestimate average impacts by 10 % (“Missing” on the graph), and attribute 33 % to area emissions from the polar RTA grids. This leaves on-road emissions with 13 %, non-road with 16 %, MAR with 10 %, “other” with 11 %, point sources with 5 % and open burning with 4 %.

Whereas EC behaves as a tracer species from source to receptor, OC is due to the combination of transport from source to receptor and formation in the atmosphere during transport. Because this paper only considers transport, we expect the model results to underestimate average concentrations: the prior time series represents 60 % of the average OC concentration. As discussed in Sect. 4, this suggests that 40 % of OC at the measurement site is from secondary formation, in line with the estimate provided in Bae et al. (2006). The largest contributor in the prior is the “other” category with 68 % of simulated impacts, followed by point sources with 12 %, on-road with 9 %, non-road with 6 %, and MAR and open burning with 3 % each. The posterior accounts for 88 % of the

average OC levels, mainly by reducing the impacts of the source groups and using the RTA grids to represent 46 % of the simulated impacts. Simulated impacts from open burning are increased in the posterior so that they make up 5 % of the total OC.

Normalized time series of biogenic precursor concentrations were included in the analysis. Because the units are non-dimensional, the results from the inverse model give an indication of the fraction of EC or OC that correlates with these emissions, without giving an estimate of the emissions themselves. As expected, none of the biogenic precursors contributed to the EC time series in the inversion, and these were therefore left out of the EC inversions. For OC, we tested different biogenic components and found that condensable gases category 5 “CG5” yielded the best inverse time series of OC compared to the measured time series. The model was therefore run just with this species as an input. The model estimated that 4 % of simulated OC at the measurement site is associated with emissions of CG5.

The biogenic tracer serves to highlight that the posterior estimate does not differentiate between direct emissions at the source and chemical formation inside a plume associated with those direct emissions. The biogenic emissions are in the gas phase, and the model obtains an estimate of OC concentrations that results from them. The same applies for the individual source categories. For example the 19 % of simulated impacts from the “other” category are the sum of both direct emissions and chemical formation resulting from those emissions. A finer grained study using an aerosol module would be required to deconvolve these two processes.

We used both Monte Carlo error propagation and bootstrapping to estimate the uncertainties in the emissions estimates. Figure 9 shows the histogram of total emissions for each of the main categories in the inversion, along with correlation scattergrams of the results for the bootstrapped simulations for EC. The standard deviation of the contributions is between 3 and 5 % of the mean contribution for all emission categories except for open burning where it is 20 %. There is little correlation in the emissions estimates from the different source groups. The highest  $r^2$  is 0.22 for realizations of the on-road and “other” emissions. Overall this suggests that our results are not excessively impacted by cross-correlation terms.

The results of the Monte Carlo error propagation are included in the Supplement. The uncertainties vary between 1.5 and 3 % except for open burning where they are 6 %. These are noticeably lower than the bootstrapping estimates as well as what we expect from knowing about emission inventories and from the values of the regularization parameters that were determined from the inversion themselves. These suggest that using block-bootstrapping provides a better estimate of the uncertainties.

The results for OC are included in the Supplement. The bootstrapped standard deviations are between 5 and 10 % of the mean contributions for all emission categories except for

open burning where they are 18 %. This suggests that the emissions estimates are robust with respect to uncertainties in the model inputs.

### 3.3 Inverse model results: temporal profiles

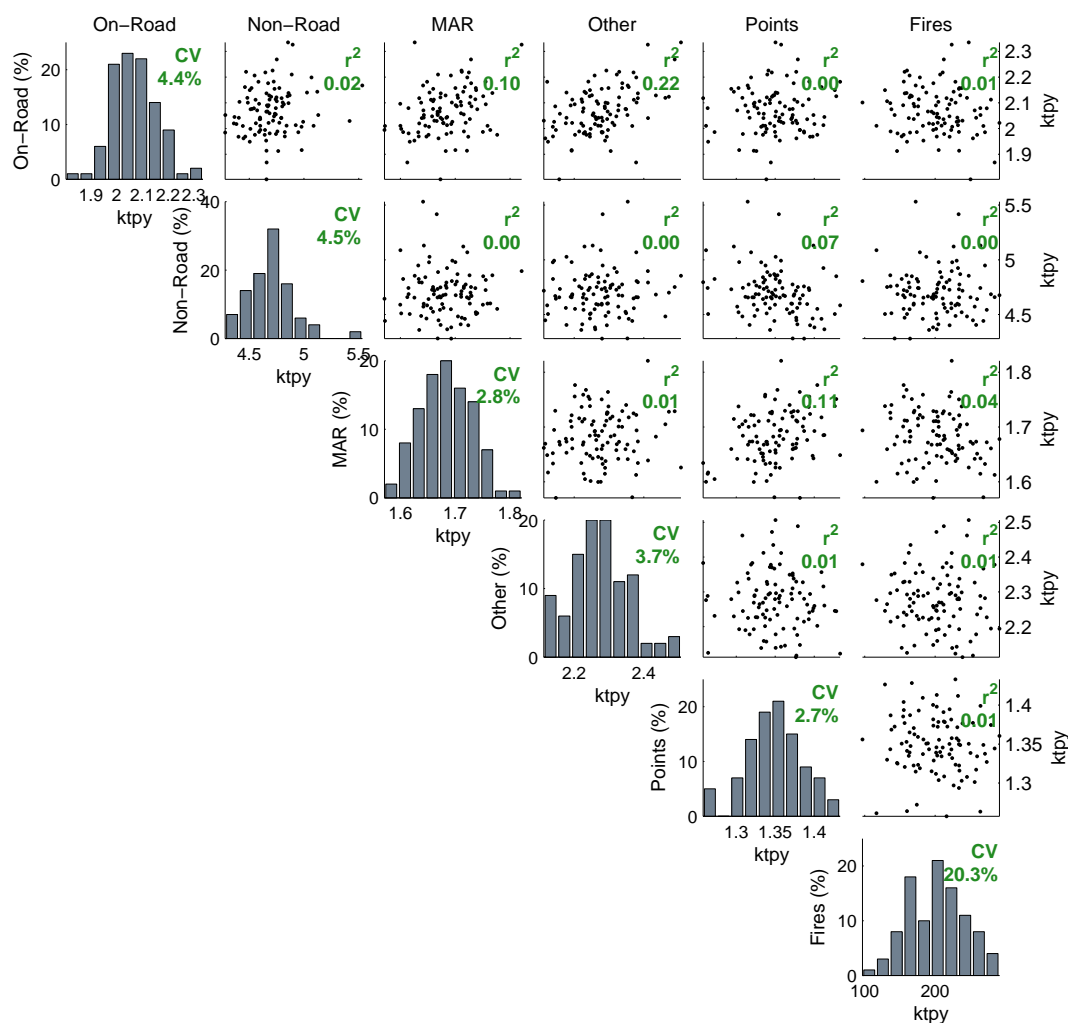
As described in Sect. 2.3, we performed the inversion using separate time series for five different time periods during the day, for weekdays and weekends, and for each month of the year. This led to  $5 \times 2 \times 12$  entries in the inverse algorithm for each of 5 source types. We now present the monthly variation and the diurnal variation for emissions of EC and OC for each of the source types for weekdays and for weekends (which include Saturdays, Sundays and holidays, SSH).

#### 3.3.1 On-road emissions

Figure 10 shows the monthly and diurnal temporal patterns for the on-road emissions. 90 % confidence intervals on the inverse model results are shown on the graphs. These were obtained from the bootstrapping which provides an estimate of the uncertainty due to episode selection and transport errors, as discussed in Sect. 2.4. On-road emissions are the category with the largest difference between inverse model results and the prior inventory.

In the prior for both EC and OC, weekday and weekend emissions are very similar, and there is only a slight annual variation from a maximum in the winter to a minimum in the summer months. The posterior levels for EC are similar to the emissions prior during the summer months for weekdays, but weekends are significantly lower. During fall and winter, the posterior emissions are very low, which is why the total emission levels shown in Table 1 went from 4300 tpy in the prior to 2100 tpy in the posterior. The monthly variation of the OC posterior is similar to the EC posterior although total OC emissions are left relatively unchanged at around 2500 tpy. The large reduction in emissions during fall and winter is unlikely to be realistic, even accounting for the fact that the measurements are from 2002 and the inventory for 2007, and so it suggests that there is an issue with the current representation of the emissions in the inventory and/or with the simulated wind transport from the sources to the receptor site.

The diurnal emissions profile of on-road EC shows a sharp increase starting at 6:00 a.m., and a peak at 3:00–4:00 p.m. followed by a gradual decline until midnight. There is a large contrast with the posterior. For weekdays EC follows the diurnal trend but has significantly lower emission levels, and has a strong reduction during the afternoon rush hour. For weekends, there is very little diurnal variation of emissions. The OC posteriors follow the diurnal profile of the priors much more closely, with slightly higher emissions during the day and lower emissions on weekends than in the prior. It would therefore seem that OC on-road emissions are better represented in the models than EC on-road emissions.



**Figure 9.** Bootstrapped estimates of uncertainties in inverse EC emissions by source group: histograms show the distribution of emission estimates, scatter plots show the cross-correlation of the estimates.  $CV = \sigma/\mu$  is the coefficient of variation.

Taken together, these results suggest that future research should seek to clarify the monthly profiles and the possibility of higher emissions during the summer rather than the winter. Furthermore, the posterior suggests that the diurnal profile could be improved as well as the difference between weekdays and weekends. It is possible that accuracy of the wind transport in the models is a function of the time of day, which could be a factor in the greater discrepancy between the prior and the posterior in the late afternoon. Finally, the large difference between the prior and the posterior could be the result of uncertainties in the current spatial distribution of the emissions.

### 3.3.2 Non-road emissions

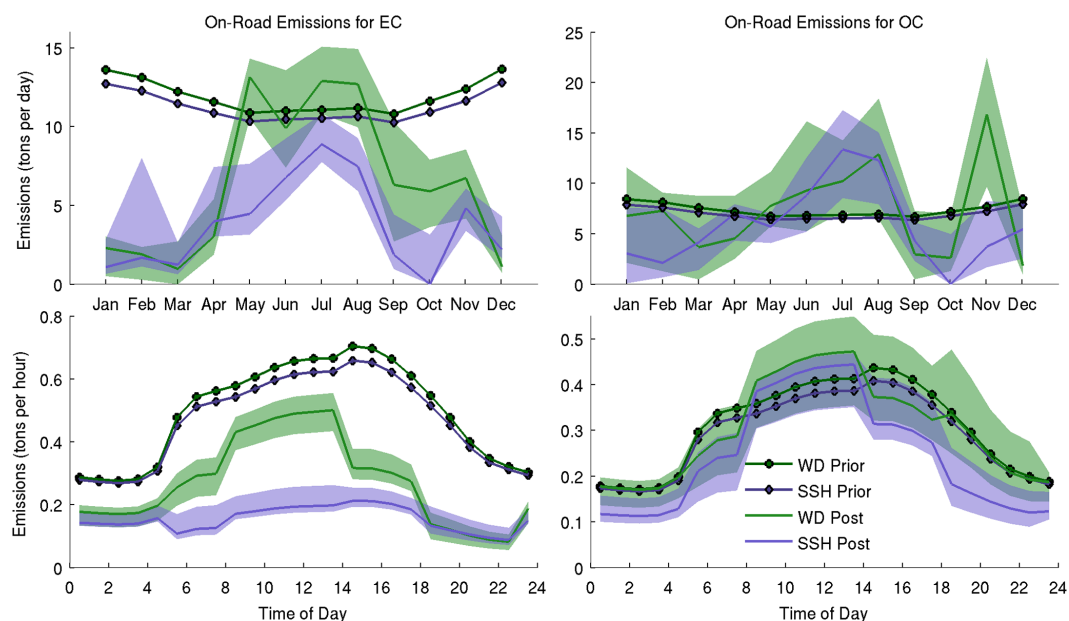
In contrast to the on-road emissions, the non-road posterior emissions follow the prior much more closely as can be seen in Fig. 11. There is a double peak, one in the early summer and a second one in the late fall. This confirms that simu-

lations can be improved by taking into account the spring and fall maximum of agricultural equipment as was done in the LADCO inventory, rather than using the default summer maximum in MOVES.

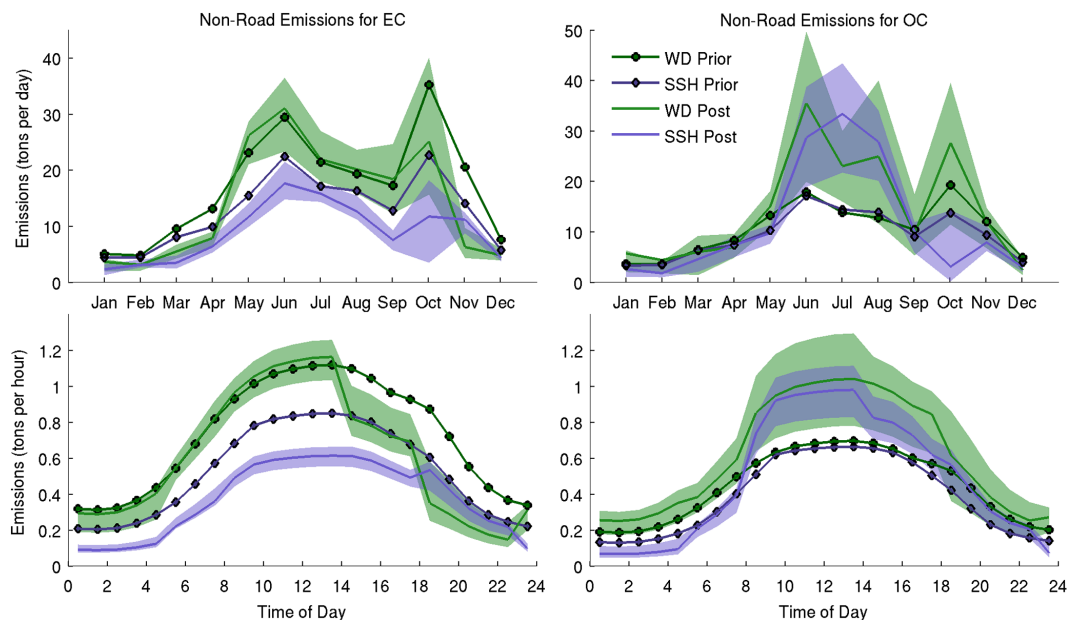
For EC, the model suggests that there is a greater decrease in emissions on weekends than is currently represented in the inventory. The diurnal profile of the posterior follows that of the prior more closely than for the on-road emissions, although there is again a sharp reduction of emissions in the posterior during the afternoon. The weekend emissions follow the diurnal profile, but are closer to 50 % lower than weekdays compared with 30 % lower in the priors.

For OC, the summer peak in the posterior is double that in the prior. We also see an enhancement of around 50 % during daylight hours. An estimate of 40 % of OC at the site being due to secondary formation (Bae et al., 2006) would account for most of the excess in OC, as discussed further in Sect. 4.





**Figure 10.** Monthly and diurnal temporal pattern of emissions of EC and OC for on-road emissions by weekday (green, WD) and weekend (blue, SSH) for St. Louis and the surrounding area. LADCO inventory results shown with solid symbols, Inverse model results shown with thin line. Shading shows the 90 % confidence interval in the inverse model results based on 100 bootstrapped inversions. Note that OC posterior totals combine primary emissions and secondary formation.

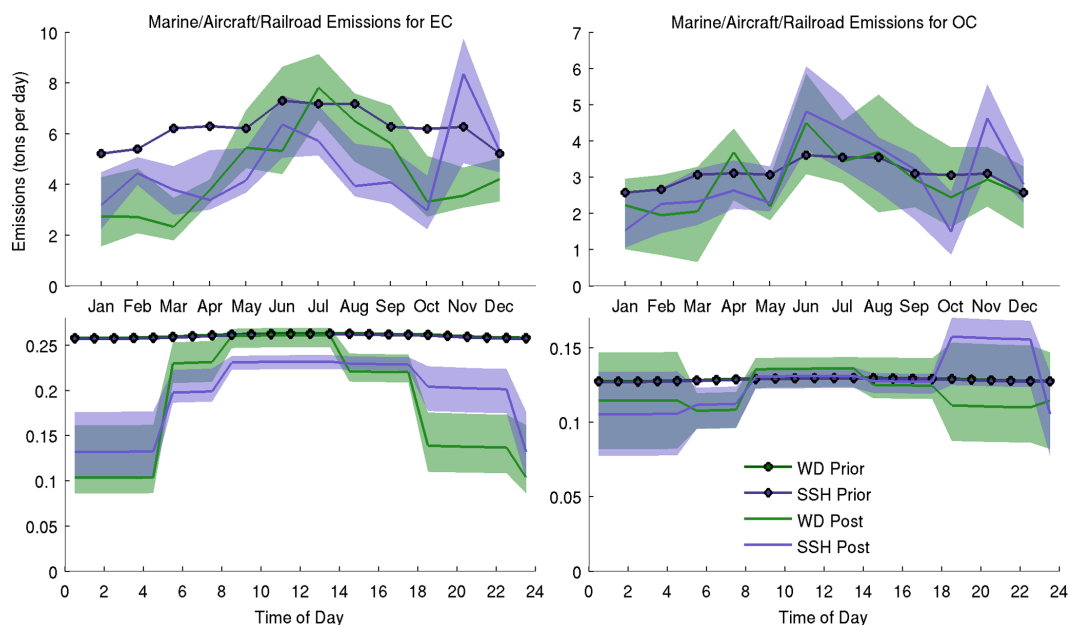


**Figure 11.** Monthly and diurnal temporal pattern of emissions of EC and OC for non-road emissions by weekday and weekend for the St. Louis region, see Fig. 10.

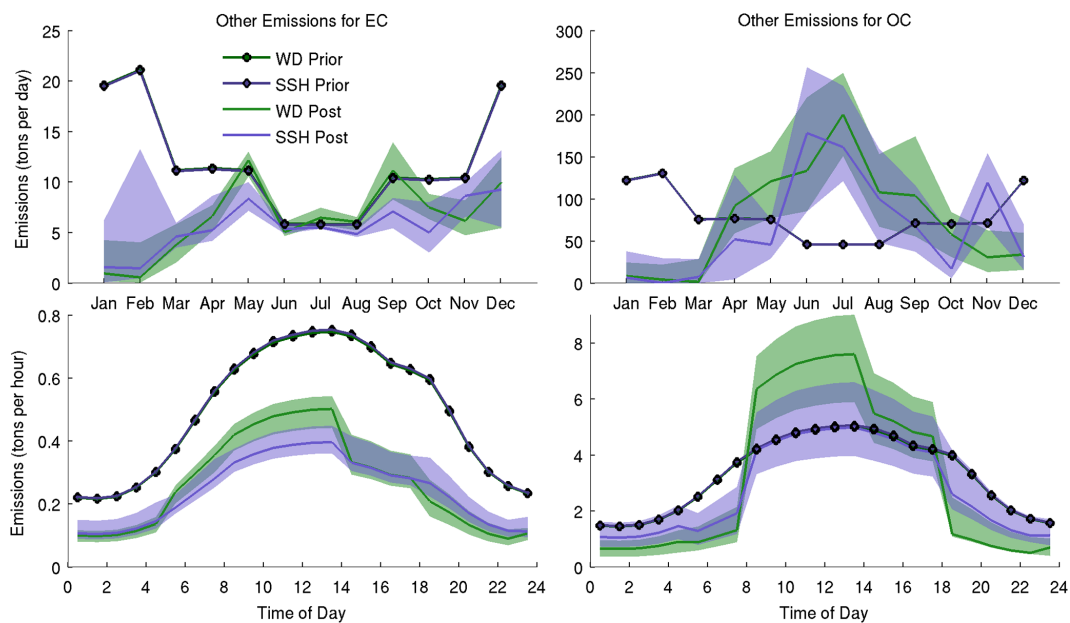
### 3.3.3 MAR emissions

The temporal profile of the MAR emissions (marine/aircraft/railroad) are shown in Fig. 12. In the prior, these are the same for weekdays and weekends and vary by 30 % throughout the year from a minimum in winter to a maxi-

imum in the summer. The posterior for EC is similar in this respect, but has a more pronounced annual variations with lower emissions in the winter months. There are differences between weekdays and weekends, but these are not systematic and could be the result of model uncertainty. The same is



**Figure 12.** Monthly and diurnal temporal pattern of emissions of EC and OC for marine/aircraft/railroad (MAR) emissions by weekday and weekend for the St. Louis region, see Fig. 10.



**Figure 13.** Monthly and diurnal temporal pattern of emissions of EC and OC for “other” emissions by weekday and weekend for the St. Louis region, see Fig. 10.

true for OC, although the levels of OC are higher in the summer which could be due to chemical formation, as discussed for non-road emissions above.

The diurnal profile is flat in the prior, but the posterior suggests that there is a definite diurnal profile with emissions of EC at night lower than daytime levels by up to 50%. There

is less difference in the OC profile, but it still suggests that the diurnal activity profile should be reconsidered.

### 3.3.4 Other emissions

Other emissions are shown in Fig. 13. In the prior the winter time emissions are three times those during the summer for both EC and OC. This is in stark contrast to the posterior

emissions. The inverse model finds that the EC concentrations at the receptor site are in good agreement with the emission patterns of spring through fall. No agreement is found, however, for the winter where the posterior estimate of both EC and OC emissions is nearly zero. For OC, the emissions are scaled up during the summer by a factor of 3 to 4, some of which is most likely due to chemical formation.

The diurnal profile of the “other” category follows those of the on-road emissions. For EC, the profile is similar although the emissions are much lower, and there is a reduction on weekends of morning emissions. For OC we see low posterior emissions at night and increased emissions during the day, as was the case for non-road emissions.

### 3.3.5 Point sources

Finally, we see the temporal profiles for point sources in Fig. 14. The monthly emissions in the prior vary from a low in the spring to a high in the fall with about 30 % changes in EC but only 15 % in OC. This is roughly reproduced in the posterior for EC albeit with a larger change from trough to peak. For OC, there are large swings in the emissions of the posterior. This suggests that there are large uncertainties in these estimates. From the perspective of the inverse model, it is a sign that there is poor agreement between the simulated and observed concentrations, but also that the estimates could be stabilized with more data, or with stronger constraints on the prior, or an improved model that considered in-plume chemistry.

The diurnal profile of the point sources is rather flat throughout the day in the prior. As for the MAR sources, the model suggests that there is a reduction in EC emissions between midnight and sunrise. There does also seem to be a slight reduction in EC emissions in the posterior on weekends compared with weekdays. The large swings in the estimates of monthly OC emissions mean that the diurnal profile should also be considered with caution. These swings are mostly contained within the 90 % confidence range displayed in the figure which suggests that they are not statistically significant. At a minimum, we can say that EC emissions from point sources seem to be reliably characterized in the inventory and the model, but that more research is needed for the OC impacts.

### 3.3.6 Uncertainty due to mixing heights

As discussed in Sect. 3.1, the WRF simulations do not have systematic errors for temperature, humidity, wind speed and direction at the surface. However, we do not have measurements of the mixing heights which could be used to evaluate errors in the vertical mixing in the model. In particular, these could contain systematic errors as a function of the time of day which would impact the diurnal profiles estimated by the inverse model. de Foy et al. (2007) found that the choice of the vertical mixing scheme in CAMx could have a significant

impact on the estimation of emissions in Mexico City. This remains a source of uncertainty in the present analysis which could be constrained in future studies if more detailed measurements of the vertical structure of wind transport in the atmosphere became available. Alternatively, the uncertainty could be estimated by running the inverse model with different sets of WRF simulations that used different options, for example by generating input meteorological fields with different boundary layer schemes.

## 3.4 Inverse model results: open burning

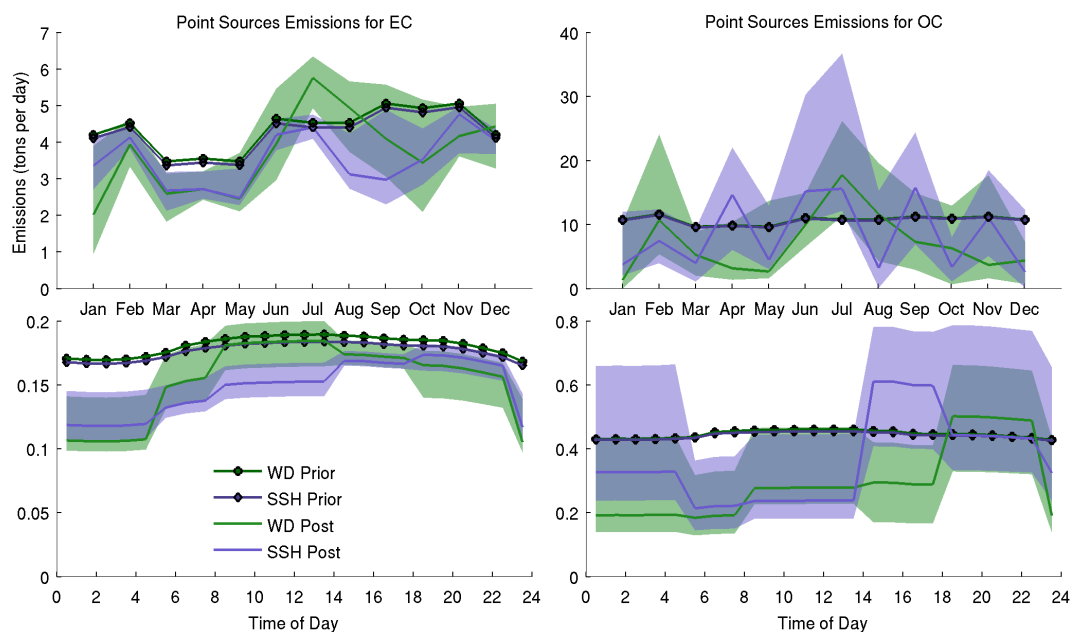
Section 3.2 showed that using emissions from FINN as the prior for CAMx simulations of open burning led to impacts of 1 % of EC and 3 % of OC. The posterior impacts were increased to 4 % for EC and 5 % for OC. Table 2 shows the emission totals by geographic sector in metric tonnes per year for the prior and for the posterior. For the Local sector (within 100 km of the receptor), the northeast sector and the southeast sector, the inverse model increases the emissions by a factor of around 30 for EC and around 20 for OC. Emissions from the southwest sector are increased by a factor of 3 for EC and by a factor of 2 for OC. The open burning emissions from the west were kept at a similar level in the posterior as in the prior. The emissions from the northwest did not match the data and were set to 0 in the posterior by the inversion.

Table 2 also shows the posterior impact fractions by sector. The largest contributions are 1.4 % of EC and 2.5 % of OC from the southeast sector, followed by the southwest and the west sector. Local fires account for 0.7 % of EC and 0.5 % of OC in the posterior.

As discussed in Sect. 2.2, the emissions in FINN are based only on the Terra MODIS sensor, as the Aqua satellite was not yet in orbit in 2002. This means that the uncertainties in these emissions are greater than those following the launch of Aqua where there is twice as much satellite data available for fire detection (Hawbaker et al., 2008). In addition to missing fires, there are uncertainties in the estimates of area burned and of the type and amount of vegetation burned. As shown in Fig. 9, uncertainty estimates based on bootstrapping are largest for open burning, with 20 %. However, adjustment factors of 20 to 30 suggest either that the uncertainties are underestimated, or that the inversion of these emissions are underconstrained. Overall, these results suggest that future work with more surface measurements and emissions estimates from more recent satellite sensors are needed to improve the inverse estimates, but that nonetheless emission factors in FINN should be revised upwards.

## 3.5 Inverse model results: residence-time-analysis impacts

The inverse model combines emission estimates using Eulerian (CAMx) and Lagrangian (FLEXPART-WRF) simula-



**Figure 14.** Monthly and diurnal temporal pattern of emissions of EC and OC for point source emissions by weekday and weekend for the St. Louis region, see Fig. 10.

tions. Polar grids of residence time analysis calculated using back trajectories are used to estimate emission sources that could be missing in the LADCO emissions inventory. The polar gridded emissions have zero prior and represent a way of decomposing the residual between the CAMx posterior and the measurements into a spatial emission signal. The inverse model includes separate grids for 3:00 to 9:00 a.m., 9:00 a.m. to 3:00 p.m., 3:00 to 9:00 p.m. and 9:00 p.m. to 3:00 a.m., as well as for weekdays and weekends, for a total of eight grids.

Note that the FLEXPART-WRF simulations do not include deposition, and that secondary OC formation is not included either. Both of these limitations would impact the estimation of actual emission amounts from the inverse model. In this section, we therefore report only impacts of different source regions on concentrations at the measurement site, which are not affected by deposition and include estimated impacts of both primary emissions and in-plume secondary formation. As will be discussed in Sect. 4, deposition is estimated to account for a 4 % loss of EC, and secondary formation is estimated to account for around 40 % of OC.

Figure 15 shows the sum of impacts from the eight grids for both EC and OC. As shown in Fig. 8, these account for 33 % of the EC posterior time series and 46 % of the OC posterior time series. The main signal is from the south, and especially the southwest for both EC and OC, indicating that these could be areas to be explored for updating the spatial distribution of emissions.

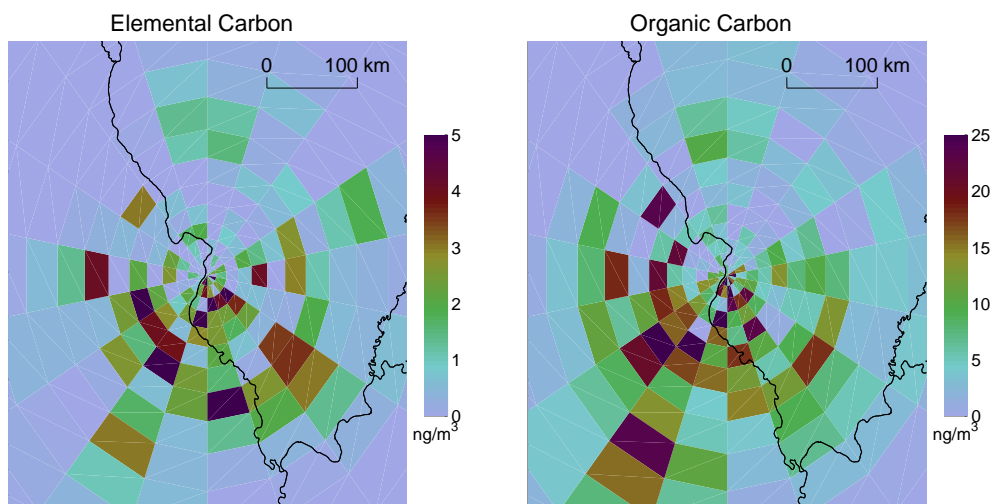
Figure 16 shows the total contribution to the average concentration for EC and OC for each of the RTA grids. For

EC, the contribution varies from a minimum of  $0.05 \text{ ng m}^{-3}$  to just above  $0.25 \text{ ng m}^{-3}$ . The contribution from the early morning to afternoon (3:00 a.m. to 3:00 p.m.) are lower than those for the late afternoon and nighttime (3:00 p.m. to 3:00 a.m.). The weekdays and weekends have a similar trends, but the diurnal variation is more pronounced on weekends. For OC, there is a similar pattern with lower contributions from 3:00 a.m. to 3:00 p.m., of around  $0.8 \text{ ng m}^{-3}$  rising to around  $1.5 \text{ ng m}^{-3}$  in the nighttime. Weekdays and weekends RTA impacts are more similar for OC than they are for EC.

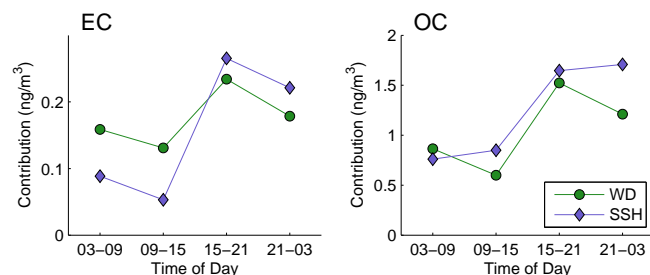
### 3.6 Inverse model results: emission totals

In this section we compare the emissions in metric tonnes per year of the different source types from the inverse model with the NEI 2008 and the LADCO inventory. Table 1 and Fig. 17 show the annual total emissions for the St. Louis Regional domain for the 2008 National Emissions Inventory, the 2007 LADCO inventory used as a prior, and for the posterior.

Overall, the LADCO inventory is slightly larger than the NEI for both EC and OC. For EC, the on-road emissions are 50 % larger, and the MAR emissions are 25 % larger while the remaining categories are similar. For OC, the largest category by far in both inventories are the “other” sources which are 17 % higher in the LADCO inventory. These include residential wood and waste combustion, non-vehicle road emissions and food cooking (estimates of agricultural burning are high in the NEI but low in the LADCO inventory). OC emissions from on-road, non-road, MAR and point sources are



**Figure 15.** Contributions to the average 2002 concentration of EC and OC in the inverse time series from the residence time analysis grids.



**Figure 16.** Total contribution to the average concentration of EC and OC in the inverse time series from the residence time analysis grids by time of day for weekdays (WD) and weekends (SSH).

all increased by up to a factor of 2 in the LADCO inventory compared with the NEI.

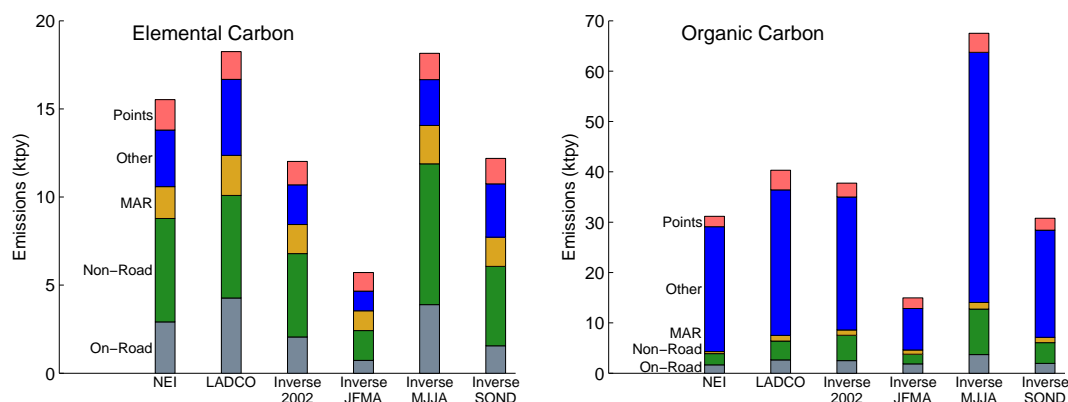
The posterior emissions are calculated from the model as departures from the LADCO prior. As discussed in Sect. 3.2, the simulated EC concentrations were too high at the site, and so the inverse model has lower emissions for all categories. The EC emission estimates from both on-road sources and “other” sources are reduced by 50% in the prior, whereas the remaining categories are only slightly reduced. For OC, there is only a slight reduction in the total emissions with a small shift in emissions from the “other” category into non-road emissions. This suggests that the inverse results are in agreement with the inventory, bearing in mind that the model does not distinguish between primary and secondary OC. As around 40% of OC is estimated to be secondary (see Sect. 4), this is a significant source of uncertainty. Nevertheless, comparison with Fig. 8 would suggest that the secondary OC is represented in the model more by the polar grid emissions or as missing carbon rather than as adjustments to the known sources.

Also shown in Fig. 17 are emissions for three time periods during the year that correspond to a natural grouping in the data: January to April, May to August and September to December. The emissions rates are annualized by multiplying the emissions in tonnes per 4 months by 3 in order to have emissions in tonnes per year. This yields the annual emission rate that would be obtained if the emissions of the 4 months continue for an entire year. Compared with the LADCO inventory, the emissions estimates are low for January–April, high for May–August and similar for September–December. This shows that there is uncertainty in the model results that depends on the time of year and that in particular simulations are in greater disagreement with the inventories for January to April. Although the variation exists for both EC and OC, it is stronger for OC because May–August is when there is the most secondary OC formation (see Sect. 4). At this stage it is not possible to say what part of this is due to limitations in the inventories, what part to measurements and especially what part due to modeling errors. Further research with more sites and longer time series would be able to better constrain the estimates.

#### 4 Conclusions

A least squares inverse model was used to estimate emissions of elemental carbon and organic carbon using hourly data for 2002 from the St. Louis–Midwest supersite, and uncertainty estimates were obtained by running the model multiple times using block-bootstrapping. The model provided information on the diurnal pattern of the emissions, the difference between weekdays and weekends and the annual variation on a month by month basis. The inversion was based on the 2007 LADCO inventory for the following source types: on-road, non-road, marine/aircraft/railroad (MAR), “other” and point sources.





**Figure 17.** Emissions of EC and OC in the Regional domain by source type for the 2008 NEI, the 2007 LADCO inventory and the posterior estimate based on using LADCO as a prior. Inverse results are shown for the entire year (2002), along with annualized emissions for January–April (JFMA), May–August (MJJA) and September–December (SOND).

There are two important limitations in our modeling. The first is that we do not include deposition in the FLEXPART back trajectories. This means that we cannot obtain emissions directly from the residence time analysis grids but instead we obtain results for the contributions of sources towards EC or OC concentrations at the measurement site. For EC, which is a passive tracer, we performed a sensitivity test on the impact of deposition using forward simulations with CAMx. The emissions based on the FLEXPART inversion were used as input into CAMx and two sets of simulations were performed: one set without deposition, and a second set with both wet and dry deposition. Wet and dry deposition in the model reduced the EC concentration at the site by 4 % on average over the whole year. The main reason this number is low is that most of the impacts are due to fairly local emissions (within 100 to 200 km). Overall, this shows that neglecting deposition in FLEXPART has a minor impact on the results.

The second limitation in our modeling is that we do not include secondary formation of OC. There is considerable formation of OC in the atmosphere (Jimenez et al., 2009; Robinson et al., 2007) and also significant uncertainties in simulations of secondary organic aerosols (Napelenok et al., 2014). These uncertainties include the complex behavior of semi-volatile and intermediate volatility organic precursors involving the evaporation of primary OC and recondensation after oxidation (Hodzic et al., 2010). In our inverse model, the OC emission results need to be interpreted as the combination of emissions and in-plume formation. To estimate the contribution of secondary formation to OC in our time series, we consider three lines of evidence. First, using the same data as our paper, Bae et al. (2006) estimate that 20 to 40 % of OC at the St. Louis–Midwest supersite was secondary organic aerosol on an annual average (see their Fig. 2). Second, Fig. 8 shows that the primary OC simulated using the LADCO inventory ( $2 \mu\text{g m}^{-3}$ ) is 40 % lower than the average OC at the measurement site ( $3.5 \mu\text{g m}^{-3}$ ). It would be reason-

able to expect that a significant fraction of this 40 % is due to secondary formation; whereas the average concentration of OC is significantly higher than the primary OC contribution from the LADCO inventory, the reverse is true for EC where the average concentration of EC is lower than the primary EC contribution. Third, Fig. 13 shows large excess peaks of OC in the summer and during the daytime which can be interpreted as consisting mainly of secondary organic aerosol. We further note that the seasonality in Fig. 13 (minimal secondary OC in the winter increasing to a majority of OC in the summer), is similar to the seasonality shown in Fig. 2 in Bae et al. (2006). Overall, these three items suggest that 40 % would be a reasonable estimate of the OC that could be due to secondary formation in the atmosphere. Consequently, the OC emissions estimates such as in Table 1 should be interpreted as being the sum of somewhat over half of primary emissions ( $\sim 60\%$ ) and a little under half of secondary formation ( $\sim 40\%$ ).

The inverse emission estimates were in agreement with the LADCO inventory for most of the source types, with a slight downward revision of the emission totals. The main discrepancies suggested by the model are as follows: (1) on-road emissions were poorly represented during the winter and on weekends. Although the results for winter remain as an outstanding question, there is a clear need to update the diurnal profile for weekends. (2) Non-road emissions need to account for actual use of agricultural equipment, which was done by LADCO but is not carried out by default in MOVES. (3) MAR and point sources do not at present have much diurnal variation in the emissions. Although their diurnal profiles are smoother than on-road and non-road emissions, the model suggests that there is a discernible drop in nighttime emissions. (4) Other emissions from the inverse model matched the inventory during the summer but not during the winter. As with on-road emissions, more research is required to constrain the sources of the discrepancy and to improve the simulations of these impacts.

In addition to these findings, the inverse model identified impacts from open burning at the measurement site, and suggests that emissions of EC and OC should be increased in the FINN model.

Finally, gridded back trajectories suggest that most of the impacts missing from the emission inventories are due to transport from the quadrants southeast and southwest of the measurement site. The contributions to the average EC and OC concentrations at the measurement site from these sources are approximately twice as large during the late afternoon and early nighttime (3:00 p.m. to 3:00 a.m.) as they are earlier in the day (3:00 a.m. to 3:00 p.m.).

**The Supplement related to this article is available online at doi:10.5194/acp-15-2405-2015-supplement.**

*Acknowledgements.* The United States Environmental Protection Agency (EPA) funded the EC and OC measurements used in this analysis through cooperative agreement R-82805901-0, and the analysis through grant number RD-83455701. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the EPA. Further, the EPA does not endorse the purchase of any commercial products or services mentioned in the publication. We thank the staff of the St. Louis–Midwest fine-particle supersite for their assistance in data collection. We are also grateful to the US EPA for making the National Emissions Inventory available, and to the US National Climatic Data Center for the meteorological data. We wish to thank the three anonymous reviewers for their thoughtful and careful reviews which helped improve the paper.

Edited by: K. Carslaw

## References

- Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crouse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, *Atmos. Chem. Phys.*, 11, 4039–4072, doi:10.5194/acp-11-4039-2011, 2011.
- Amato, F. and Hopke, P. K.: Source apportionment of the ambient PM<sub>2.5</sub> across St. Louis using constrained positive matrix factorization, *Atmos. Environ.*, 46, 329–337, 2012.
- Ashbaugh, L. L., Malm, W. C., and Sadeh, W. Z.: A residence time probability analysis of sulfur concentrations at grand-canyon-national-park, *Atmos. Environ.*, 19, 1263–1270, 1985.
- Aster, R. C., Borchers, B., and Thurber, C. H.: *Parameter Estimation and Inverse Problems*, Academic Press, Oxford, UK, 2. edn., 2012.
- Austin, E., Coull, B. A., Zanobetti, A., and Koutrakis, P.: A framework to spatially cluster air pollution monitoring sites in US based on the PM<sub>2.5</sub> composition, *Environ. Int.*, 59, 244–254, 2013.
- Bae, M. S., Schauer, J. J., DeMinter, J. T., and Turner, J. R.: Hourly and daily patterns of particle-phase organic and elemental carbon concentrations in the urban atmosphere, *J. Air Waste Manage.*, 54, 823–833, 2004a.
- Bae, M. S., Schauer, J. J., DeMinter, J. T., Turner, J. R., Smith, D., and Cary, R. A.: Validation of a semi-continuous instrument for elemental carbon and organic carbon using a thermal-optical method, *Atmos. Environ.*, 38, 2885–2893, 2004b.
- Bae, M. S., Schauer, J. J., and Turner, J. R.: Estimation of the monthly average ratios of organic mass to organic carbon for fine particulate matter at an urban site, *Aerosol Sci. Technol.*, 40, 1123–1139, 2006.
- Bell, M. L., Ebisu, K., Peng, R. D., Samet, J. M., and Dominici, F.: Hospital Admissions and Chemical Composition of Fine Particle Air Pollution, *American Journal of Respiratory and Critical Care Medicine*, 179, 1115–1120, 2009.
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kaercher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, *J. Geophys. Res.-Atmos.*, 118, 5380–5552, 2013.
- Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollock, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T., and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NO<sub>x</sub> and CO<sub>2</sub> and their impacts, *Atmos. Chem. Phys.*, 13, 3661–3677, doi:10.5194/acp-13-3661-2013, 2013a.
- Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., Angevine, W., Evan, S., Dingwell, A., Fast, J. D., Easter, R. C., Pisso, I., Burkhardt, J., and Wotawa, G.: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1, *Geosci. Model Dev.*, 6, 1889–1904, doi:10.5194/gmd-6-1889-2013, 2013b.
- Brioude, J., Kim, S.-W., Angevine, W. M., Frost, G. J., Lee, S.-H., McKeen, S. A., Trainer, M., Fehsenfeld, F. C., Holloway, J. S., Ryerson, T. B., Williams, E. J., Petron, G., and Fast, J. D.: Top-down estimate of anthropogenic emission inventories and their interannual variability in Houston using a mesoscale inverse modeling technique, *J. Geophys. Res.-Atmos.*, 116, D20305, doi:10.1029/2011JD016215, 2011.
- Brunner, D., Henne, S., Keller, C. A., Reimann, S., Vollmer, M. K., O'Doherty, S., and Maione, M.: An extended Kalman-filter for regional scale inverse emission estimation, *Atmos. Chem. Phys.*, 12, 3455–3478, doi:10.5194/acp-12-3455-2012, 2012.
- Carroll, M., Townshend, J., Hansen, M., DiMiceli, C., Sohlberg, R., and Wurster, K.: MODIS Vegetative Cover Conversion and Vegetation Continuous Fields, in: *Land Remote Sensing and Global Environmental Change: NASA's Earth Observing System and the Science of Aster and MODIS*, edited by: Ramachandran, B., Justice, C. O., and Abrams, M., 725–746, Springer-Verlag, New York, NY, 2011.

- de Foy, B., Lei, W., Zavala, M., Volkamer, R., Samuelsson, J., Melqvist, J., Galle, B., Martínez, A.-P., Grutter, M., Retama, A., and Molina, L. T.: Modelling constraints on the emission inventory and on vertical dispersion for CO and SO<sub>2</sub> in the Mexico City Metropolitan Area using Solar FTIR and zenith sky UV spectroscopy, *Atmos. Chem. Phys.*, 7, 781–801, doi:10.5194/acp-7-781-2007, 2007.
- de Foy, B., Zavala, M., Bei, N., and Molina, L. T.: Evaluation of WRF mesoscale simulations and particle trajectory analysis for the MILAGRO field campaign, *Atmos. Chem. Phys.*, 9, 4419–4438, doi:10.5194/acp-9-4419-2009, 2009.
- de Foy, B., Smyth, A. M., Thompson, S. L., Gross, D. S., Olson, M. R., Sager, N., and Schauer, J. J.: Sources of nickel, vanadium and black carbon in aerosols in Milwaukee, *Atmos. Environ.*, 59, 294–301, 2012a.
- de Foy, B., Wiedinmyer, C., and Schauer, J. J.: Estimation of mercury emissions from forest fires, lakes, regional and local sources using measurements in Milwaukee and an inverse method, *Atmos. Chem. Phys.*, 12, 8993–9011, doi:10.5194/acp-12-8993-2012, 2012b.
- de Foy, B., Heo, J., and Schauer, J. J.: Estimation of direct emissions and atmospheric processing of reactive mercury using inverse modeling, *Atmos. Environ.*, 85, 73–82, 2014.
- Edick, S. and Janssen, M.: Temporally Allocating Emissions with CEM Data for Chemical Transport and SIP Modeling, in: 15th International Emission Inventory Conference “Reinventing Inventories - New Ideas in New Orleans”, edited by: Solomon, D., United States Environmental Protection Agency, New Orleans, LA, 2006.
- Efron, B.: Bayes’ theorem in the 21st century, *Science*, 340, 1177–1178, 2013.
- ENVIRON: CAMx User’s Guide, Comprehensive Air quality Model with eXtensions, Tech. Rep. Version 6.0, ENVIRON International Corporation, 2013.
- EPA, U.: Motor Vehicle Emissions Simulator (MOVES), Tech. Rep. EPA-420-B-12-001b, US EPA, Research Triangle Park, NC, 2012.
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sens. Environ.*, 114, 168–182, 2010.
- Gentner, D. R., Isaacman, G., Worton, D. R., Chan, A. W., Dallmann, T. R., Davis, L., Liu, S., Day, D. A., Russell, L. M., Wilson, K. R., Weber, R., Guha, A., Harley, R. A., and Goldstein, A. H.: Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions, *P. Natl. Acad. Sci. USA*, 109, 18318–18323, 2012.
- Giglio, L., Descloitres, J., Justice, C., and Kaufman, Y.: An enhanced contextual fire detection algorithm for MODIS, *Remote Sens. Environ.*, 87, 273–282, 2003.
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, 6, 3181–3210, doi:10.5194/acp-6-3181-2006, 2006.
- Hand, J. L., Schichtel, B. A., Pitchford, M., Malm, W. C., and Frank, N. H.: Seasonal composition of remote and urban fine particulate matter in the United States, *J. Geophys. Res.-Atmos.*, 117, D05209, doi:10.1029/2011JD017122, 2012.
- Hand, J. L., Schichtel, B. A., Malm, W. C., and Frank, N. H.: Spatial and Temporal Trends in PM<sub>2.5</sub> Organic and Elemental Carbon across the United States, *Advances in Meteorology*, 367674, doi:10.1155/2013/367674, 2013.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R. A.: Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm, *Earth Interact.*, 7, 1–15, 2003.
- Hansen, M. C., Townshend, J. R. G., Defries, R. S., and Carroll, M.: Estimation of tree cover using MODIS data at global, continental and regional/local scales, *Int. J. Remote Sens.*, 26, 4359–4380, 2005.
- Hawbaker, T. J., Radeloff, V. C., Syphard, A. D., Zhu, Z., and Stewart, S. I.: Detection rates of the MODIS active fire product in the United States, *Remote Sens. Environ.*, 112, 2656–2664, 2008.
- Henze, D. K., Seinfeld, J. H., and Shindell, D. T.: Inverse modeling and mapping US air quality influences of inorganic PM<sub>2.5</sub> precursor emissions using the adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, 9, 5877–5903, doi:10.5194/acp-9-5877-2009, 2009.
- Hodzic, A., Jimenez, J. L., Madronich, S., Canagaratna, M. R., DeCarlo, P. F., Kleinman, L., and Fast, J.: Modeling organic aerosols in a megacity: potential contribution of semi-volatile and intermediate volatility primary organic compounds to secondary organic aerosol formation, *Atmos. Chem. Phys.*, 10, 5491–5514, doi:10.5194/acp-10-5491-2010, 2010.
- Hoelzemann, J., Schultz, M., Brasseur, G., Granier, C., and Simon, M.: Global Wildland Fire Emission Model (GWEM): Evaluating the use of global area burnt satellite data, *J. Geophys. Res.-Atmos.*, 109, D14S04, doi:10.1029/2003JD003666, 2004.
- Jaekels, J. M., Bae, M.-S., and Schauer, J. J.: Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols, *Environ. Sci. Technol.*, 41, 5763–5769, 2007.
- Janssen, N. A. H., Hoek, G., Simic-Lawson, M., Fischer, P., van Bree, L., ten Brink, H., Keuken, M., Atkinson, R. W., Anderson, H. R., Brunekreef, B., and Cassee, F. R.: Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM<sub>10</sub> and PM<sub>2.5</sub>, *Environ. Health Persp.*, 119, 1691–1699, 2011.
- Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of Organic Aerosols in the Atmosphere, *Science*, 326, 1525–1529, 2009.

- LADCO: Regional Air Quality Analyses for Ozone, PM<sub>2.5</sub>, and Regional Haze: Base C Emissions Inventory, Tech. Rep. September 12, 2011, Lake Michigan Air Directors Consortium, Rosemont, IL, 2011.
- Lee, J. H. and Hopke, P. K.: Apportioning sources of PM<sub>2.5</sub> in St. Louis, MO using speciation trends network data, *Atmos. Environ.*, 40, S360–S377, 2006.
- Lee, J. H., Hopke, P. K., and Turner, J. R.: Source identification of airborne PM<sub>2.5</sub> at the St. Louis-Midwest Supersite, *J. Geophys. Res.-Atmos.*, 111, D10S10, doi:10.1029/2005JD006329, 2006.
- Lewandowski, M., Jaoui, M., Offenberg, J. H., Kleindienst, T. E., Edney, E. O., Sheesley, R. J., and Schauer, J. J.: Primary and secondary contributions to ambient PM in the midwestern United States, *Environ. Sci. Technol.*, 42, 3303–3309, 2008.
- Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J. Geophys. Res.-Atmos.*, 108, 4493, doi:10.1029/2002JD003161, 2003.
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, *B. Am. Meteorol. Soc.*, 87, 343–360, 2006.
- Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, *Geosci. Model Dev.*, 7, 303–315, doi:10.5194/gmd-7-303-2014, 2014.
- Napelenok, S. L., Simon, H., Bhave, P. V., Pye, H. O. T., Pouliot, G. A., Sheesley, R. J., and Schauer, J. J.: Diagnostic Air Quality Model Evaluation of Source-Specific Primary and Secondary Fine Particulate Carbon, *Environ. Sci. Technol.*, 48, 464–473, 2014.
- Ramanathan, V. and Carmichael, G.: Global and regional climate changes due to black carbon, *Nature Geosci.*, 1, 221–227, 2008.
- Reche, C., Querol, X., Alastuey, A., Viana, M., Pey, J., Moreno, T., Rodríguez, S., González, Y., Fernández-Camacho, R., de la Rosa, J., Dall'Osto, M., Prévôt, A. S. H., Hueglin, C., Harrison, R. M., and Quincey, P.: New considerations for PM, Black Carbon and particle number concentration for air quality monitoring across different European cities, *Atmos. Chem. Phys.*, 11, 6207–6227, doi:10.5194/acp-11-6207-2011, 2011.
- Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, *Atmos. Chem. Phys.*, 11, 9887–9898, doi:10.5194/acp-11-9887-2011, 2011.
- Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking organic aerosols: Semivolatile emissions and photochemical aging, *Science*, 315, 1259–1262, 2007.
- Rödenbeck, C., Gerbig, C., Trusilova, K., and Heimann, M.: A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, *Atmos. Chem. Phys.*, 9, 5331–5342, doi:10.5194/acp-9-5331-2009, 2009.
- Rohr, A. C. and Wyzga, R. E.: Attributing health effects to individual particulate matter constituents, *Atmos. Environ.*, 62, 130–152, 2012.
- Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmos. Chem. Phys.*, 4, 51–63, doi:10.5194/acp-4-51-2004, 2004.
- Seibert, P., Kromp-Kolb, H., Baltensperger, U., Jost, D. T., and Schwikowski, M.: Trajectory analysis of high-alpine air pollution data, in: *Air Pollution Modelling and its Application X*, edited by: Gryning, S.-E. and Millan, M. M., 595–596, Plenum Press, New York, 1994.
- Sheesley, R. J., Schauer, J. J., Meiritz, M., DeMinter, J. T., Bae, M.-S., and Turner, J. R.: Daily variation in particle-phase source tracers in an urban atmosphere, *Aerosol Sci. Technol.*, 41, 981–993, 2007.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 2, Tech. Rep. NCAR/TN-468+STR, NCAR, Boulder, CO, 2005.
- Snyder, D. C., Schauer, J. J., Gross, D. S., and Turner, J. R.: Estimating the contribution of point sources to atmospheric metals using single-particle mass spectrometry, *Atmos. Environ.*, 43, 4033–4042, 2009.
- Snyder, D. C., Rutter, A. P., Worley, C., Olson, M., Plourde, A., Bader, R. C., Dallmann, T., and Schauer, J. J.: Spatial variability of carbonaceous aerosols and associated source tracers in two cities in the Midwestern United States, *Atmos. Environ.*, 44, 1597–1608, 2010.
- Spak, S. N. and Holloway, T.: Seasonality of speciated aerosol transport over the Great Lakes region, *J. Geophys. Res.-Atmos.*, 114, D08302, doi:10.1029/2008JD010598, 2009.
- Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmos. Chem. Phys.*, 5, 2461–2474, doi:10.5194/acp-5-2461-2005, 2005.
- Stohl, A., Seibert, P., Arduini, J., Eckhardt, S., Fraser, P., Grealley, B. R., Lunder, C., Maione, M., Mühle, J., O'Doherty, S., Prinn, R. G., Reimann, S., Saito, T., Schmidbauer, N., Simmonds, P. G., Vollmer, M. K., Weiss, R. F., and Yokouchi, Y.: An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons, *Atmos. Chem. Phys.*, 9, 1597–1620, doi:10.5194/acp-9-1597-2009, 2009.
- Sullivan, A. P., Weber, R. J., Clements, A. L., Turner, J. R., Bae, M. S., and Schauer, J. J.: A method for on-line measurement of water-soluble organic carbon in ambient aerosol particles: Results from an urban site, *Geophys. Res. Lett.*, 31, L13105, doi:10.1029/2004GL019681, 2004.
- Thesing, K. B., Bollman, A. D., Strait, R., VanBruggen, J., and Janssen, M.: Improvements to nonroad model inputs for midwestern states, in: *13th International Emission Inventory Conference "Working for Clean Air in Clearwater"*, edited by: Lorang, P. A., United States Environmental Protection Agency, Clearwater, FL, 2004.
- Wang, G., Hopke, P. K., and Turner, J. R.: Using highly time resolved fine particulate compositions to find particle sources in St. Louis, MO, *Atmospheric Pollution Research*, 2, 219–230, 2011.
- Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, *Geosci. Model Dev.*, 4, 625–641, doi:10.5194/gmd-4-625-2011, 2011.

Wunsch, C.: Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications, Cambridge University Press, Cambridge, UK, p. 43, 2006.

Zhang, L., Brook, J. R., and Vet, R.: A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, 3, 2067–2082, doi:10.5194/acp-3-2067-2003, 2003.