Supplement of Atmos. Chem. Phys., 16, 1729–1746, 2016
http://www.atmos-chem-phys.net/16/1729/2016/
doi:10.5194/acp-16-1729-2016-supplement

Supplement of

What do correlations tell us about anthropogenic–biogenic interactions and SOA formation in the Sacramento plume during CARES?

L. Kleinman et al.

Correspondence to: L. Kleinman (kleinman@bnl.gov)

S1. Determination of statistical significance

An F-test can be used to determine the significance (p value) of the increase in R^2 that accompanies an additional explanatory variable. The F test statistic for determining the significance of adding Bio to the OA vs. CO regression is

$$F_{BIO} = (R_4^2 - R_1^2)/(k-g) / ((1-R_4^2)/(N - (k+1)) \quad \text{(S1)}$$

where k=2 and g=1 are the number of explanatory variables in models M4 and M1 and N is the number of data points. As statistical significance varies in the same direction as F, it can be seen from Eq. S1 that significance levels for the improvement in R_4^2 caused by adding Bio to the OA vs. CO regression decrease as the difference between R_4^2 and R_1^2 becomes small and as N becomes small. The F statistic, F_{CO}, for adding CO to the Bio regression is obtained by replacing R_1^2 with R_2^2 in Eq. S1.

Equation S1 is based on the assumption that data points are independent and residuals are normally distributed with equal standard deviations. Independence is the most serious concern as the effective number of independent data points, N_{eff}, can be many-fold fewer than N because data are from plumes which are coherent structures, much larger than the spacing between data points. Because of autocorrelation between data points (Trenberth, 1984; Eq. 2.8), the value of N in Eq. 2 is several-fold too large and statistical significance is over-estimated. It has been shown theoretically and by numerical examples that it is difficult to obtain a numerically stable value of N_{eff} and that simply changing N to N_{eff} in Eq. S1 does not generate a valid F statistic (Thiébaux and Zwiers, 1984; Trenberth, 1984). Corrections to the usual statistical tests have been proposed and implemented with success for test problems (e.g. Dale and Fortin, 2009). The scant reference to autocorrelation in the atmospheric chemistry literature leads us to believe that determining significance levels for regressions of plume constituents does not have a universal solution.

Dale, M.R., and M.-J. Fortin, Spatial autocorrelation and statistical tests: Some solutions, *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 188-206, doi:10.1198/jabes.2009.0012, 2009.

Thiébaux, H.J., and F.W. Zwiers, The interpretation and estimation of effective sample size, *J. Climate Appl. Meteor.*, 23, 800-811, 1984.

Trenberth, K.E., Some effects of finite sample size and persistence on meteorological statistics. Part 1: Autocorrelation, *Mon. Wea. Rev.*, 112, 2359-2368, 1984.

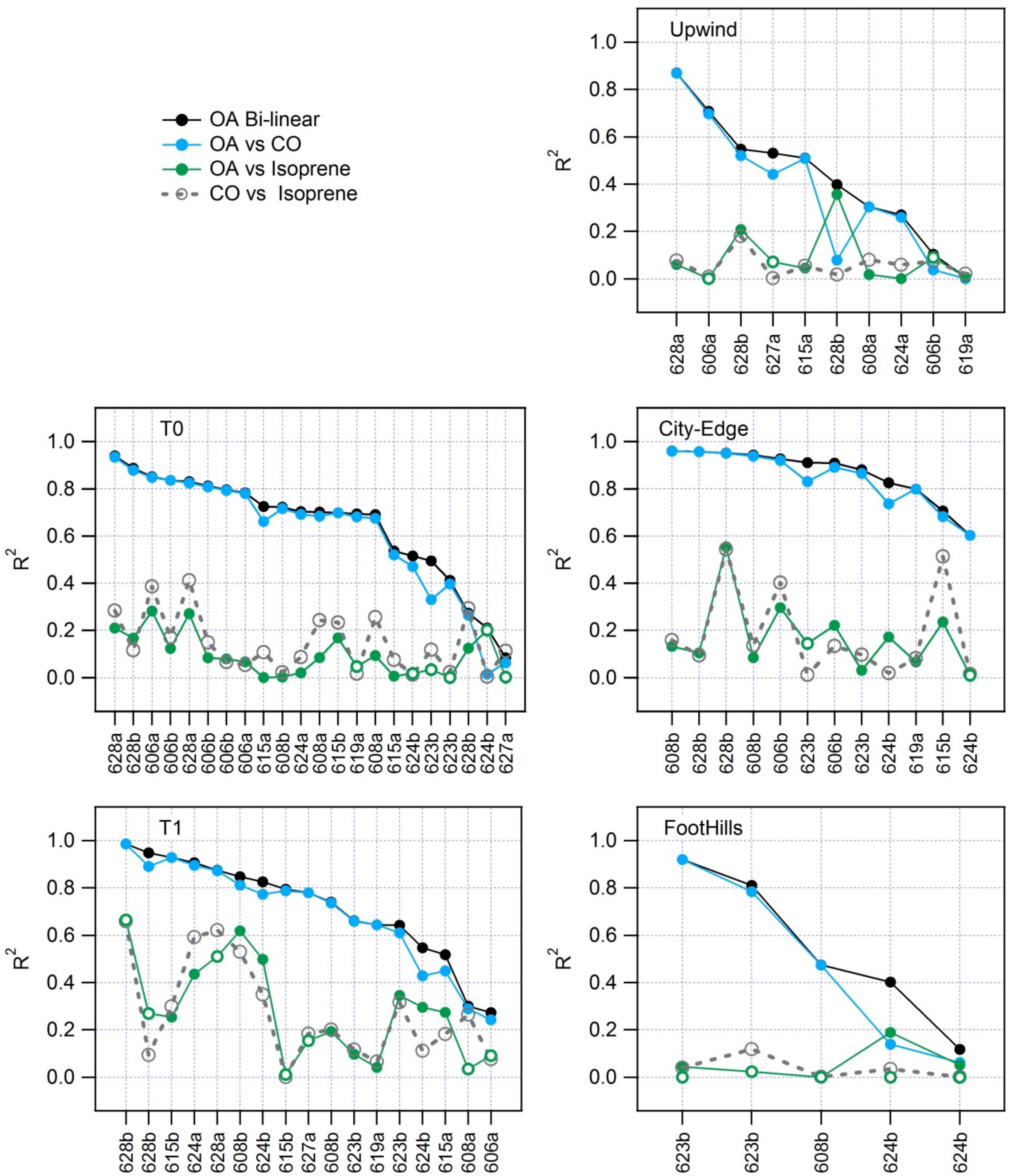

Figure S1. Coefficient of determination for transects at 5 locations, with explanatory variables, CO and isoprene. Results rank ordered according to R^2 of bilinear model, M4-isoprene. Legend in upper left panel. Open green circles indicates transects in which OA is anti-correlated with isoprene.

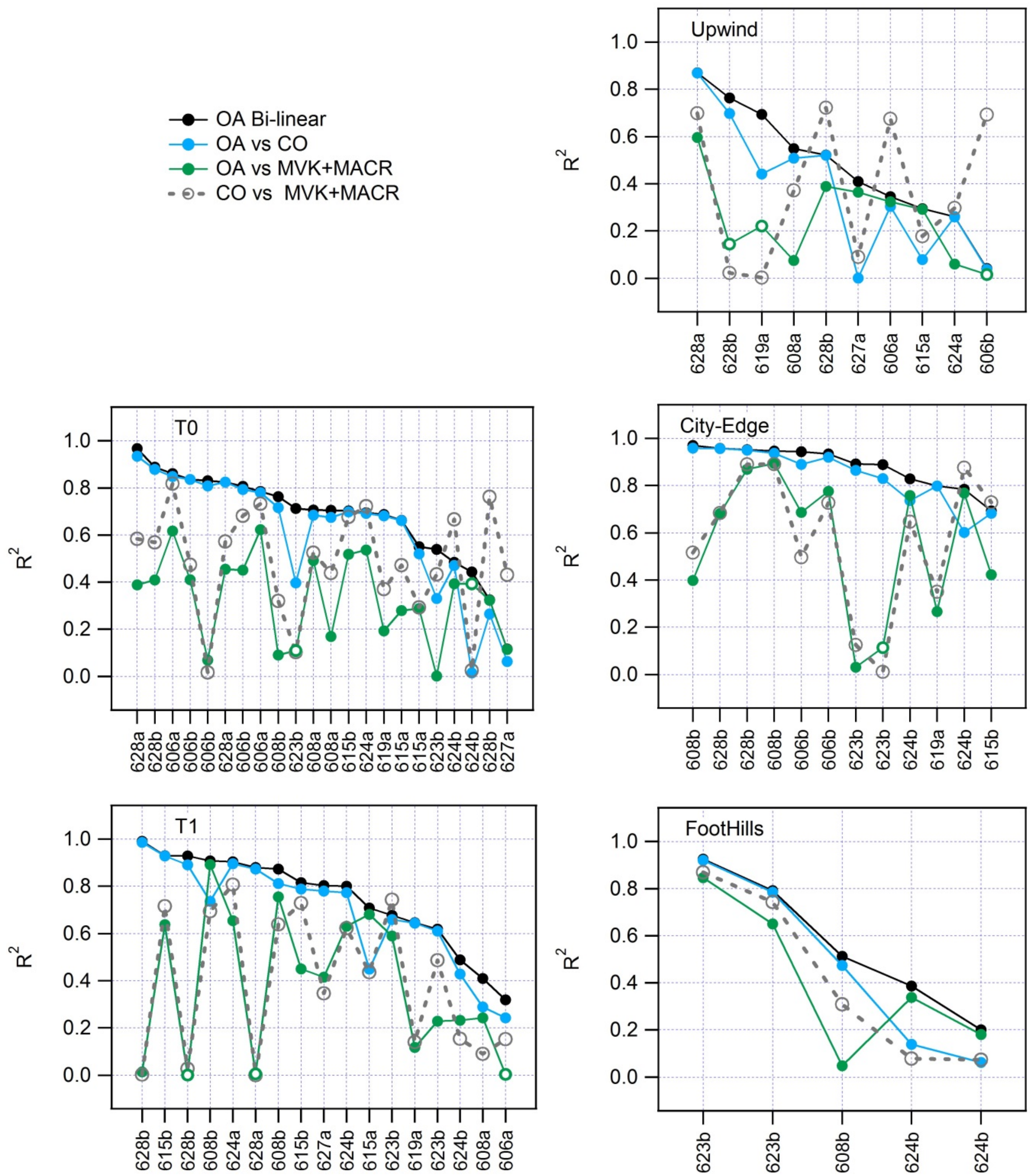

Figure S2. Coefficient of determination for transects at 5 locations, with explanatory variables, CO and MVK+MACR. Same Format as Fig. S1.

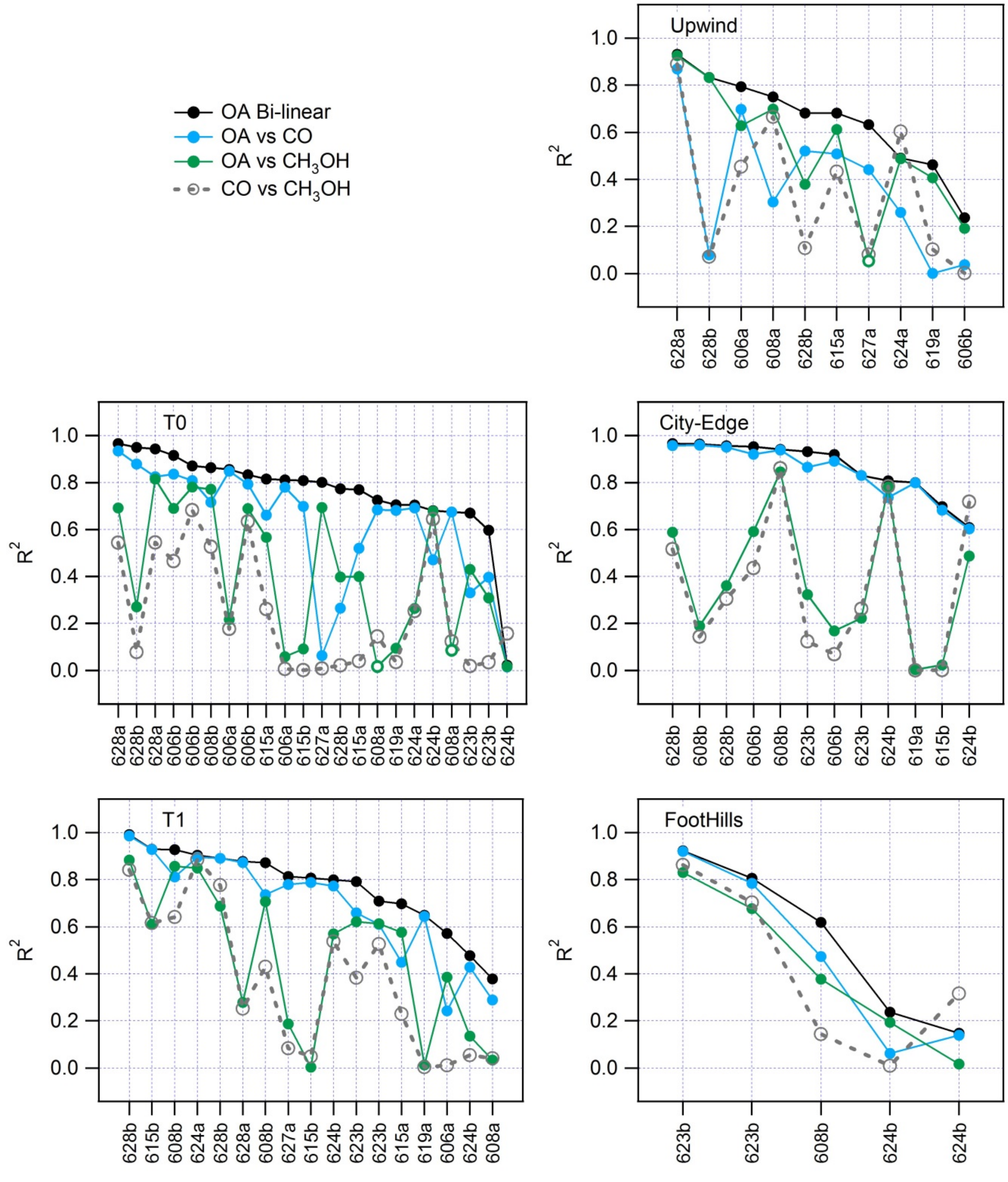

Figure S3. Coefficient of determination for transects at 5 locations, with explanatory variables, CO and CH_3OH. Same Format as Fig. S1.