Washington State Regional Haze State Implementation Plan

Appendix D

Western Regional Air Partnership Interagency Monitoring Protected Visual Environments Monitoring Data Substitutions

WRAP IMPROVE Data Substitutions
04/03/07

In the Western Regional Air Partnership (WRAP) states, data substitution was performed for nine IMPROVE monitoring sites to achieve RHR data completeness, or to fully populate 2002, WRAP's selected modeling year. These data substitutions included estimating missing species from other on-site measurements and appropriately scaling data collected at selected donor sites which had favorable long-term comparisons. This document outlines the data substitution methods used at these sites.

Data Completeness Requirements

Regional Haze Rule (RHR) guidance outlines IMPROVE aerosol data completeness requirements including the following conditions:

- Individual samples must contain all species required for the calculation of light extinction (sulfate, nitrate, organic carbon, elemental carbon, soil, coarse mass, and, for the new IMPROVE algorithm, chloride or chlorine)
- Individual seasons must contain at least 50% of all possible daily samples
- Individual years must contain at least 75% of all possible daily samples
- Individual years must not contain more than 10 consecutive missing daily samples
- The baseline period (2000-04) must contain at least 3 complete years of data

Further details can be found in the RHR guidance document for tracking progress: http://www.epa.gov/ttn/oarpg/t1/memoranda/rh_tpurhr_gd.pdf.

Routine Data Substitutions

RHR guidelines provide provisions to fill in missing data under specific circumstances. There are currently two methods routinely used in preparing the RHR data set to substitute data for missing samples:

- The use of a surrogate in the data set:
    - Total sulfate is generally determined as 3 times the sulfur measured on the A module filter. If sulfur is missing, the sulfur measurement from the B module filter is used to calculate sulfate.
    - For the new IMPROVE algorithm, sea salt is calculated from chloride measured on the B module filter. If chloride is missing or below detection limit, the chlorine measurement from the A module filter is used to calculate sea salt.

- The application of "patching" missing data described by the RHR guidance:

- Missing samples not substituted using a surrogate as described above can be patched, or replaced, by a seasonal average if the patching exercise passes a series of tests outlined in the guidance document.

Once these methods have been applied to the data, the resulting complete years are eligible for use in calculation of baseline conditions and tracking progress under the Regional Haze Rule. These methods have been applied to all IMPROVE data.

Sites Not Meeting Data Completeness Requirements

After routine data substitutions were made, some WRAP sites still failed to meet data completeness requirements for the baseline period. These sites are listed in Table 1. Sites were candidates for substitution for two reasons:

- The sites had fewer than 3 complete years of data, thus RHR visibility metrics for the baseline period could not be calculated.

- The sites had at least 3 years of complete data, but were missing 2002, the year selected for regional modeling. If this year is missing, then the worst 20% visibility days from 2002 cannot be determined, and the relative response factors (RRFs), which are used to predict visibility metrics in 2018, cannot be calculated.

Sites that did not meet data completeness requirements were not necessary for submittal of State Implementation Plans (SIPs) are indicated with an asterisk (*) in Table 1. Additional data substitutions for these sites have not been applied.

Table 1
WRAP Sites Failing RHR Data Completeness Requirements

State	Site	<3 years	Missing 2002
AZ	BALD1	X	X
	INGA1*	X	X
	TONT1		X
CA	KAIS1	X	X
	RAFA1	X	X
	SEQU1		X
	TRIN1		X
MT	FLAT1*	X	X
	FOPE1*	X	X
	GLAC1		X
	NOCH1*	X	X
UT	CAPI1	X	X
WA	NOCA1	X	

* Indicates additional substitution is not required for a SIP.

Additional Data Substitutions

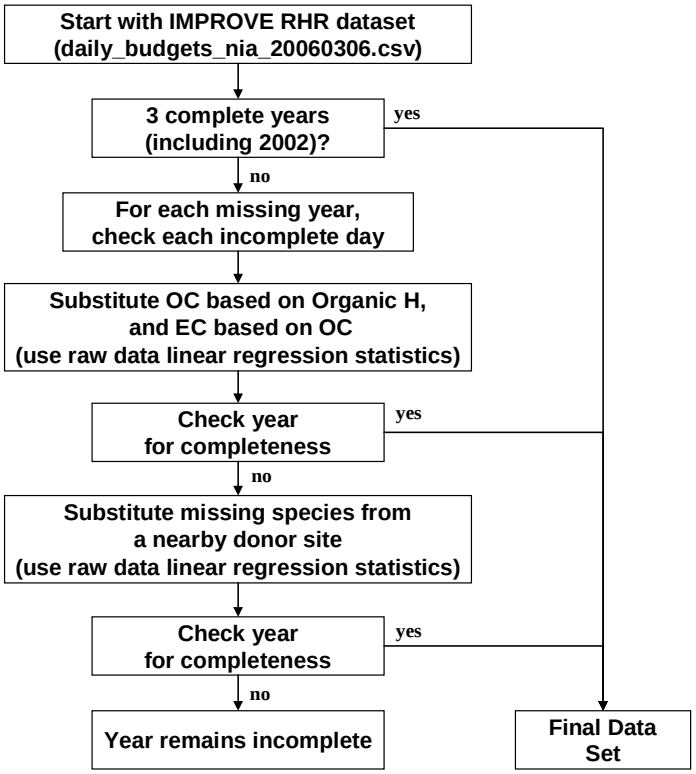
This section outlines the WRAP methods for additional data substitutions designed to address the problems at sites listed in Table 1. Similar methods have been used at IMPROVE sites with incomplete data records in other RPOs.

Figure 1 presents a flow chart of the WRAP data substitution methods. The starting data set was the RHR IMPROVE data using the "New IMPROVE Algorithm," updated March 2006, (http://vista.cira.colostate.edu/views/Web/IMPROVE/SummaryData.aspx). This data set includes the routine surrogate and patched data substitutions allowed by RHR guidance. Note that only years deemed incomplete under RHR guidance were candidates for additional data substitutions. Years deemed complete were not changed, even thought there may have been missing samples during those years.

The first of the additional substitution methods used organic hydrogen as a surrogate for organic carbon, and resultant organic carbon as a surrogate for elemental carbon. If the carbon data substitution was not sufficient to complete the required years, measured mass for individual species from nearby IMPROVE sites with favorable long-term comparisons were scaled appropriately and used as surrogates. IMPROVE donor sites were selected in consultation with individual states. These methods are described in detail below.

Figure 1

Flow Chart of Data Substitution Methods Used



All substitutions were made using quarterly specific Kendall-Theil linear regressions statistics. These statistics were chosen because they are more resistant to outliers than the standard linear least squares statistics. Kendall-Theil slopes and intercepts were used to calculate substituted values from surrogates.

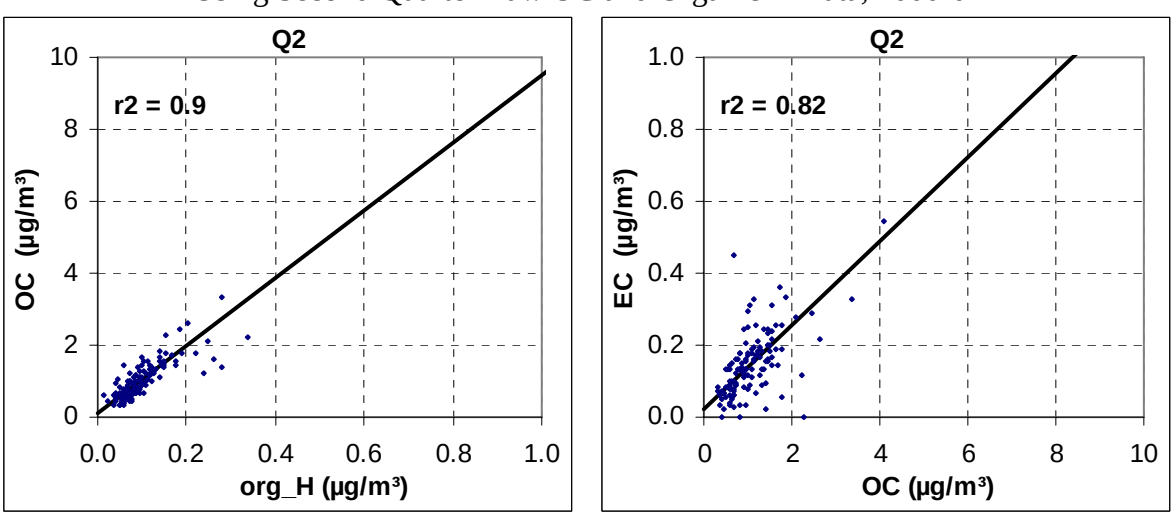
1. Carbon Substitutions

The first substitution method relied on using a surrogate for carbon mass measurements when the C module data is not available. Hydrogen (H) is measured on the A module filter, and is assumed to be primarily associated with organic carbon and inorganic compounds such as ammonium sulfate. Therefore, organic carbon (OC) can be estimated using the historical comparison between estimated organic H and OC. Organic H is estimated by subtracting the portion of H that is assumed to be associated with the inorganic compounds from the total H (Org_H = H – 0.24*S).

Figure 2 presents a sample comparison for data collected at the Tonto National Monument site in Arizona during the second quarter between 2000-04 for OC and organic H. Once OC has been estimated using this method, elemental carbon (EC) mass is determined using long-term comparisons between OC and EC at the site. Statistics were calculated and applied quarterly to account for seasonal variations.

Figure 2

Comparison of OC and Estimated Organic H, and EC and OC at Tonto National Monument, AZ Using Second Quarter Raw OC and Organic H Data, 2000-04



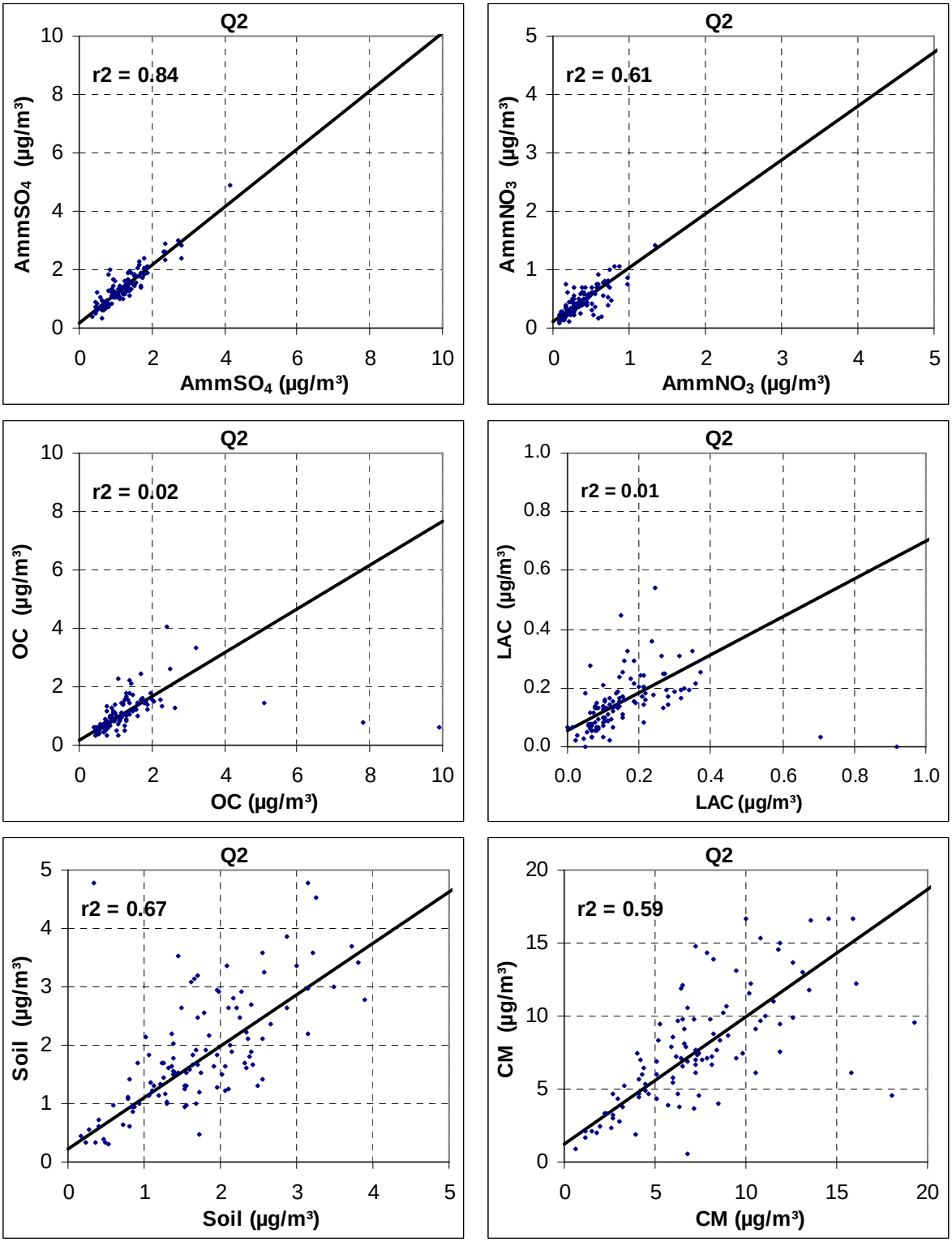
2. Donor Site Substitutions

In the WRAP, the carbon data substitution methods were not sufficient to complete the required years. A second method involved identification of another nearby IMPROVE site which had favorable long-term comparisons and similar regional characteristics to be used as a donor site. Candidate sites were identified, and final donor sites for surrogate mass were selected in consultation with states.

Figure 3 presents a sample inter-site mass comparison by species for data collected during the second quarter, 2000-04, between the Tonto National Monument site and the Sierra Ancha site in Arizona. Component specific correlations were calculated and applied quarterly. Note that only species missing in a given sample were substituted based on donor site data. Species collected at the site under investigation were never replaced with data from a donor site.

Figure 3

Comparison of Aerosol Species Mass Between
Tonto National Monument, AZ (y-axis) and Sierra Ancha, AZ (x-axis)
Using Second Quarter Raw Data, 2000-04



2. Data Completeness Following Substitutions

Table 2 indicates which years required some degree of substitution, where a 2 indicates a substituted year, a 1 indicates the year was already complete under RHR guidelines, and dashes indicate the year did not meet RHR guidelines and no additional substitutions were made. The table also lists sites that were selected as donor sites.

The minimum data requirement of 3 complete years (including 2002) was met for each site, and additional substitutions beyond these requirements were made on a case by case basis in consultation with individual states. For example, at the KAIS1 site, substitutions were made only for the 2002 year even though substituted data (from the YOSE1 donor site) was available for other years. In this case, the years 2000 and 2001 had less than 50% of the original RHR data. In contrast, additional substitutions were applied for all incomplete years (2000-2002) at the RAFA1 site. For the RAFA1 site, the original RHR data was more substantial (73-86% available) and substitutions had less of an impact on the worst days distributions.

Table 2
Data Completeness at WRAP Sites Following Data Substitution

State	Site	<3 years	Missing 2002	Donor	2000	2001	2002	2003	2004
AZ	BALD1	X	X	TONT1	--	2	2	1	1
	TONT1		X	SIAN1	--	1	2	1	1
CA	KAIS1	X	X	YOSE1	--	--	2	1	1
	RAFA1	X	X	PINN1	2	2	2	1	1
	SEQU1		X	DOME1	1	1	2	2	1
	TRIN1		X	LAVO1	--	1	2	1	1
MT	GLAC1		X	FLAT1	1	1	2	2	1
UT	CAPI1	X	X	CANY1	2	2	2	1	1
WA	NOCA1	X		SNPA1	--	1	1	2	2

-- indicates an incomplete year with no substitutions made
1 indicates a complete RHR year
2 indicates a year is considered complete with some substituted values

Availability and Archival of Data Sets

A dedicated page on the VIEWS database will act as the repository of all site-specific substitute data sets: http://vista.cira.colostate.edu/views/web/documents/substitutedata.aspx. Table 3 presents a key to the substituted data files. All materials prepared in the data substitution work (descriptive narrative, tables of regression statistics, graphics, etc.) will be posted on this site for review by states, tribes, and other data users. This information will also be made accessible through the TSS.

Table 3
Key to Substituted Data Files

Column Header	Description
site	IMPROVE site code
year	
month	
day	
QUARTER	1 = Jan. – Mar., 2 = Apr.-Jun., 3 = Jul. – Sept., 4 = Oct. – Dec.
date	
Group	10 = One of the 20% best visibility days; 90 = One of the 20% worst visibility days
good_year	0 = incomplete year, 1 = complete RHR year, 2 = complete year with substitutions
ss_rayleigh	Site specific Rayleigh value (clean air extinction)
fsrh	f(RH) value for small sulfate, nitrate and organic mass
flrh	f(RH) value for large sulfate, nitrate and organic mass
fssrh	f(RH) value for sea salt mass
Sea_Salt	Sea salt mass (µg/m3)
Soil	Soil Mass (µg/m3)
Amm_NO3	Ammonium nitrate mass (µg/m3)
OMC	Organic mass by carbon (µg/m3)
LAC	Light absorbing carbon (aka EC/Elemental Carbon) (µg/m3)
CM	Coarse mass (µg/m3)
Amm_SO4	Ammonium sulfate mass (µg/m3)
Large_OMC	Large organic mass (µg/m3)
Small_OMC	Small organic mass (µg/m3)
Large_Amm_SO4	Large ammonium sulfate mass (µg/m3)
Small_Amm_SO4	Small ammonium sulfate mass (µg/m3)
Large_Amm_NO3	Large ammonium nitrate mass (µg/m3)
Small_Amm_NO3	Small ammonium nitrate mass (µg/m3)
EAmm_SO4	Extinction due to ammonium sulfate (Mm-1)
EAmm_NO3	Extinction due to ammonium nitrate (Mm-1)
EOMC	Extinction due to organic carbon mass (Mm-1)
ELAC	Extinction due to light absorbing carbon mass (Mm-1)
ESoil	Extinction due to soil mass (Mm-1)
ECM	Extinction due to coarse mass (Mm-1)
ESea_Salt	Extinction due to sea salt mass (Mm-1)
RBext	Reconstructed aerosol extinction (Mm-1)
TBext	Reconstructed total extinction (Mm-1)
OC_SUB1	OC substituted using OC vs. organic H correlations
EC_SUB1	EC substituted using EC vs. OC correlations
(NH4)2SO4_SUB2	Ammonium sulfate substituted using site donor correlations
(NH4)NO3_SUB2	Ammonium nitrate substituted using site donor correlations
OM_SUB2	Organic mass substituted using site donor correlations
EC_SUB2	Elemental carbon (aka light absorbing carbon) substituted using site donor correlations
Soil_SUB2	Soil substituted using site donor correlations
CM_SUB2	Coarse mass substituted using site donor correlations
SeaSalt_SUB2	Sea salt substituted using site donor correlations

Results for Washington Sites

The charts and tables on the following pages detail the substitutions that were used for the Washington NOCA1 site. The following charts/tables are provided for each site:

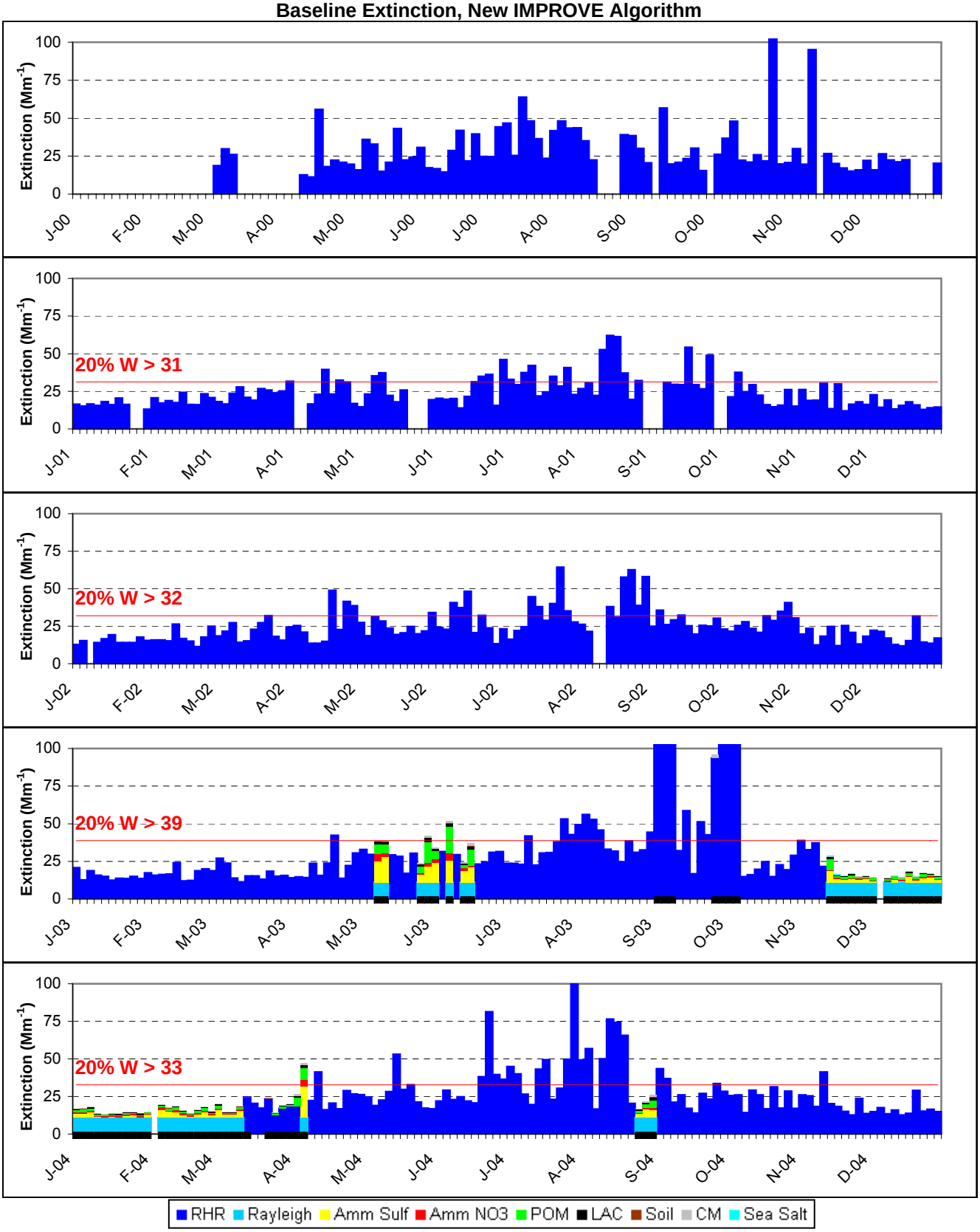
- **1st chart:** Bar charts by year indicating original RHR data in blue, and substituted data by species in the standard IMPROVE colors. Substituted days are also indicated by a black bar underneath the day. The red line indicates the threshold above which days are counted in the 20% worst days for that year. A red line is not included for any year that was incomplete and not substituted.

- **2nd chart:** Bar charts by year indicating speciation of all data. Days in which all or part of the day was substituted are indicated by a black bar underneath the day. The red line indicates the threshold above which days are counted in the 20% worst days.

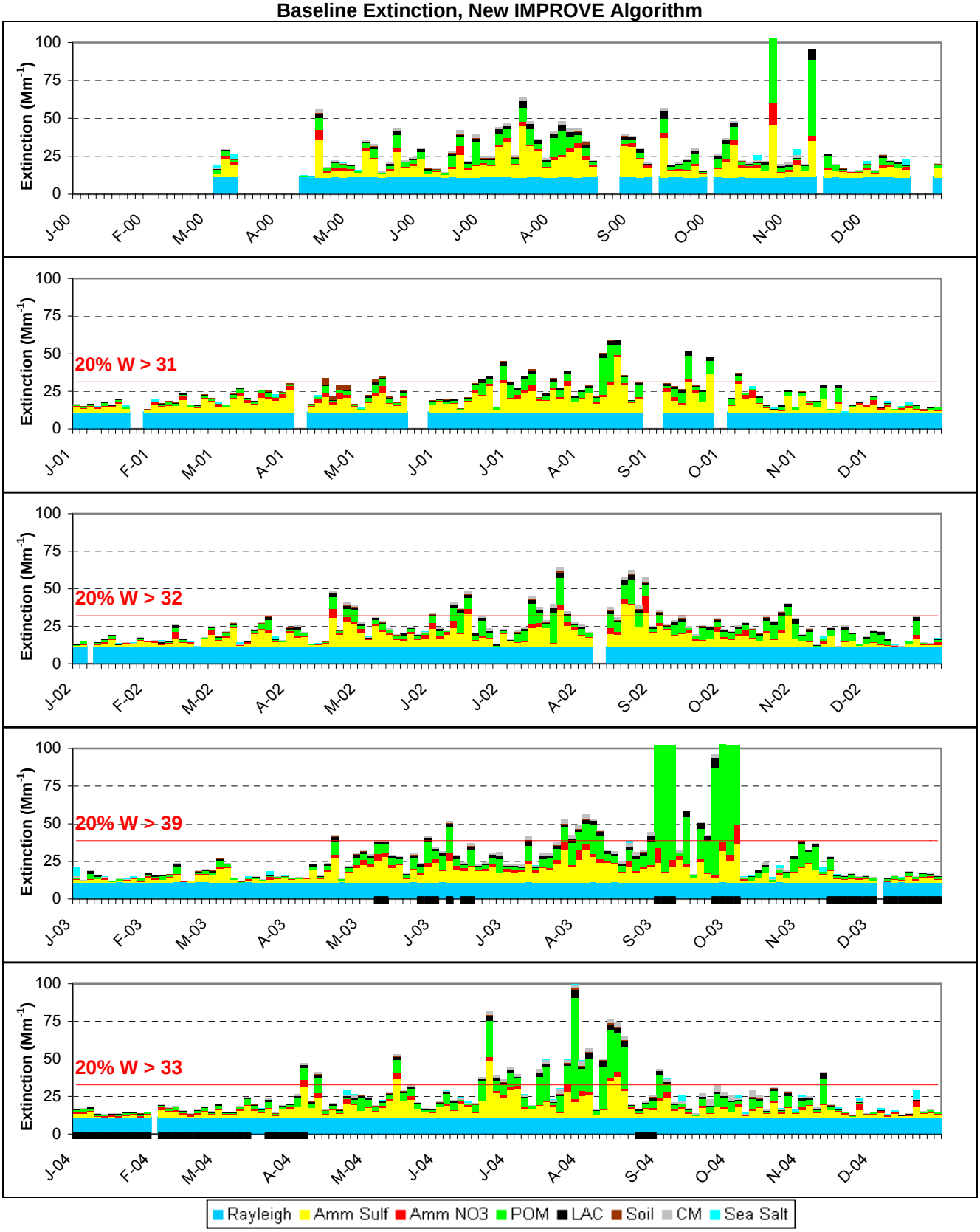
- **Table:** The table lists the Kendall-Theil regression statistics used to calculate substituted values. Data represent raw mass value correlations for aerosol collected during 2000-04 for all days, and by quarter. The median absolute deviation (MAD) statistic was used to characterize the degree of correlation, where the closer the MAD statistic is to zero, the better the line fit.

- **Scatter plots:** A series of nine scatter plots, showing the distribution of points and the line fits for the regression statistics that were used to calculate substituted values. Plots indicate raw data collected at the sites during 2000-04, with donor sites represented on the x-axis.

NOCA1 (SNPA1 as donor)
Baseline Extinction, New IMPROVE Algorithm



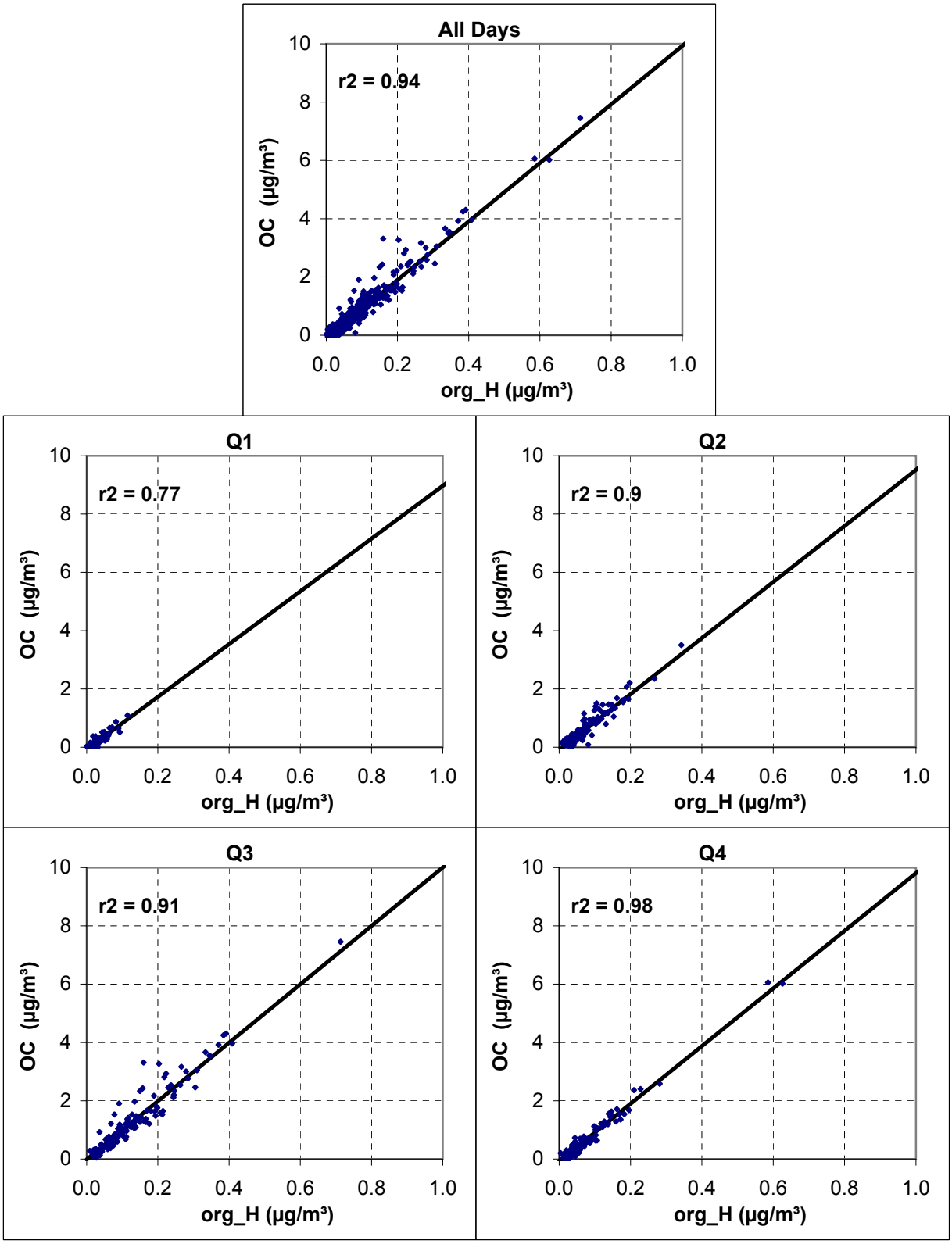
NOCA1 (SNPA1 as donor)
Baseline Extinction, New IMPROVE Algorithm



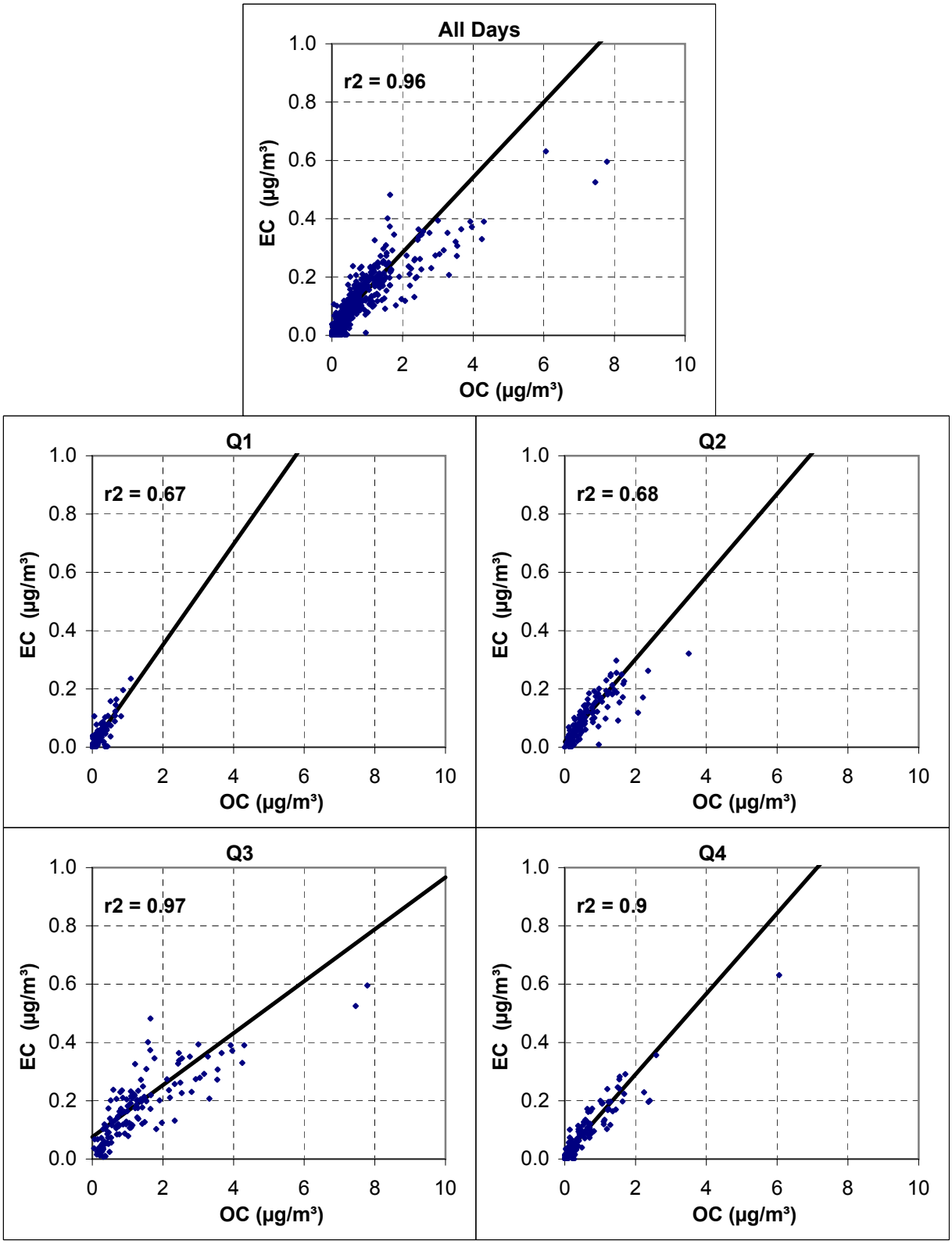
Statistics for North Cascades, WA Substitutions

NOCA1						
Components	**KT Stats**	**All**	**1**	**2**	**3**	**4**
OC vs. org_H	**Slope**	10.04	9.05	9.62	10.02	9.89
	Intercept	-0.11	-0.08	-0.10	-0.02	-0.08
	MAD	0.09	0.06	0.08	0.15	0.10
EC vs. OC	**Slope**	0.13	0.17	0.14	0.09	0.14
	Intercept	0.03	0.01	0.02	0.08	0.01
	MAD	0.03	0.01	0.02	0.05	0.02
NOCA1 vs. SNPA1						
Components	**KT Stats**	**All**	**1**	**2**	**3**	**4**
SO4	**Slope**	0.80	0.67	0.70	0.76	0.70
	Intercept	-0.01	0.00	0.20	0.04	0.00
	MAD	0.17	0.09	0.26	0.36	0.09
NO3	**Slope**	0.17	0.04	0.39	0.32	0.00
	Intercept	0.03	0.05	0.03	0.03	0.05
	MAD	0.05	0.03	0.05	0.06	0.03
OC	**Slope**	0.69	0.35	0.70	0.77	0.31
	Intercept	-0.01	0.06	0.01	0.18	0.15
	MAD	0.17	0.09	0.12	0.27	0.22
LAC	**Slope**	0.38	0.20	0.44	0.41	0.19
	Intercept	0.01	0.01	0.01	0.06	0.03
	MAD	0.04	0.02	0.03	0.06	0.04
Soil	**Slope**	0.68	0.52	0.70	0.76	0.36
	Intercept	0.01	0.02	0.03	0.00	0.03
	MAD	0.03	0.02	0.05	0.05	0.02
CM	**Slope**	0.69	0.16	0.54	0.81	0.39
	Intercept	0.30	0.22	0.58	0.69	0.32
	MAD	0.46	0.16	0.40	0.64	0.28
Sea Salt	**Slope**	0.00	0.00	0.00	0.00	0.08
	Intercept	0.00	0.00	0.00	0.00	0.00
	MAD	0.00	0.00	0.00	0.00	0.00

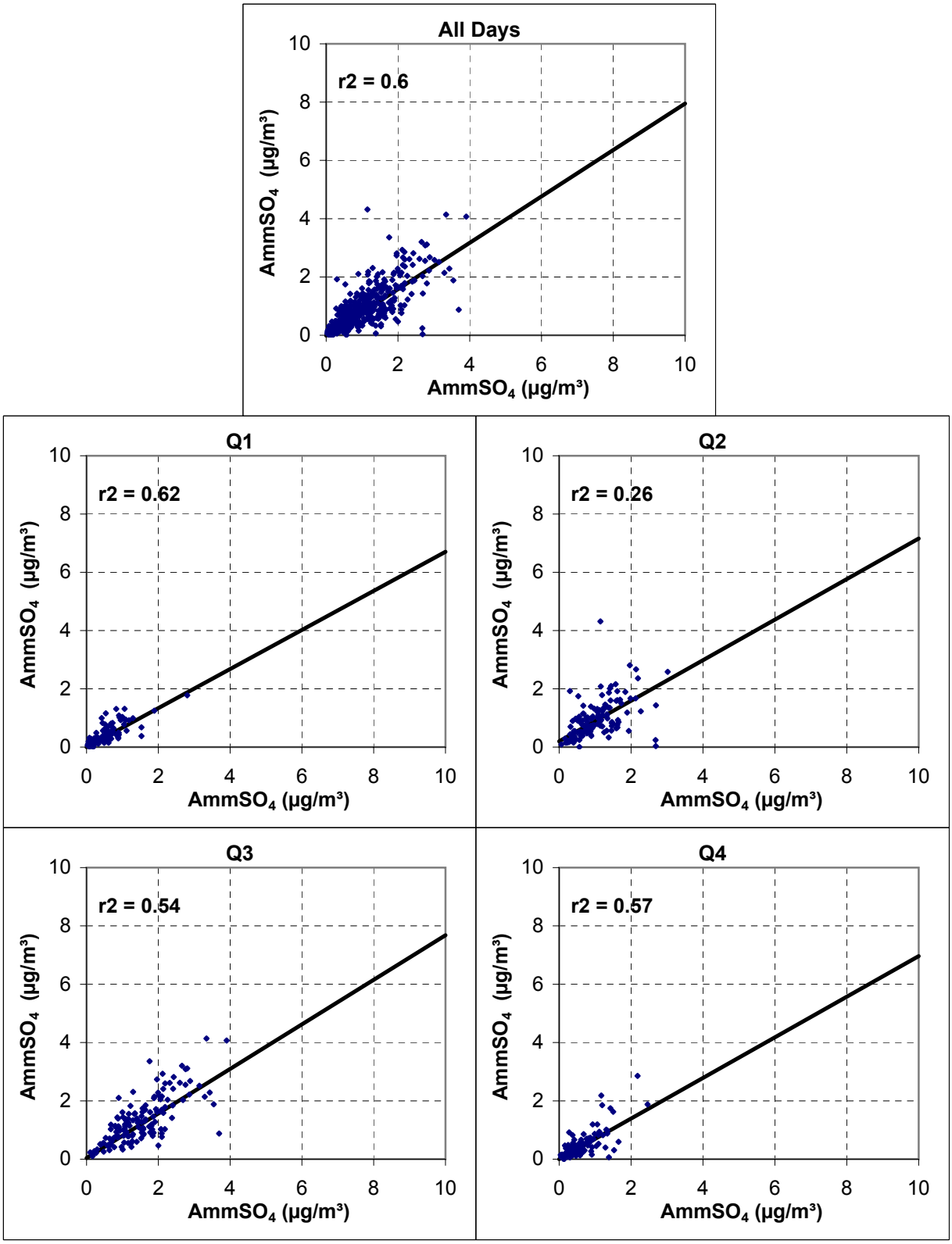
NOCA1
OC vs. org_H



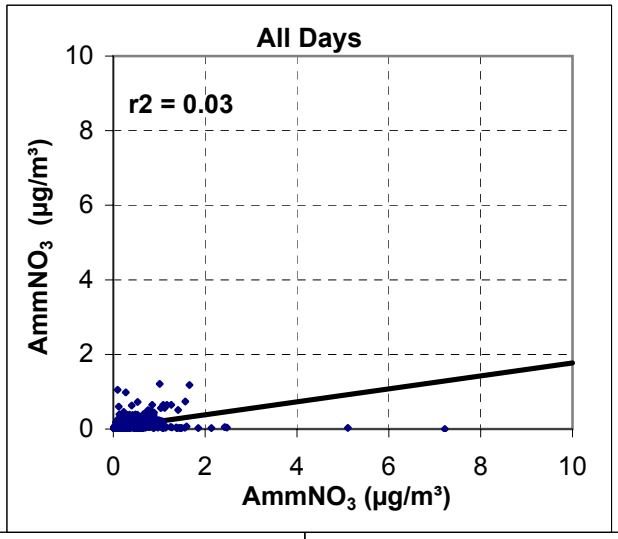
NOCA1
EC vs. OC

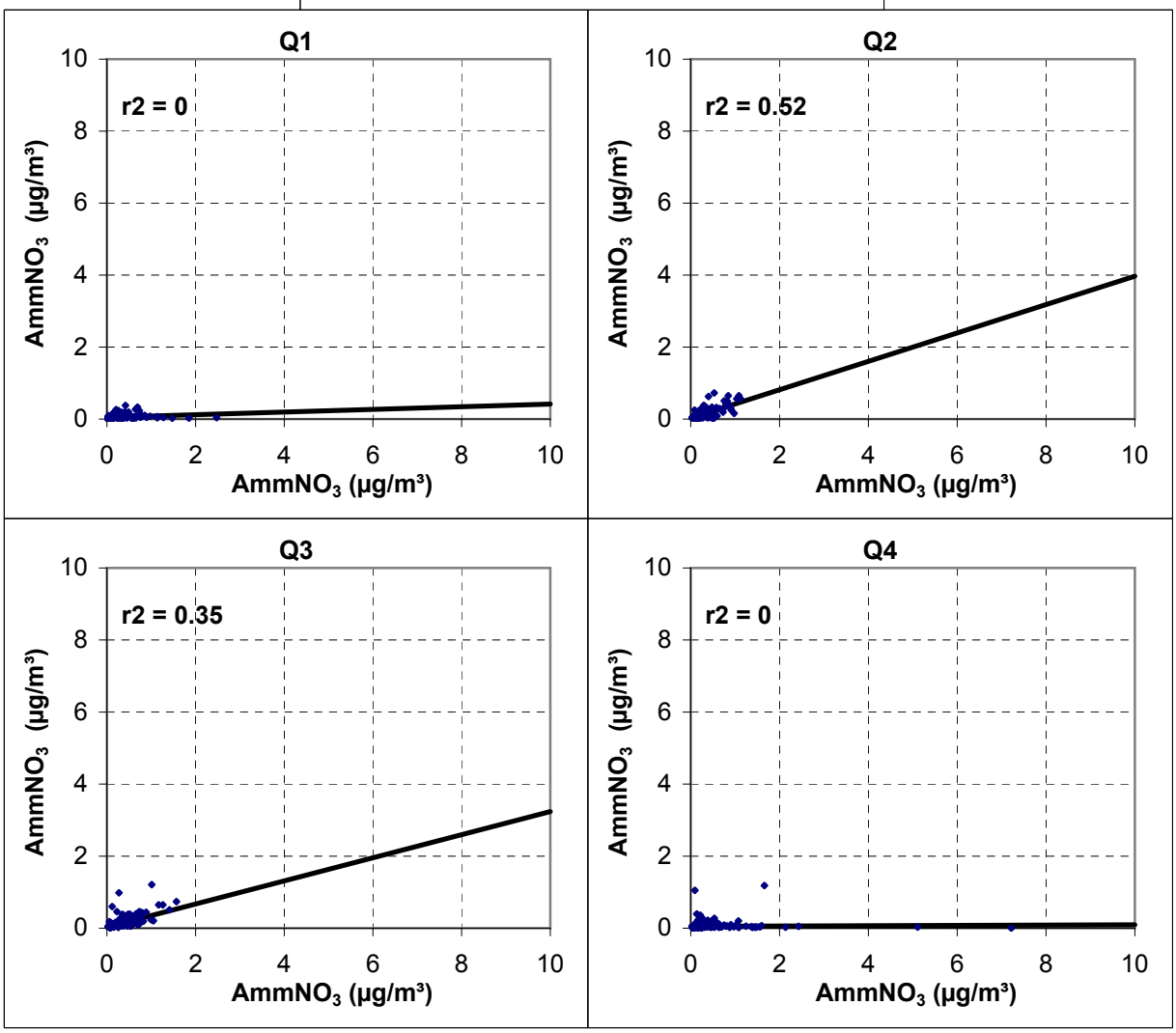


NOCA1 vs. SNPA1
Ammonium Sulfate

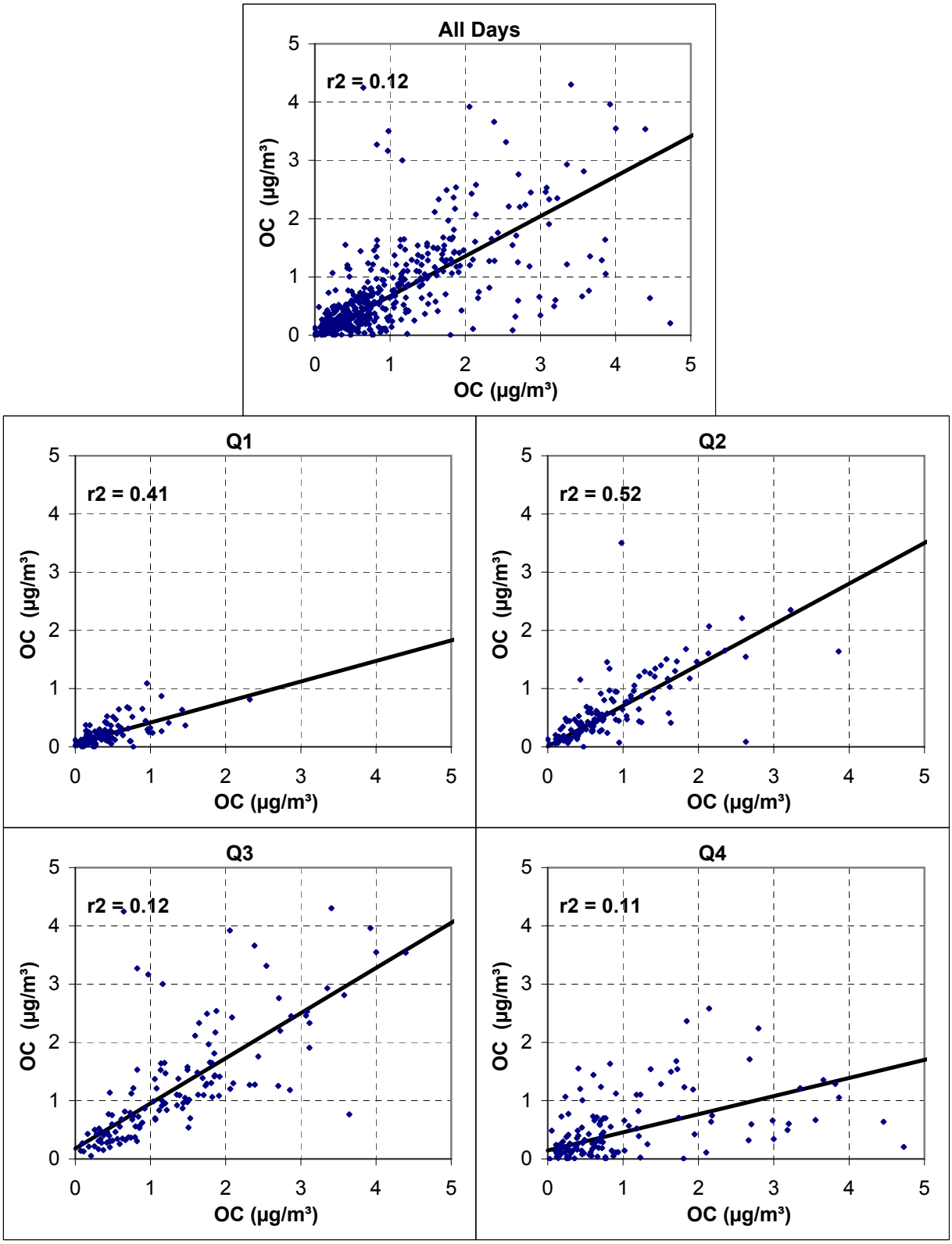


NOCA1 vs. SNPA1
Ammonium Nitrate

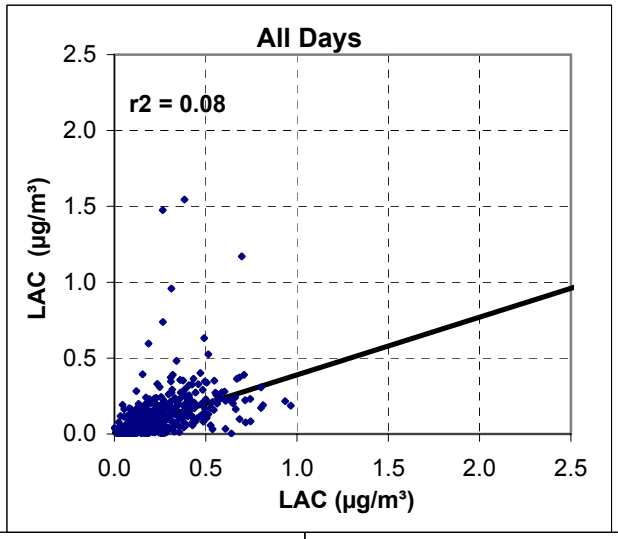


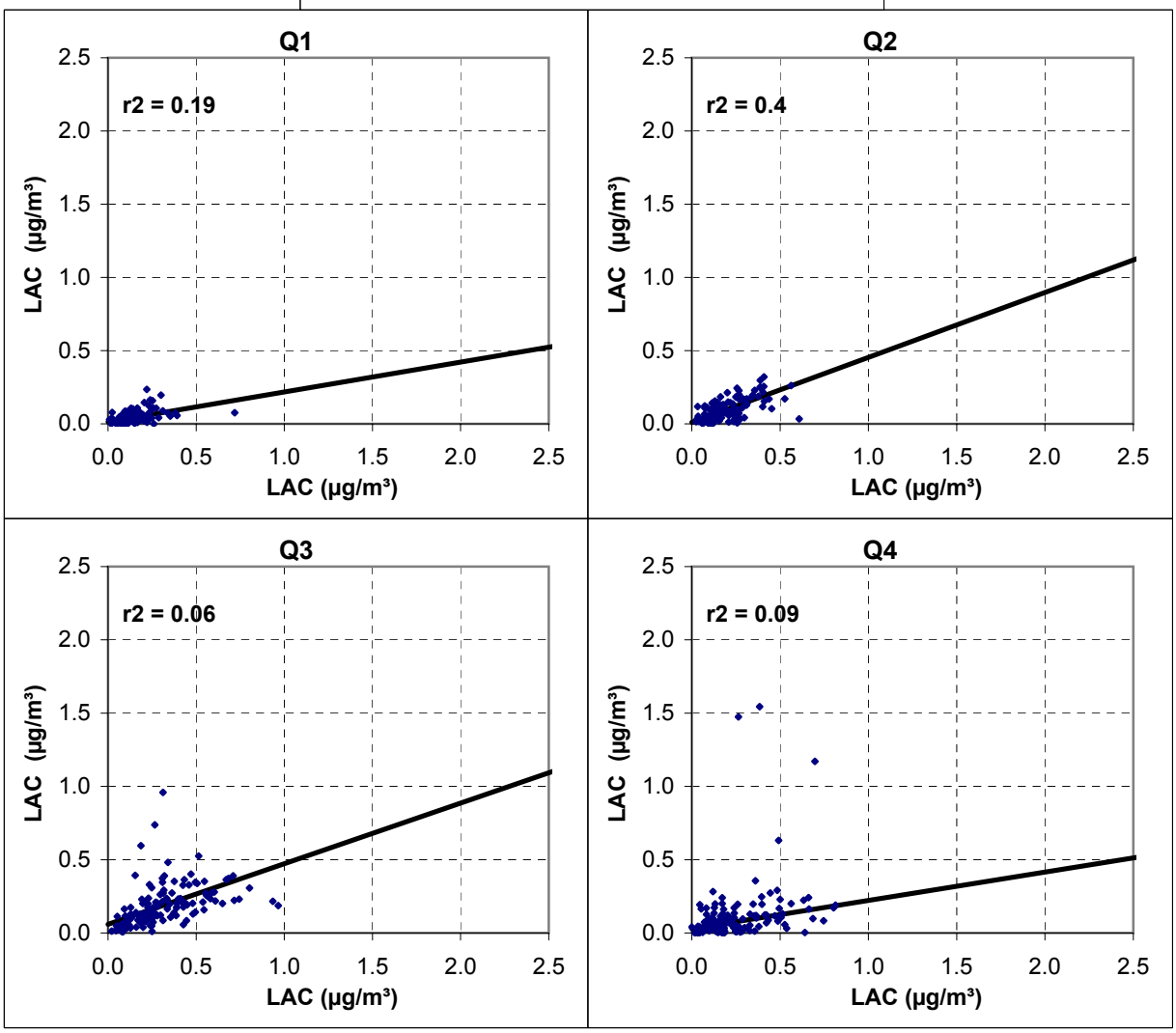


NOCA1 vs. SNPA1
Organic Carbon

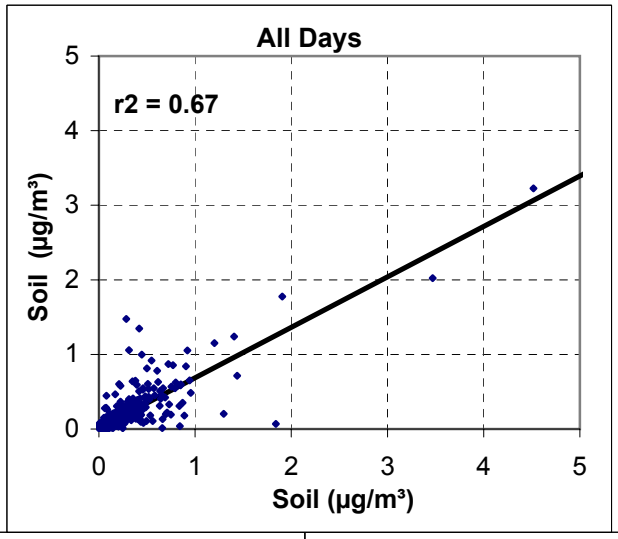


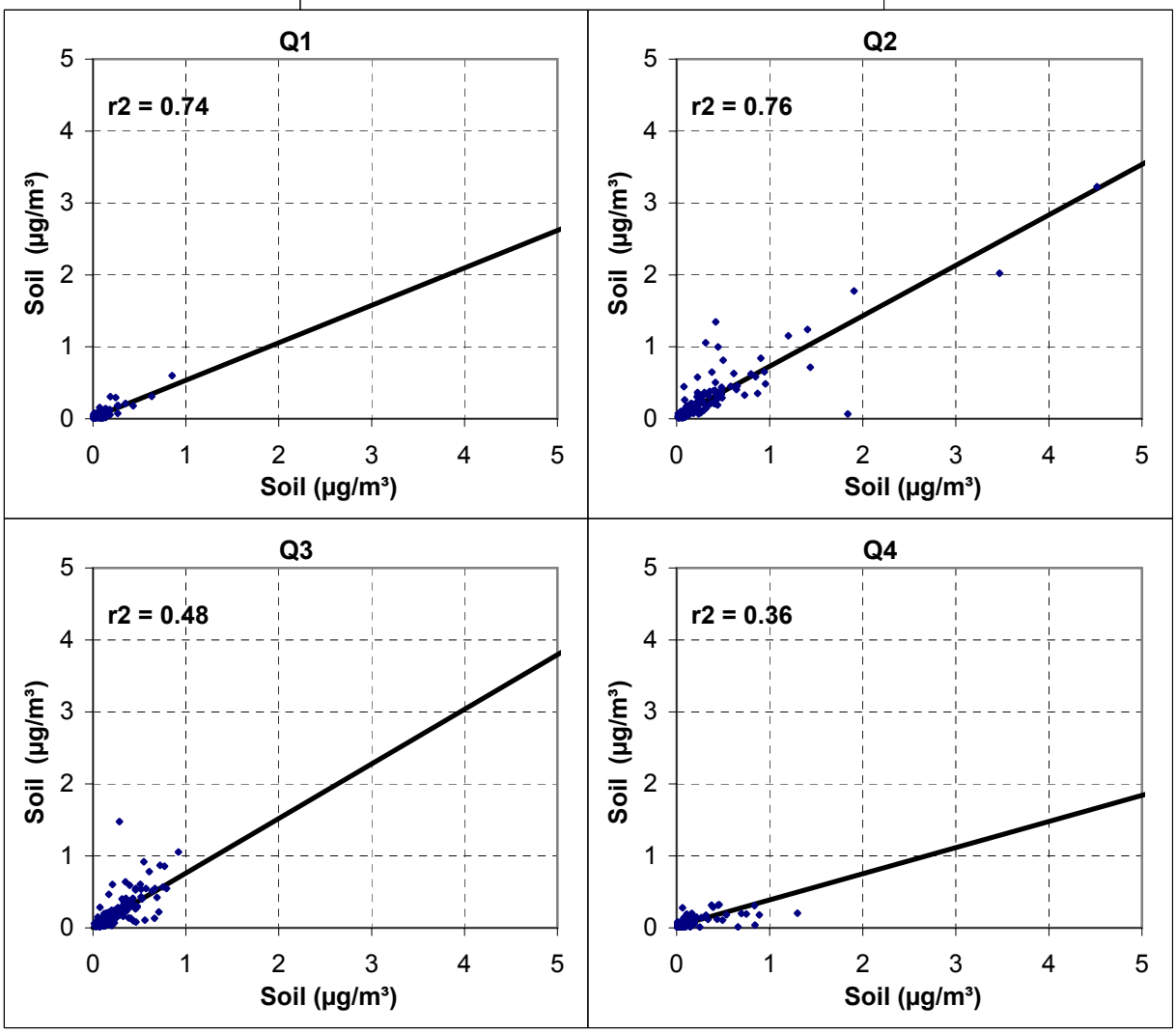
NOCA1 vs. SNPA1
Light Absorbing Carbon



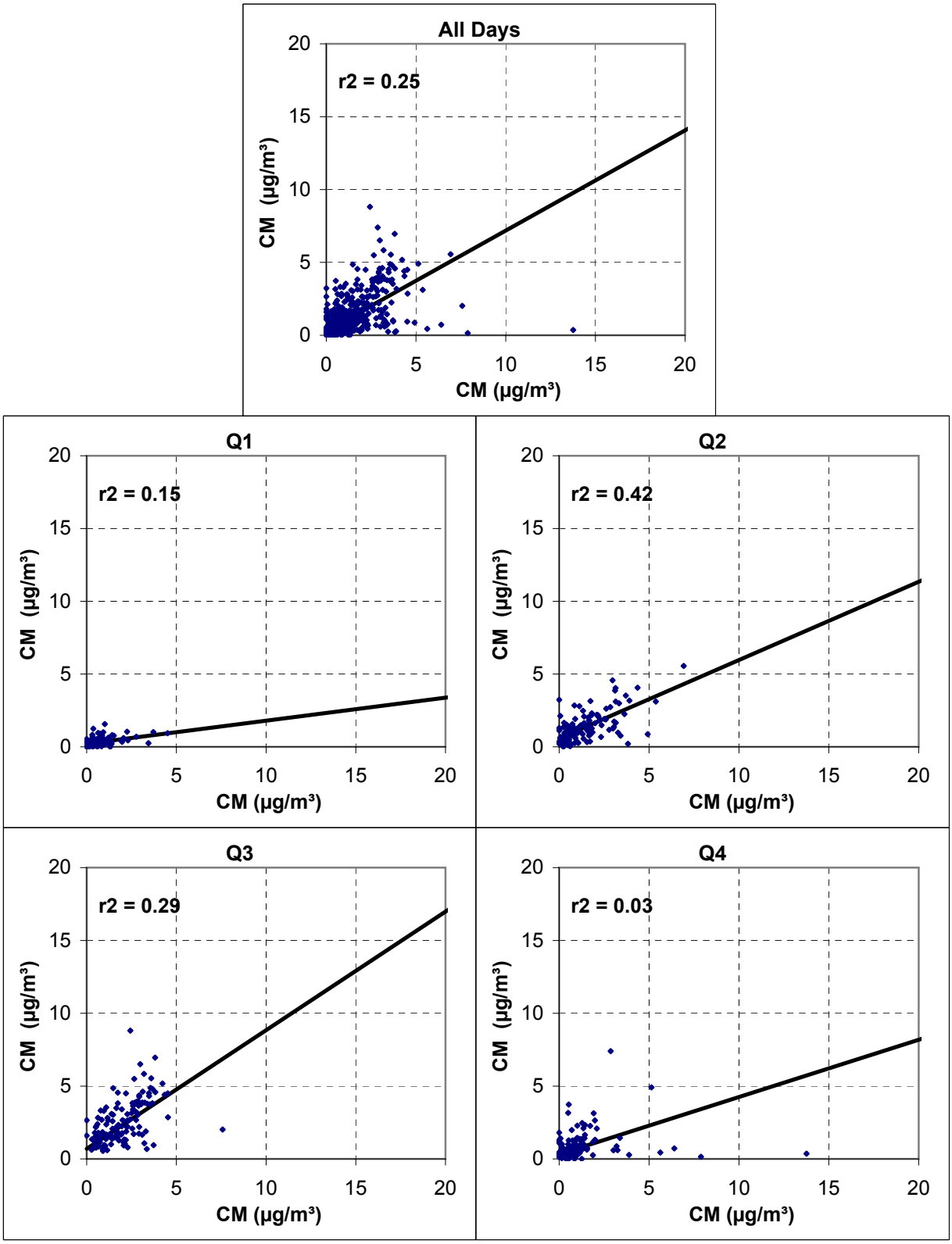
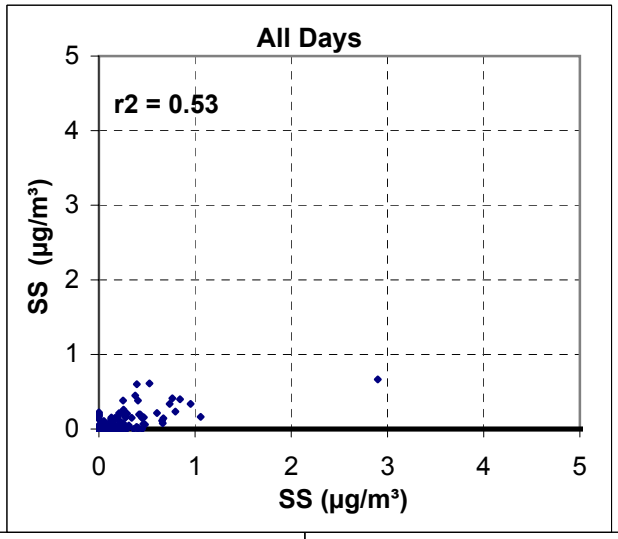
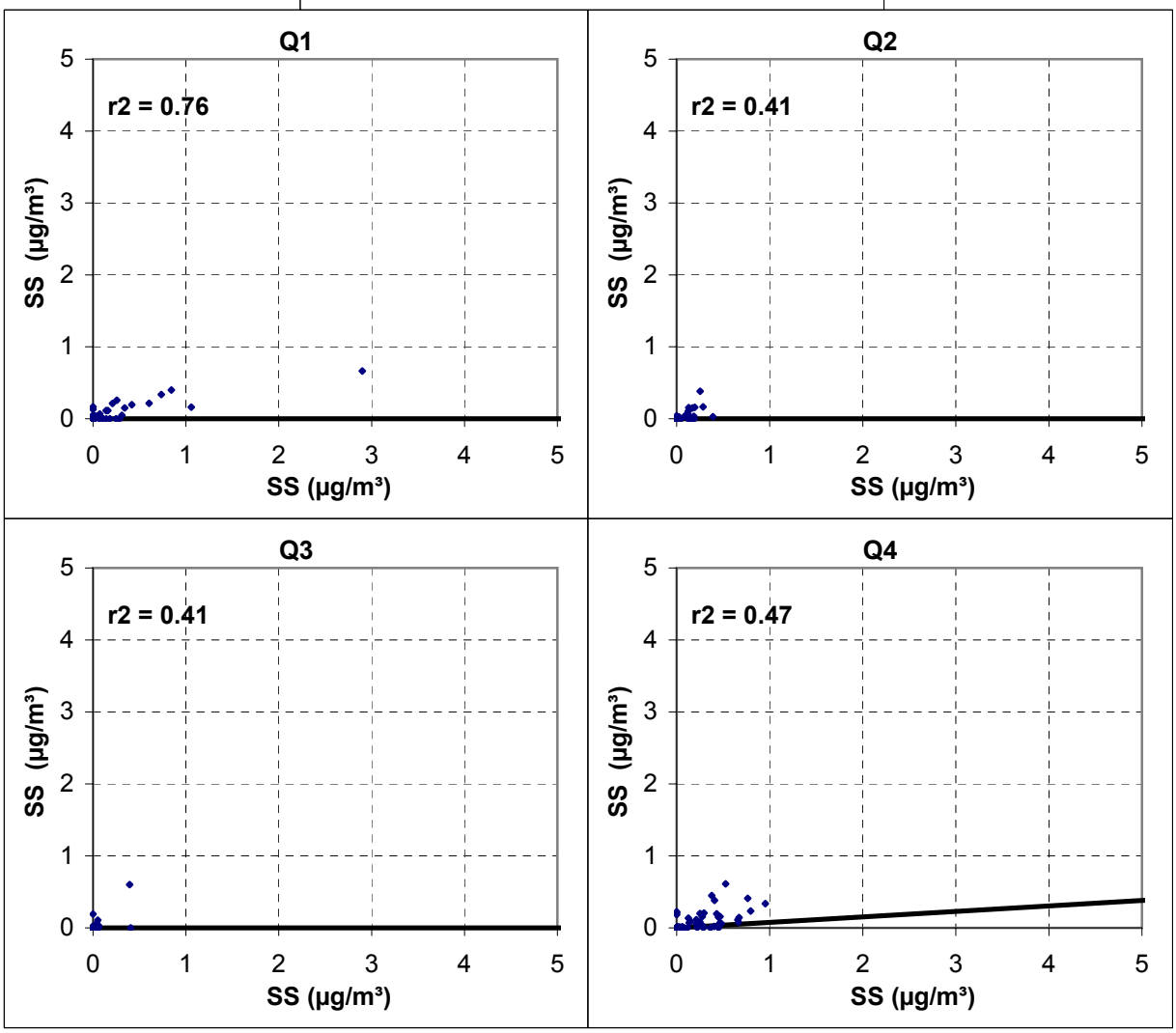


NOCA1 vs. SNPA1
Soil





NOCA1 vs. SNPA1
Coarse Mass



NOCA1 vs. SNPA1
Sea Salt