Supplement of Atmos. Chem. Phys., 14, 5089–5110, 2014
http://www.atmos-chem-phys.net/acp-14-5089-2014/
doi:10.5194/acp-14-5089-2014-supplement
© Author(s) 2014. CC Attribution 3.0 License.

Atmospheric Chemistry and Physics
Open Access

Supplement of

Primary and secondary biomass burning aerosols determined by proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy during the 2008 EU-CAARI campaign in the Po Valley (Italy)

M. Paglione et al.

Correspondence to: M. Paglione (m.paglione@isac.cnr.it)

SUPPLEMENT

Criteria for choosing factor number in NMR factor analysis

Several mathematical metrics could be used to aid determination of factor number, even if metadata analysis is also used to this aim (e.g., Lanz et al. 2008):

- *Q-value Analysis*

A first standardized criterion is the calculation of **Q-value**, the total sum of the squares of scaled residuals (Paatero et al., 2002). Q is expected to decrease with the number of factor. However, spurious solutions provide only minor decreases in Q, whereas genuine factors explain a significant fraction of the total variance and their inclusion is generally reflected by a marked decrease in Q. Therefore, the visual inspection of the curve Q versus number of factors often provides a straightforward manner to highlight to number of "genuine factors" (Paatero and Tapper, 1993). In this study, the Q-values for the NMR factor analysis (averaged between all methods, Fig. S1) suggest that a number of factors higher than five does not significantly improve the goodness of fit.

- *Principal Components Analysis (PCA)*

Preliminary PCA can be run to identify the best number of factors as a function of explained variance. For the SPC 2008 IOP NMR-dataset, the PCA model with 5 factors already explains 93.9% of the total variance. A sixth factor explains only a further 1.8% indicating a probable spurious solution.

- *Uniqueness of NMR spectral profiles and contribution*

As discussed in the text, the number of factors (p) was chosen to be 5 for the NMR dataset (Fig. 5 and Fig. 6) but solutions resulting from p from 2 up to 8 were explored with all 4 factor analysis algorithms listed in the text. Comparisons between results from different algorithms were made, both to evaluate any differences between the models and to try to determine the best number of factors that can decompose the NMR-data, supposing their best agreement around the right number of factors.

In the solution with p = 6 (Fig. S2), the spectra of two-three of the factors were very similar and difficult to distinguish from each other (e.g. F3, F6 and F4 in Fig. S2). Table S1 showed the correlation between profiles for the p=4 solution: very high values of correlation coefficient (R) between F3 and F6, F4 and F6, F3 and F4 (respectively 0.91, 0.98 and 0.84) suggested that, probably, division into six factors was forced based on the idea that if two profiles have a high correlation are not well defined from the factorization process. Even for p=5 the correlation between F3 and F4 was high (0.90), but a more detailed analysis of the spectrum and of their time series provided a different interpretation of these two profiles.

Moreover results from the different algorithms show a significant deterioration of their agreement from the p=6 solution with respect to that from p=5, confirming the chosen number of factor.

Table S1: Pearson correlation coefficients (R) between spectral profiles of NMR-factors for p=5 and p=6 solutions. In yellow were highlighted correlation coefficients higher than 0.8.

p=5

R	F1	F2	F3	F4	F5
F1	1				
F2	0.22	1			
F3	0.55	0.60	1		
F4	0.50	0.65	0.90	1	
F5	0.62	0.47	0.85	0.76	1

p=6

R	F1	F2	F3	F4	F5	F6
F1	1					
F2	0.27	1				
F3	0.57	0.52	1			
F4	0.61	0.60	0.84	1		
F5	0.54	0.36	0.77	0.60	1	
F6	0.62	0.59	0.91	0.98	0.64	1

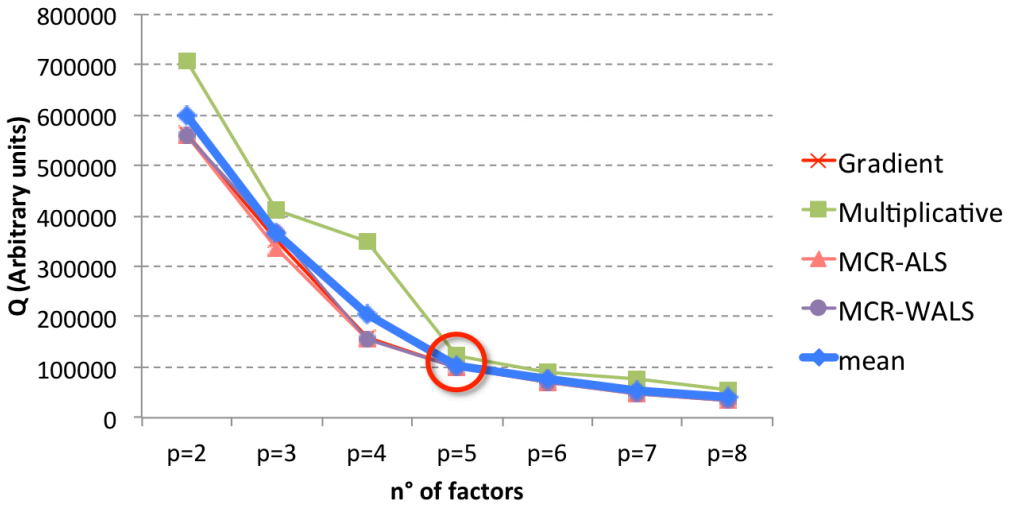

Figure S1: Q values versus the number of factors p. Blue line represents average values between four methods applied. Red circle denotes the chosen solution (p=5)**.**

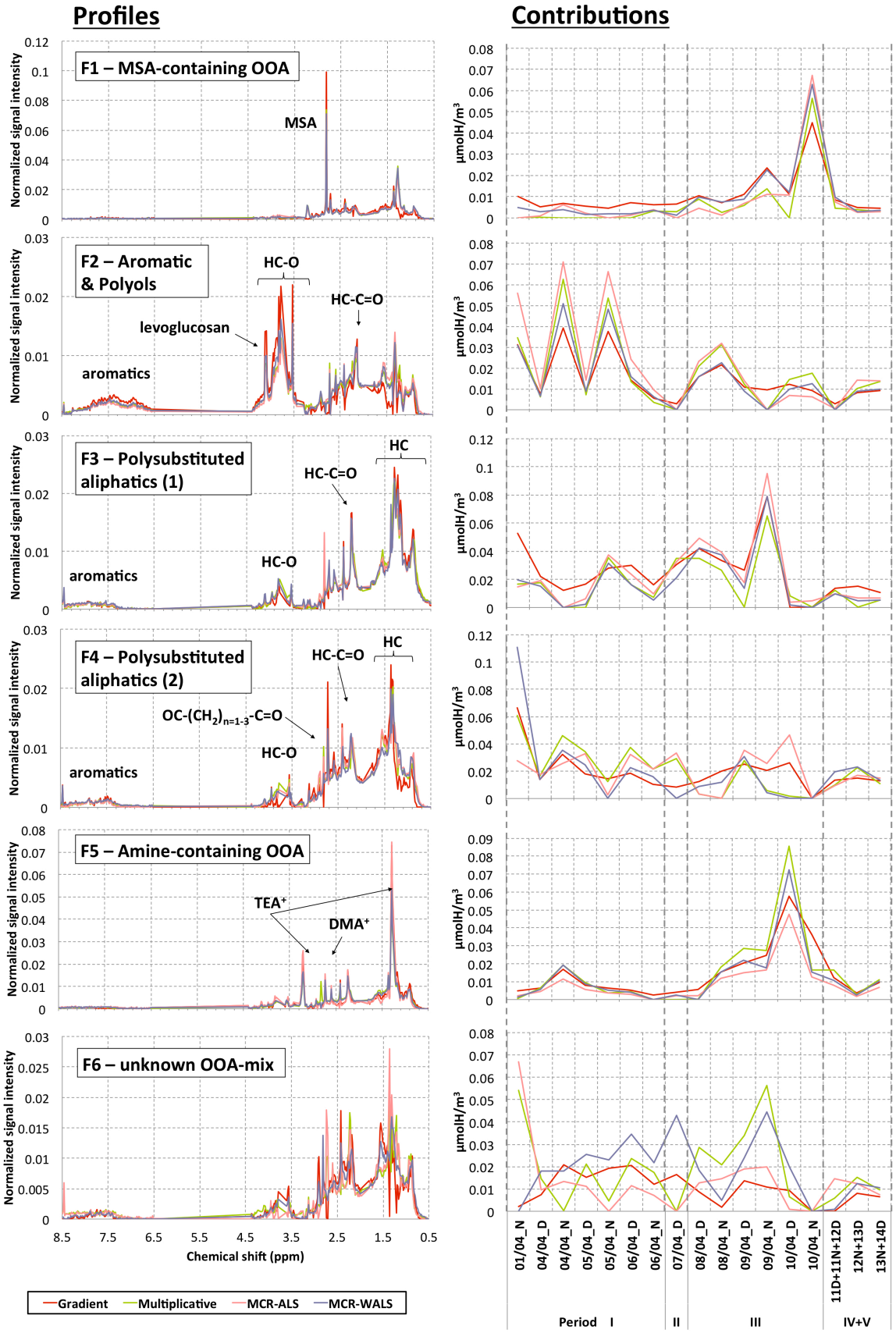

Figure S2: 6-factors solution for the NMR-dataset, spectral profiles (left side) and time series (right side). Results from all 4 different algorithms: Projected Gradient (red line), Multiplicative (green line), MCR-ALS (pink line) and MCR-WALS (blu line) methods.