Supporting Information

Winiger et al. 10.1073/pnas.1613401114

SI Methods

Carbon Isotope End-Member Determination. The end-member values (source-specific signatures) of the carbon isotopes were compiled from the literature (Table S2). The fossil $\Delta^{14}C$ sources (liquid fossil, coal, and gas flaring) are well constrained at −1,000 ± 0‰. The main source of biomass in Russia is wood burning, estimated at +225 ± 60‰ (13, 31). The $\delta^{13}C$ end-member values for coal (−23.4 ± 1.3‰) and biomass (C3 plants, −26.7 ± 1.8‰) were collected from the literature review conducted by Andersson et al. (28). The liquid fossil end-member was taken from typical liquid fossil sources used in Russia (−31.4 ± 1‰) (40). The perhaps most uncertain end-member is the $\delta^{13}C$ value for gas flaring. Here a wide variability of 3‰ is used, but the average position may be dependent on the relative contribution from the different gas components, e.g., methane, ethane, and propane, which all have quite varying $\delta^{13}C$-value signatures, although methane is expected to dominate the mixtures ($\delta^{13}C$ for methane ≈ –60‰). Notable is that BC formation, in general, is associated with an enrichment in the $\delta^{13}C$ isotope (higher $\delta^{13}C$), because the lighter ^{12}C is more prone to CO_2 formation. However, an enrichment of more than a few per mill is not expected. In general, $\delta^{13}C$ lower than –38‰ is expected for certain situations (corresponding to an even lower posterior flaring contribution). Here, –38‰ was used in agreement with the (to our knowledge) only published $\delta^{13}C$ characterization of BC from gas flaring (44).

Domestic is a mixed source (60% biomass, 39% coal, and 1% liquid fossil) (24). The end-member values (for both $\delta^{13}C$ and $\Delta^{14}C$) were estimated assuming normal distribution mixing, where the mean (μ) and SD (σ) were estimated as:

$$\mu_D = 0.6 \times \mu_B + 0.39 \times \mu_C + 0.01 \times \mu_L \quad \text{[S1a]}$$

$$\sigma_D = \sqrt{(0.6 \times \sigma_B)^2 + (0.39 \times \sigma_C)^2 + (0.01 \times \sigma_L)^2}, \quad \text{[S1b]}$$

where D = domestic, B = biomass, C = coal, and L = liquid fossil.

Estimating the SD of the Priors. The fractional source contribution priors for the Bayesian analysis are assumed to be normally distributed with a mean (μ) equal to the FEG model results. The second parameter in the normal distribution is the SD (σ), which also needs to be estimated. BC bottom-up EIs typically report large uncertainties (e.g., 125 to 500%) in the flux (14, 15). However, the uncertainties for the fractional contributions from different sources need to be lower (<100%), as constrained by the mass balance criterion. To fully estimate the uncertainties of the relative source contributions to Tiksi BC levels from the FEG model would mean a vast modeling effort, i.e., coupled uncertainty propagation from FLEXPART, ECLIPSE, and GFED. Instead, given the comparably large expected uncertainties of the FEG modeling results, we estimated the uncertainties (σ) using the weakest assumptions of prior knowledge: Each possible fractional source combination is equally probable. The distribution that describes this scenario for n sources (here $n = 5$) is the n-dimensional Dirichlet distribution where the shape factors (α) are all equal to 1 (a.k.a., the n-dimensional standard simplex). The SD of the marginal distribution (the distribution of one of the fractional source contributions) of this symmetric Dirichlet distribution is given by:

$$\sigma = \sqrt{\frac{n-1}{n^2 \times (n+1)}}. \quad \text{[S2]}$$

Five sources thus give an uncertainty of roughly 16.3%. For comparison, the two-source case of this Dirichlet distribution is the standard uniform distribution, with $\sigma \approx 28.9\%$.

MCMC Analysis. The MCMC analysis was conducted using in-house scripts written in Matlab (ver. 2014b). A Metropolis–Hastings algorithm (63, 64) was used for sampling the parameter space, using 1,000,000 iterations, a burn-in of 10,000, and a data thinning of 10. The jump size of the stochastic perturbation was adjusted to obtain an acceptance rate of ∼0.23. The end-member distributions were described with normal distributions (28). The different samples were given different weight in the fitting, depending on observational sampling duration. The prior relative source distributions before the Bayesian modeling were assigned with normal distributions with (mean) values from the FEG model, and with an SD of 16.3%. For the sources where a general shift was introduced, the prior was exchanged from one per sample to one parameter affecting all samples, where the shift (here exemplified by gas flaring, x_F) was allowed to take any positive number, exemplified as:

$$f_F(j)_{posterior} = \frac{x_F \times f_F(j)_{prior}}{x_F \times f_F(j)_{prior} + f_{OF}(j)_{prior} + f_C(j)_{prior} + f_L(j)_{prior} + f_D(j)_{prior}} \quad \text{[S3a]}$$

$$f_{OF}(j)_{posterior} = \frac{f_{OF}(j)_{prior}}{x_F \times f_F(j)_{prior} + f_{OF}(j)_{prior} + f_C(j)_{prior} + f_L(j)_{prior} + f_D(j)_{prior}} \quad \text{[S3b]}$$

$$f_C(j)_{posterior} = \frac{f_C(j)_{prior}}{x_F \times f_F(j)_{prior} + f_{OF}(j)_{prior} + f_C(j)_{prior} + f_L(j)_{prior} + f_D(j)_{prior}} \quad \text{[S3c]}$$

$$f_L(j)_{posterior} = \frac{f_L(j)_{prior}}{x_F \times f_F(j)_{prior} + f_{OF}(j)_{prior} + f_C(j)_{prior} + f_L(j)_{prior} + f_D(j)_{prior}} \quad \text{[S3d]}$$

$$f_D(j)_{posterior} = \frac{f_D(j)_{prior}}{x_F \times f_F(j)_{prior} + f_{OF}(j)_{prior} + f_C(j)_{prior} + f_L(j)_{prior} + f_D(j)_{prior}}, \quad \text{[S3e]}$$

where F = gas flaring, OF = open fires, C = coal, L = liquid fossil, D = domestic, and f denotes the corresponding fractional contributions. The index j denotes sample number. Thus, if x takes on a value larger than 1, the overall fractional contribution of source gas flaring increases (and all others

decrease), and, if it is smaller than 1, it decreases (and all others increase).

Bayesian Model Comparison. For comparing the significance of the data fit for different models, evaluation of the Bayes factor is a powerful approach. In the present case, the different models that were compared shifted the importance of one or more sources to better fit with the observational data, i.e., the $\Delta^{14}C$ and $\delta^{13}C$ data for EC. The Bayes factor (K) is defined as the conditional probability (P) of data (D) given model 1 (the evidence for model 1, M_1) divided by the conditional probability for data given model 2 (the evidence for model 2, M_2):

$$K = \frac{P(D|M_1)}{P(D|M_2)}. \quad \textbf{[S4]}$$

The evidence for a model was computed by integrating (marginalizing) over the parameters assigned to that model. As an example, the Bayes factor comparing shifting both flaring (F) and open fires (OF) vs. shifting only flaring is given by the ratio between the evidence for the first model divided by the evidence for the second:

$$K = \frac{\iint P(D|x_F, x_{OF},\ M_{F,OF}) \times P(x_F, x_{OF}|M_{F,OF})dx_F dx_{OF}}{\int P(D|x_F,\ M_F) \times P(x_F|M_F)dx_F}, \quad \textbf{[S5]}$$

where x denotes the shifting factor. For the present example, two shifting factors are associated with model 1 and one shifting factor for model 2. The evidence for each model was computed separately (within model approach) using MCMC sampling over the parameter space, and the Bayes factor was computed through combinatorial comparison of the different models. For this particular case (comparing the two models with the best fits), the Bayes factor is >100, which is a "decisive" or "very strong" in favor of the combination of combining gas flaring and open fires. Adding a third parameter, e.g., coal, does not significantly improve the fit, as the Bayes factor is ~1. An advantage with using Bayes factors for model comparisons is that the degrees of freedom, e.g., the number of sources that are shifted, are naturally incorporated in the integral in Eq. **S5**.

Winiger et al. www.pnas.org/cgi/content/short/1613401114

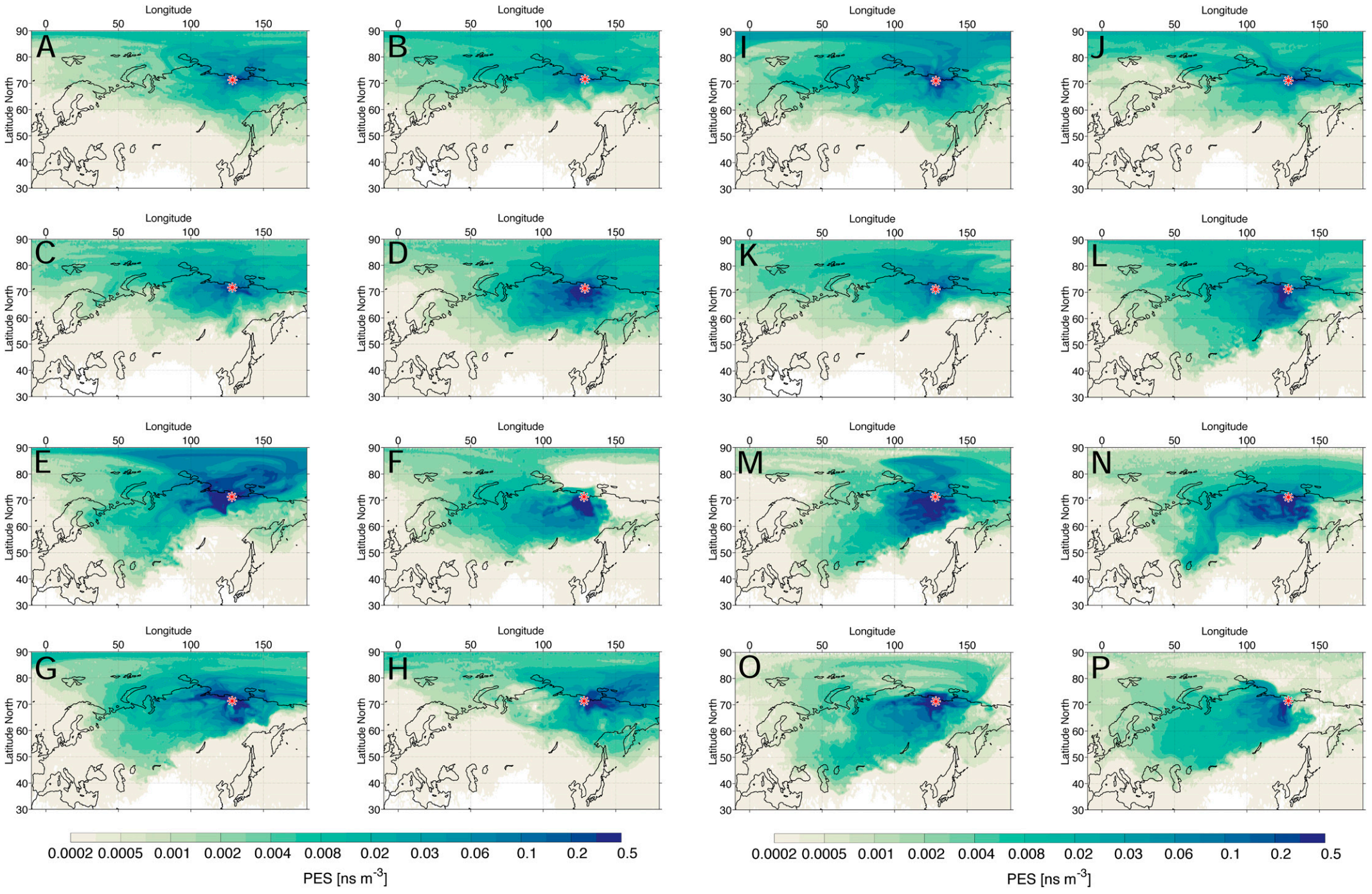

Fig. S1. FLEXPART footprints. FLEXPART potential footprint emission sensitivity (PES) for the BC aerosol tracer arriving at Tiksi, Russia (red star). (*A*) Isotope data point (given as range with yyyy-mm-dd) 2012-04-16 to 2012-06-21. (*B*) Isotope data point 2012-06-21 to 2012-08-24. (*C*) Isotope data point 2012-08-24 to 2012-10-25. (*D*) Isotope data point 2012-10-25 to 2012-12-27. (*E*) Isotope data point 2012-12-27 to 2013-01-14. (*F*) Isotope data point 2013-02-06 to 2013-02-27. (*G*) Isotope data point 2013-02-27 to 2013-03-20. (*H*) Isotope data point 2013-03-20 to 2013-04-10. (*I*) Isotope data point 2013-04-10 to 2013-05-02. (*J*) Isotope data point 2013-05-02 to 2013-05-23. (*K*) Isotope data point 2013-07-25 to 2013-10-17. (*L*) Isotope data point 2013-10-17 to 2014-01-02. (*M*) Isotope data point 2014-01-02 to 2014-01-23. (*N*) Isotope data point 2014-01-23 to 2014-02-14. (*O*) Isotope data point 2014-02-14 to 2014-03-07. (*P*) Isotope data point 2014-03-07 to 2014-03-28.

Winiger et al. www.pnas.org/cgi/content/short/1613401114

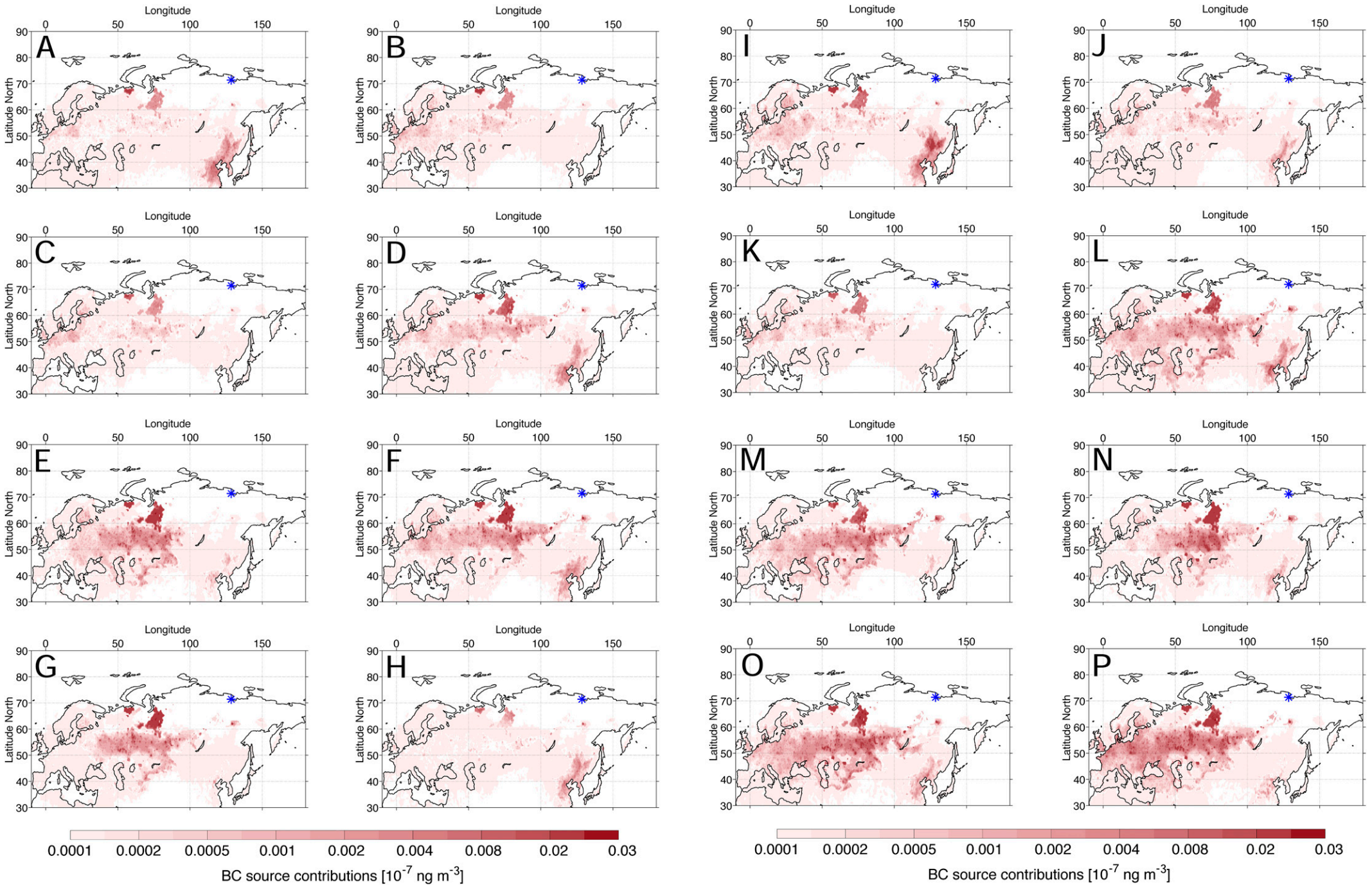

Fig. S2. FLEXPART BC source contributions. FLEXPART geographical distribution of the anthropogenic BC source contribution to the simulated mixing ratio at Tiksi, Russia (blue star). (*A*) Isotope data point (given as range with yyyy-mm-dd) 2012-04-16 to 2012-06-21. (*B*) Isotope data point 2012-06-21 to 2012-08-24. (*C*) Isotope data point 2012-08-24 to 2012-10-25. (*D*) Isotope data point 2012-10-25 to 2012-12-27. (*E*) Isotope data point 2012-12-27 to 2013-01-14. (*F*) Isotope data point 2013-02-06 to 2013-02-27. (*G*) Isotope data point 2013-02-27 to 2013-03-20. (*H*) Isotope data point 2013-03-20 to 2013-04-10. (*I*) Isotope data point 2013-04-10 to 2013-05-02. (*J*) Isotope data point 2013-05-02 to 2013-05-23. (*K*) Isotope data point 2013-07-25 to 2013-10-17. (*L*) Isotope data point 2013-10-17 to 2014-01-02. (*M*) Isotope data point 2014-01-02 to 2014-01-23. (*N*) Isotope data point 2014-01-23 to 2014-02-14. (*O*) Isotope data point 2014-02-14 to 2014-03-07. (*P*) Isotope data point 2014-03-07 to 2014-03-28.

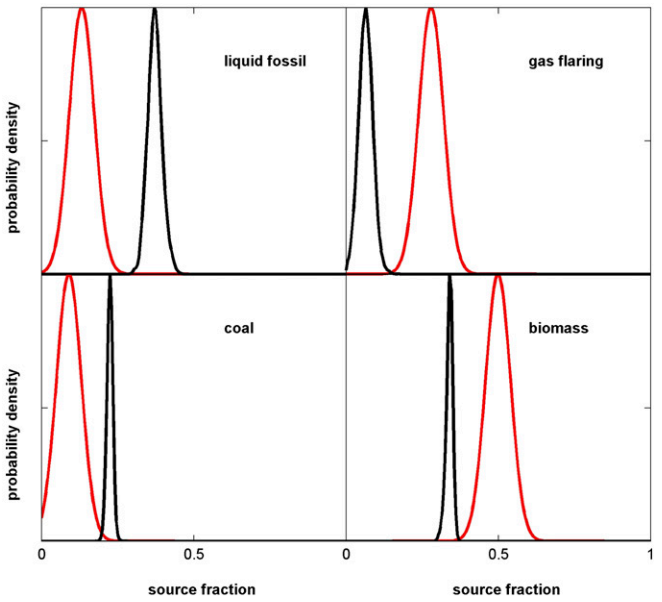

Fig. S3. Mean Bayesian source contributions. Prior (red) and posterior (black) mean fractions of the four fuel-type emissions sources from the Bayesian modeling for the full campaign (2 y). Note that the SD for the prior in these plots is not 16.3%, as they are the average of the 16 data points. Instead, it is calculated as 16.3 divided by $16^{0.5}$ (~4%), reflecting the temporal averaging.

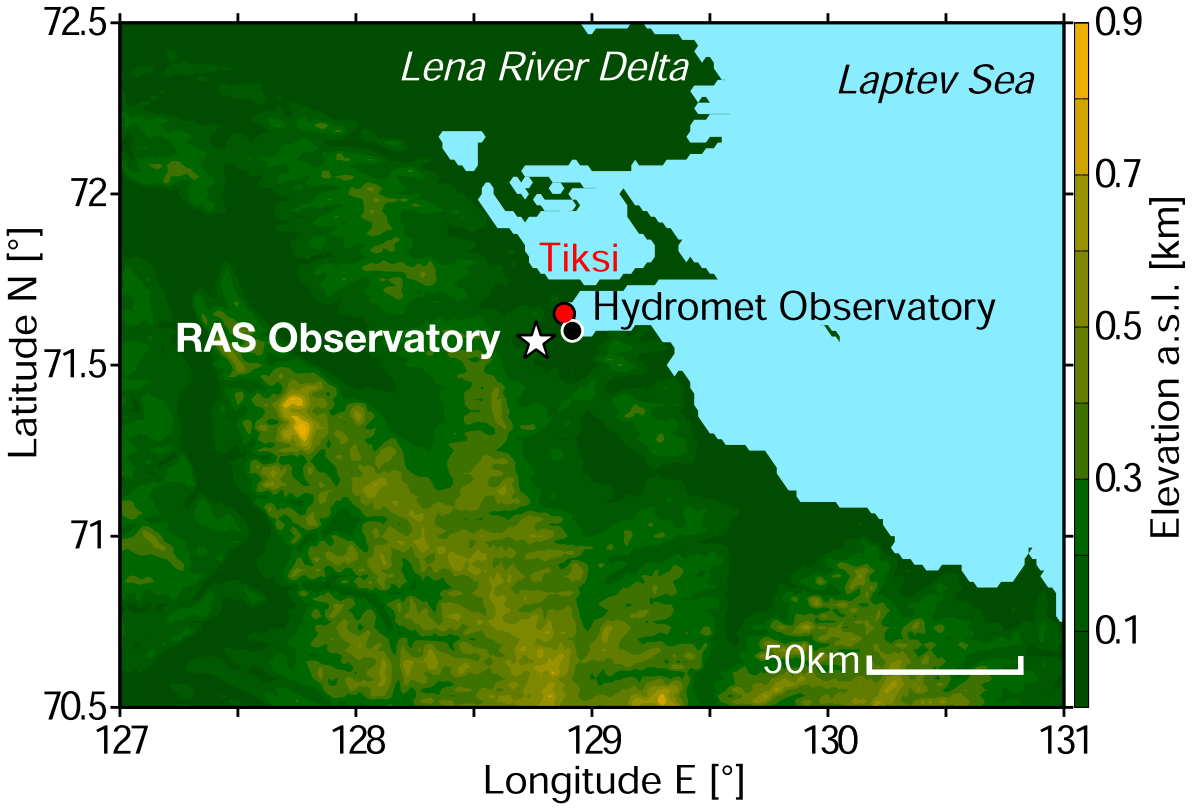

Fig. S4. The study site. Elevation map of the RAS Polar Geocosmophysical Observatory (star) and Hydromet observatory sites (black circle) in relation to Tiksi village (red circle); a.s.l., above sea level.

Table S1. Observation vs. model

Start yyyy-mm-dd	Sampling time, d	Observation				FLEXPART–ECLIPSE				
		TSP EC, ng C·m^{-3}			f_{bb}	Fossil EC, ng·m^{-3}	Biofuels EC, ng·m^{-3}	Fire EC, ng·m^{-3}	Σ EC, ng·m^{-3}	f_{bb}
2012-04-16	66.9	67.7	± 5.5		0.408	6.6	1.0	12.3	19.9	0.670
2012-06-21	63.0	37.8	± 4.3		0.731	5.1	0.6	44.8	50.5	0.899
2012-08-24	62.0	19.8	± 3.1		0.619	5.3	0.6	5.5	11.4	0.536
2012-10-25	63.0	21.2	± 3.3		0.253	17.1	1.0	0.2	18.2	0.061
2012-12-27	17.9	73.5	± 6.4		0.194	42.3	1.5	0.0	43.8	0.035
2013-02-06	21.0	302.1	± 16.2		0.080	44.9	2.2	0.0	47.1	0.046
2013-02-27	21.0	82.5	± 5.3		0.119	38.9	0.8	0.0	39.8	0.021
2013-03-20	21.0	85.8	± 5.5		0.187	4.9	0.8	0.3	6.0	0.191
2013-04-10	22.0	103.2	± 6.5		0.141	22.0	3.2	3.1	28.3	0.222
2013-05-02	21.0	63.0	± 4.4		0.239	7.4	0.9	3.0	11.3	0.340
2013-05-23	63.0	40.8	± 3.3		0.362	*	*	*	*	*
2013-07-25	84.0	8.0	± 1.5		0.600	6.0	0.4	83.8	90.2	0.934
2013-10-17	78.0	19.2	± 2.1		0.605	25.9	2.5	0.3	28.7	0.097
2014-01-02	20.1	41.0	± 3.1		0.359	32.5	1.8	0.0	34.3	0.052
2014-01-23	21.9	47.2	± 3.4		0.313	67.3	1.3	0.0	68.6	0.019
2014-02-14	21.0	43.0	± 3.2		0.300	51.6	3.0	0.2	54.8	0.058
2014-03-07	21.0	37.3	± 3.0		0.263	58.3	5.2	4.6	68.1	0.144
ALL (analog to FLEXPART)		46.7			0.304	19.8	1.4	18.1	39.3	0.398
stdev		66.7			0.200	21.2	1.3	22.6	23.8	0.316
2013-02-06 to 2014-02-14	310	55.4			0.210	23.1	1.5	23.2	47.8	0.339
stdev		87.7			0.197	20.8	0.9	27.7	26.6	0.291
2013-02-27 to 2014-03-07	310	37.4			0.290	23.5	1.5	23.2	48.3	0.339
stdev		31.6			0.180	21.6	1.0	27.7	27.0	0.290
2013-03-20 to 2014-03-28	310	34.4			0.315	24.8	1.8	23.6	50.2	0.348
stdev		30.2			0.166	23.6	1.5	27.5	28.5	0.283
2013-02-06 to 2014-02-14	373	53.3			0.227					
stdev		83.8			0.187					
2013-02-27 to 2014-03-07	373	37.9			0.302					
stdev		30.1			0.170					
2013-03-20 to 2014-03-28	373	35.4			0.323					
stdev		28.6			0.156					

TSP EC compared with FLEXPART data; stdev, standard deviation.
*No model calculations were available, i.e., observation data were excluded for the total and yearly averages.

Table S2. Stable carbon ($\delta^{13}C$) end-members for different BC sources

BC source	Liquid fossil	Coal	Gas flaring	Domestic*	Biomass†	R fossil‡
$\Delta^{14}C$, ‰	−1,000 ± 0	−1,000 ± 0	−1,000 ± 0	−265 ± 1.2	+225 ± 60	−1,000 ± 0
$\delta^{13}C$, ‰	−31.4 ± 1	−23.4 ± 1.3	−38 ± 3	−25.5 ± 36	−26.7 ± 1.8	−25.5 ± 1.3

The end-members are site-/region-specific. Applied end-members are according to Andersson et al. (28) for coal and R fossil, Mašalaitė et al. (40) for liquid fossil, and Widory (44) for gas flaring. For biomass, the $\delta^{13}C$ end-member is used according to Andersson et al. (28) and Winiger et al. (13), reflecting the photosynthesis pathway of trees. Although we applied a $\Delta^{14}C$ end-member of 225 ± 25‰ for the European Arctic (13, 31), an end-member with higher variability was used for the Russian Arctic, to account for possible differences in biota or lumbering behavior (e.g., trees older than 60 y).
*Estimated as 60% biomass, 39% coal, and 1% liquid fossil (24).
†Biomass (open fires) were used interchangeably with GFED, even though GFED captures also other events.
‡Liquid fossil fuels of regular origin, i.e., consumed in, e.g., Western Europe or China.

Table S3. Isotope analysis

Start yyyy-mm-dd	Sampling duration, d	TSP_{EC} $\Delta^{14}C$, ‰	f_{bb}	$\delta^{13}C$, ‰
2012-04-16	66.9	−500 ± 2	0.408 ± 0.026	−28.2 ± 0.2
2012-06-21	63.0	−105 ± 3	0.731 ± 0.049	−25.8 ± 0.2
2012-08-24	62.0	−242 ± 2	0.619 ± 0.046	−27.5 ± 0.2
2012-10-25	63.0	−690 ± 1	0.253 ± 0.018	−30.1 ± 0.2
2012-12-27	17.9	−762 ± 2	0.194 ± 0.029	−29.2 ± 0.2
2013-02-06	21.0	−902 ± 3	0.080 ± 0.023	−30.7 ± 0.2
2013-02-27	21.0	−855 ± 3	0.119 ± 0.028	−29.2 ± 0.2
2013-03-20	21.0	−771 ± 2	0.187 ± 0.031	−29.0 ± 0.2
2013-04-10	22.0	−827 ± 2	0.141 ± 0.031	−29.4 ± 0.2
2013-05-02	21.0	−707 ± 1	0.239 ± 0.034	−28.6 ± 0.2
2013-05-23	63.0	−556 ± 1	0.362 ± 0.038	−28.1 ± 0.2
2013-07-25	84.0	−265 ± 2	0.600 ± 0.034	−27.6 ± 0.2
2013-10-17	78.0	−259 ± 2	0.605 ± 0.015	−28.0 ± 0.2
2014-01-02	20.1	−560 ± 1	0.359 ± 0.027	−27.5 ± 0.2
2014-01-23	21.9	−616 ± 1	0.313 ± 0.023	−28.2 ± 0.2
2014-02-14	21.0	−633 ± 1	0.300 ± 0.027	−27.0 ± 0.2
2014-03-07	21.0	−677 ± 1	0.263 ± 0.028	−26.9 ± 0.2
All			0.308 ± 0.194	

Shown are isotopic data from ambient aerosol samples of the EC fraction of TSP. The uncertainties for the isotope data are based on AMS measurement errors (1 SD), and the f_{bb} uncertainty is based on MCMC calculations (including measurement and sampling uncertainties).

Table S4. Emission partitioning of the ECLIPSE EI data

Source type	Emission type
Biofuels	Residential and commercial
	Industry (combustion and processing)
	Power plants
Fossil fuels	Residential and commercial
	Residential and commercial; nonfuel activity
	Power plants, energy conversion, extraction
	Industry (combustion and processing)
	Industry (combustion and processing); nonfuel activity
	Power plants
	Power plants; nonfuel activity
	Surface transportation
	Waste
Open fires (wild fires not included via ECLIPSE)	Agricultural waste burning

All available (anthropogenic) ECLIPSE emissions were split according to their source type. Agricultural waste burning (AGW, e.g., on fields) is included in the Global Fire Emissions Database (GFED). AGW was hence excluded from the ECLIPSE emissions and implemented via GFED, to avoid double counting of AGW emissions.