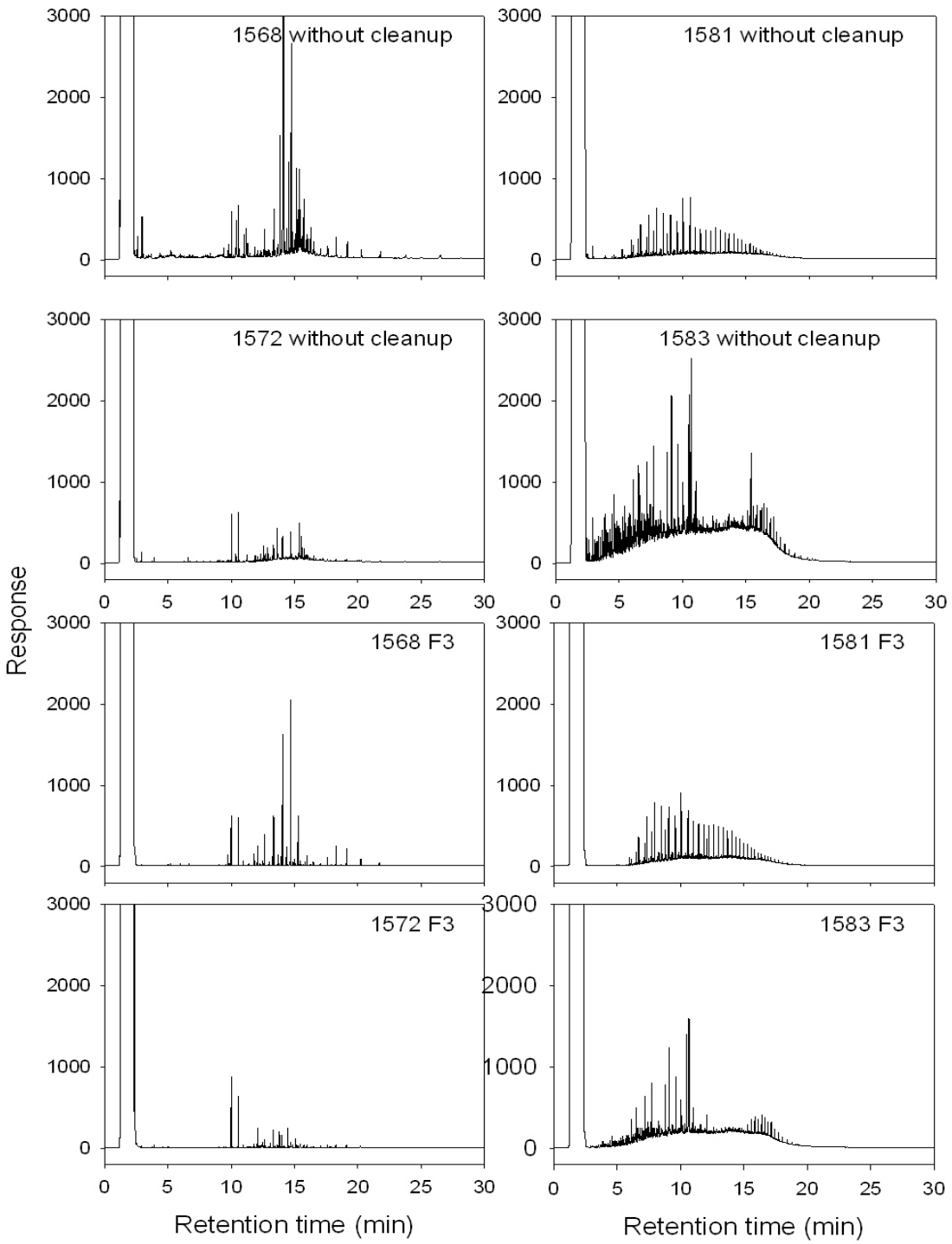


Figure S1 GC/FID chromatograms of two pairs of representative samples (1568 and 1581 for Swan Hills phytoremediation plant and oil/diesel contaminated soil; and 1572 and 1583 for Orphan Well Association background soil and highly oil contaminated soil) for determination of the total PHC or called TPH (***Top: 1A, without column cleanup; and Bottom: 1B, after column cleanup***).

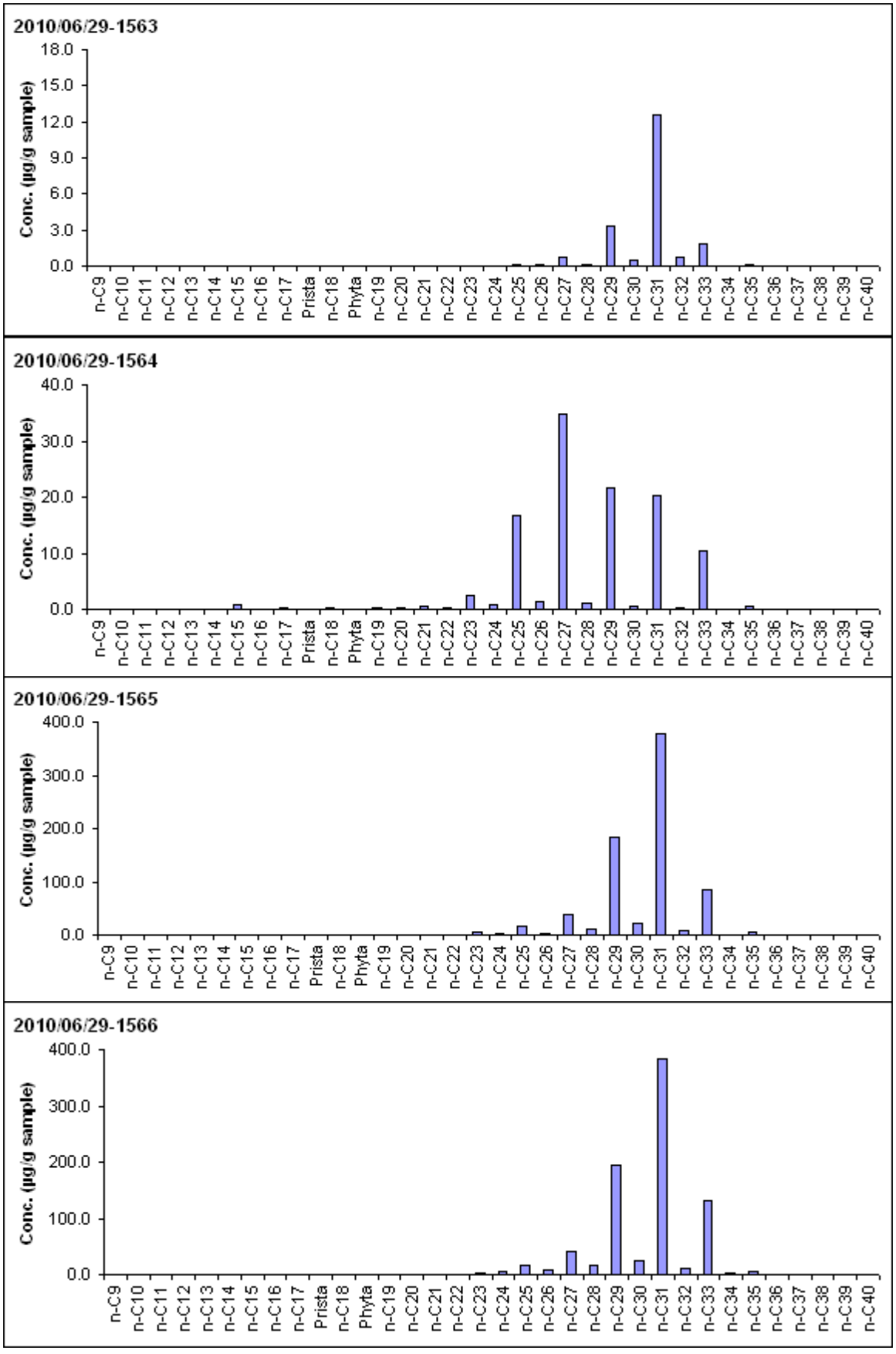
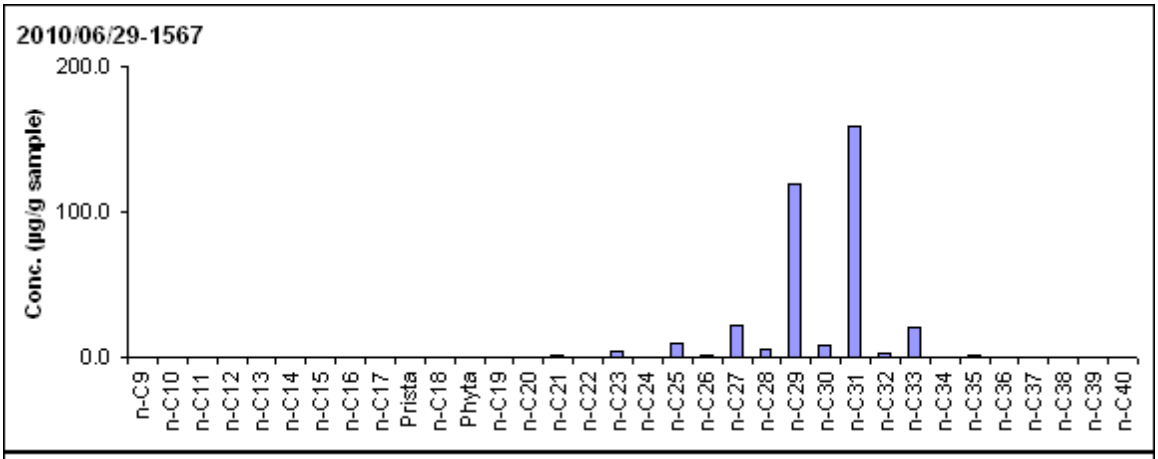
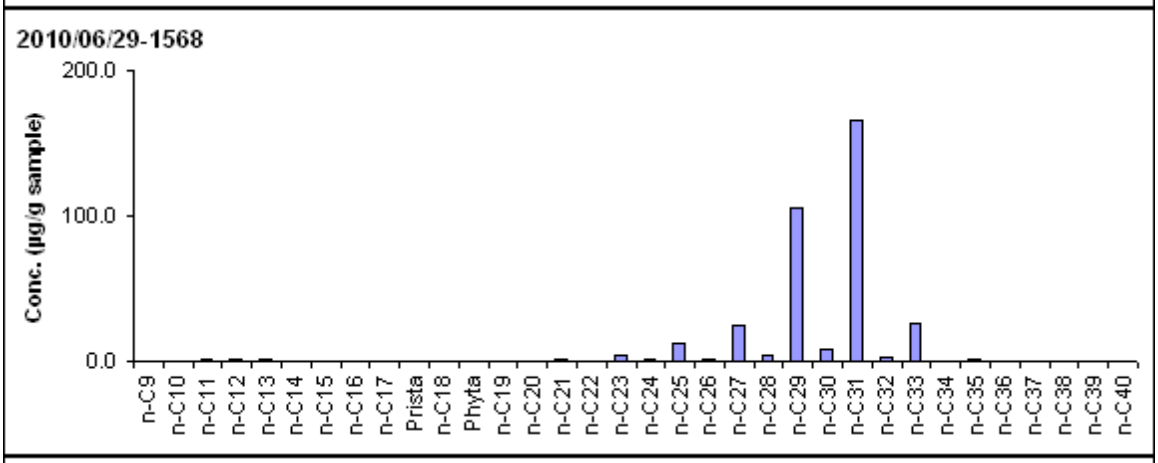
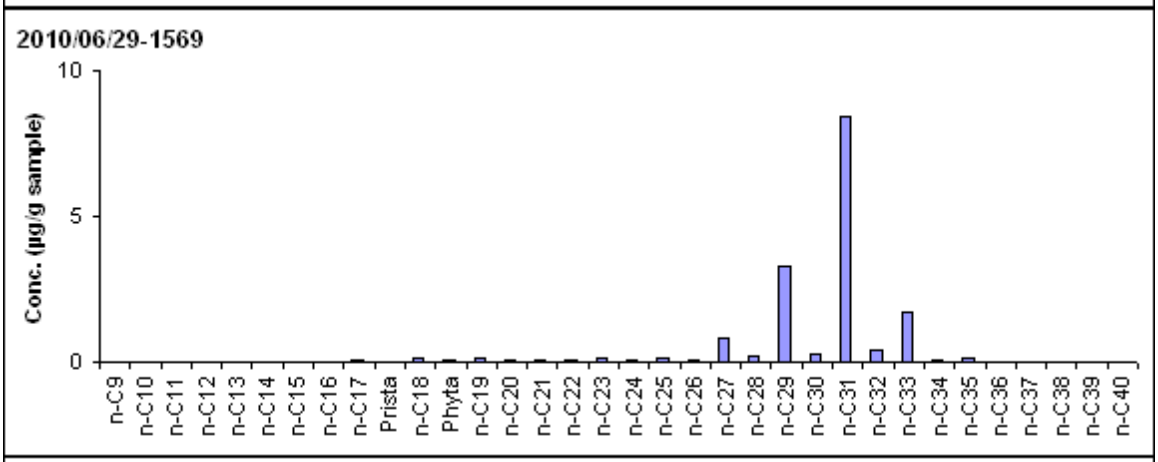


Figure S2 n-Alkanes distribution in samples







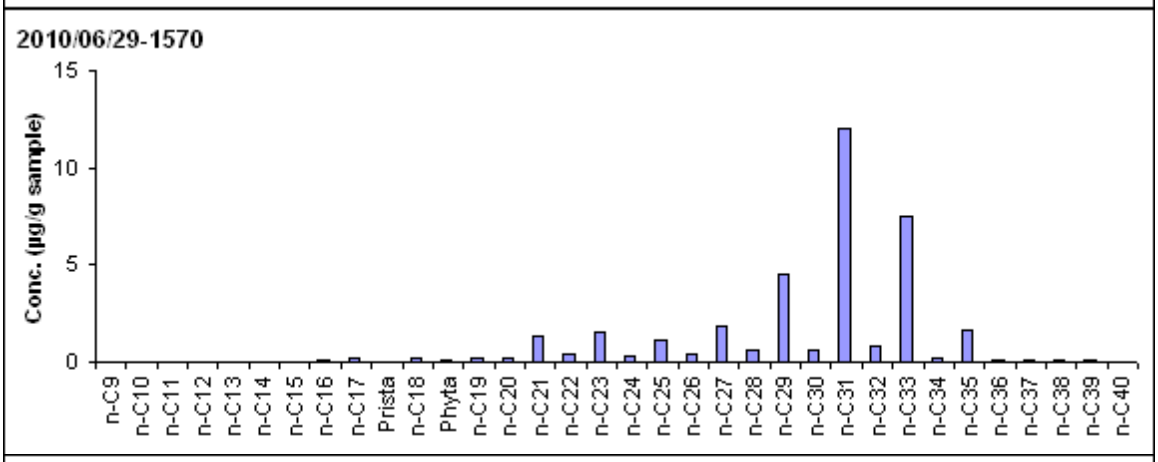


Figure S2 n-Alkanes distribution in samples

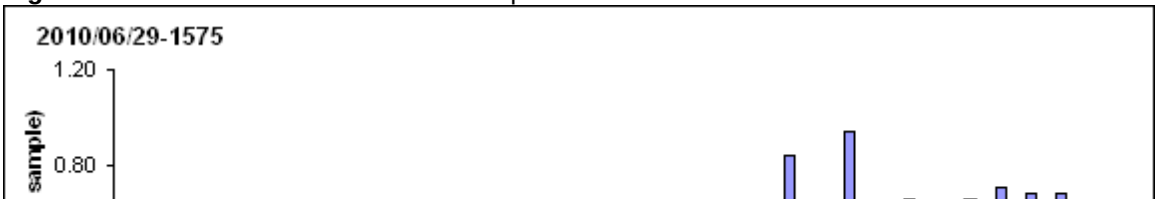


Figure S2 n-Alkanes distribution in samples

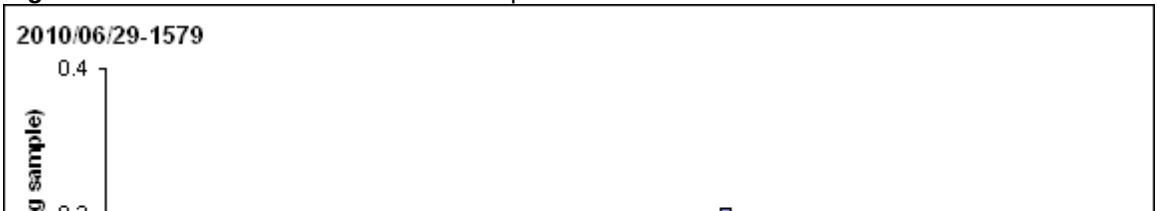


Figure S2 n-Alkanes distribution in samples

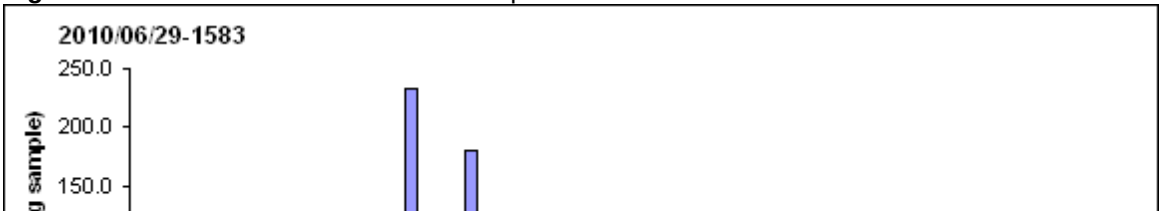


Figure S2 n-Alkanes distribution in samples

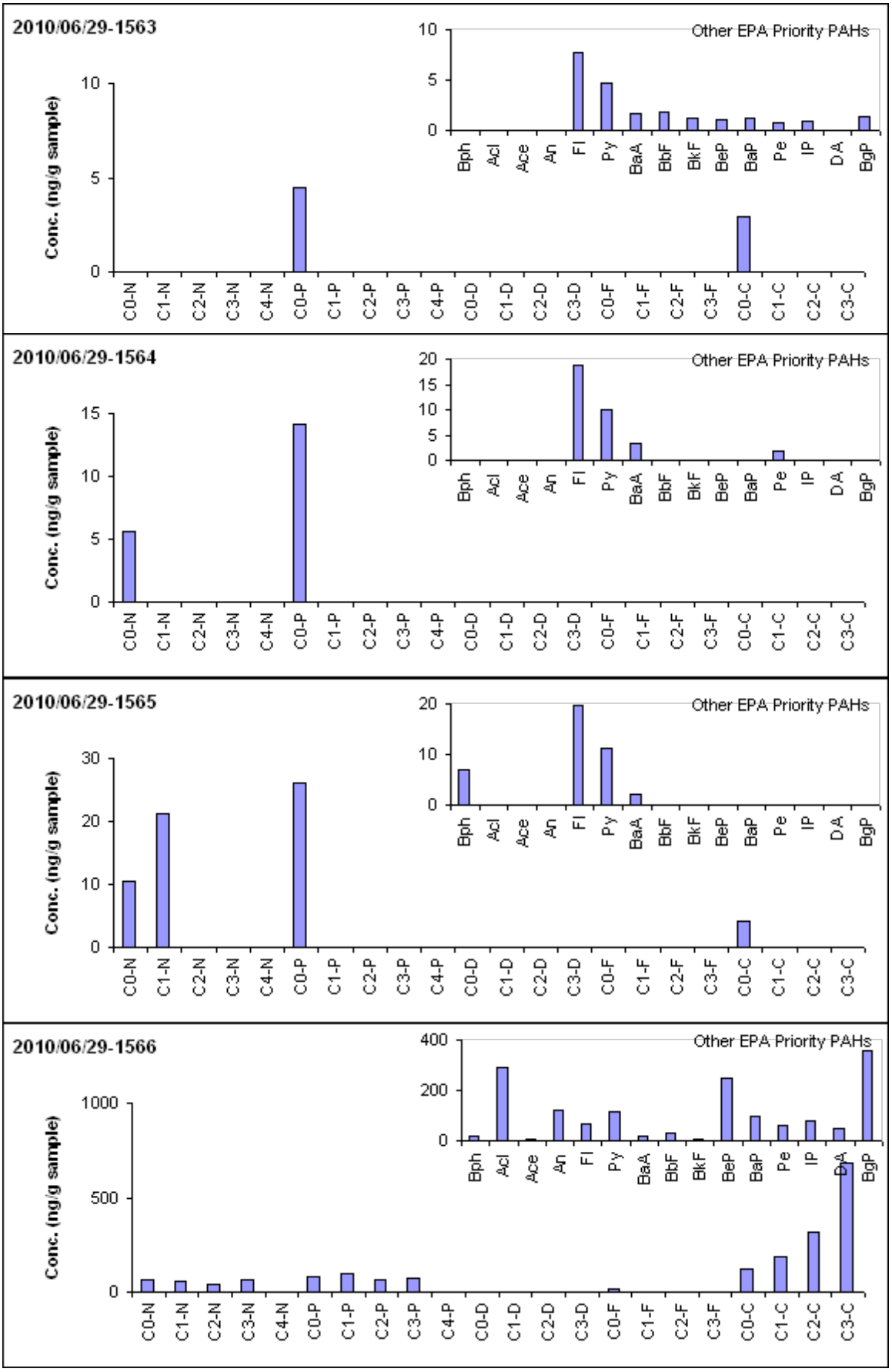


Figure S3 Distribution of target PAHs in samples

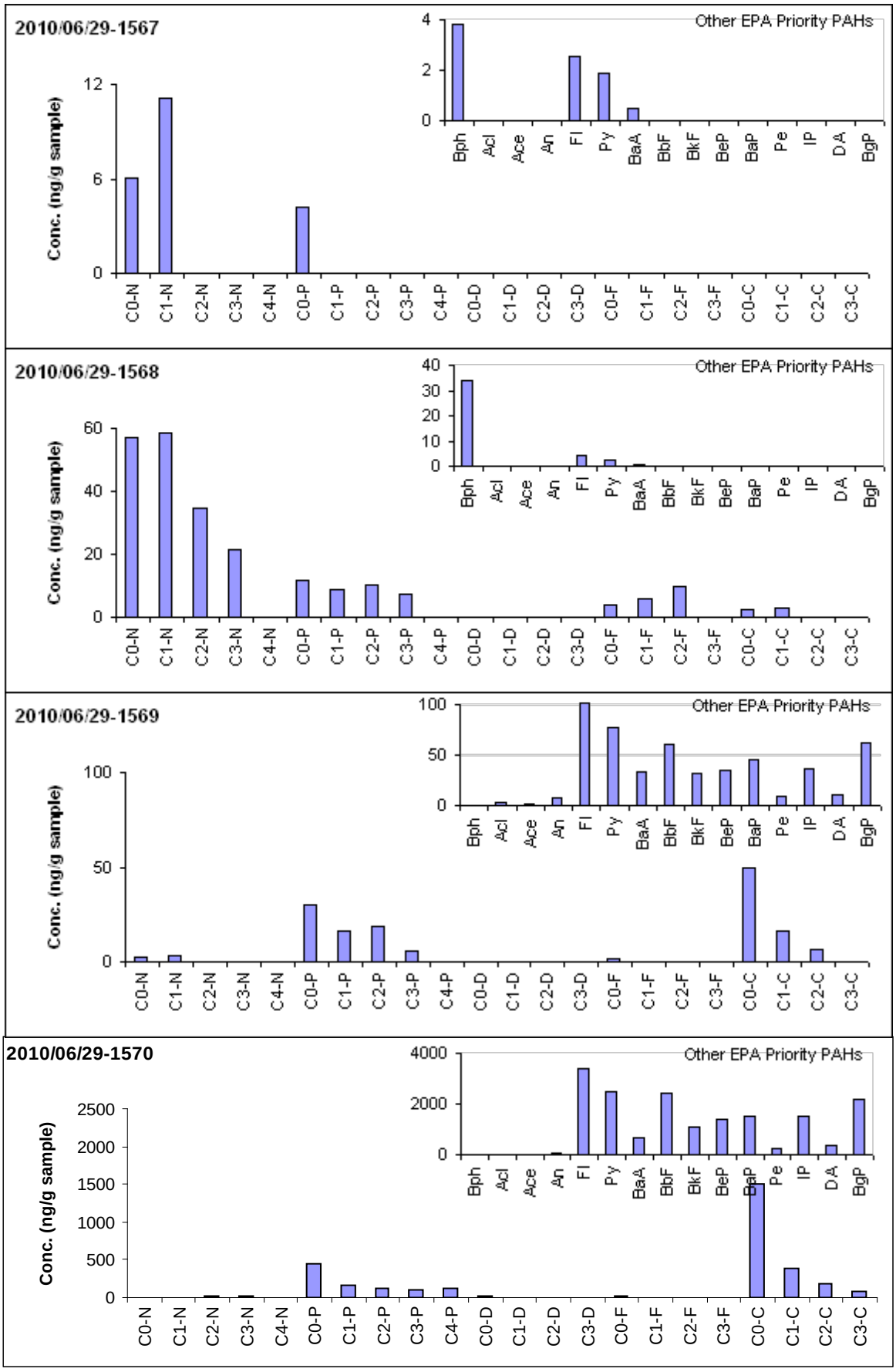


Figure S3 Distribution of target PAHs in samples

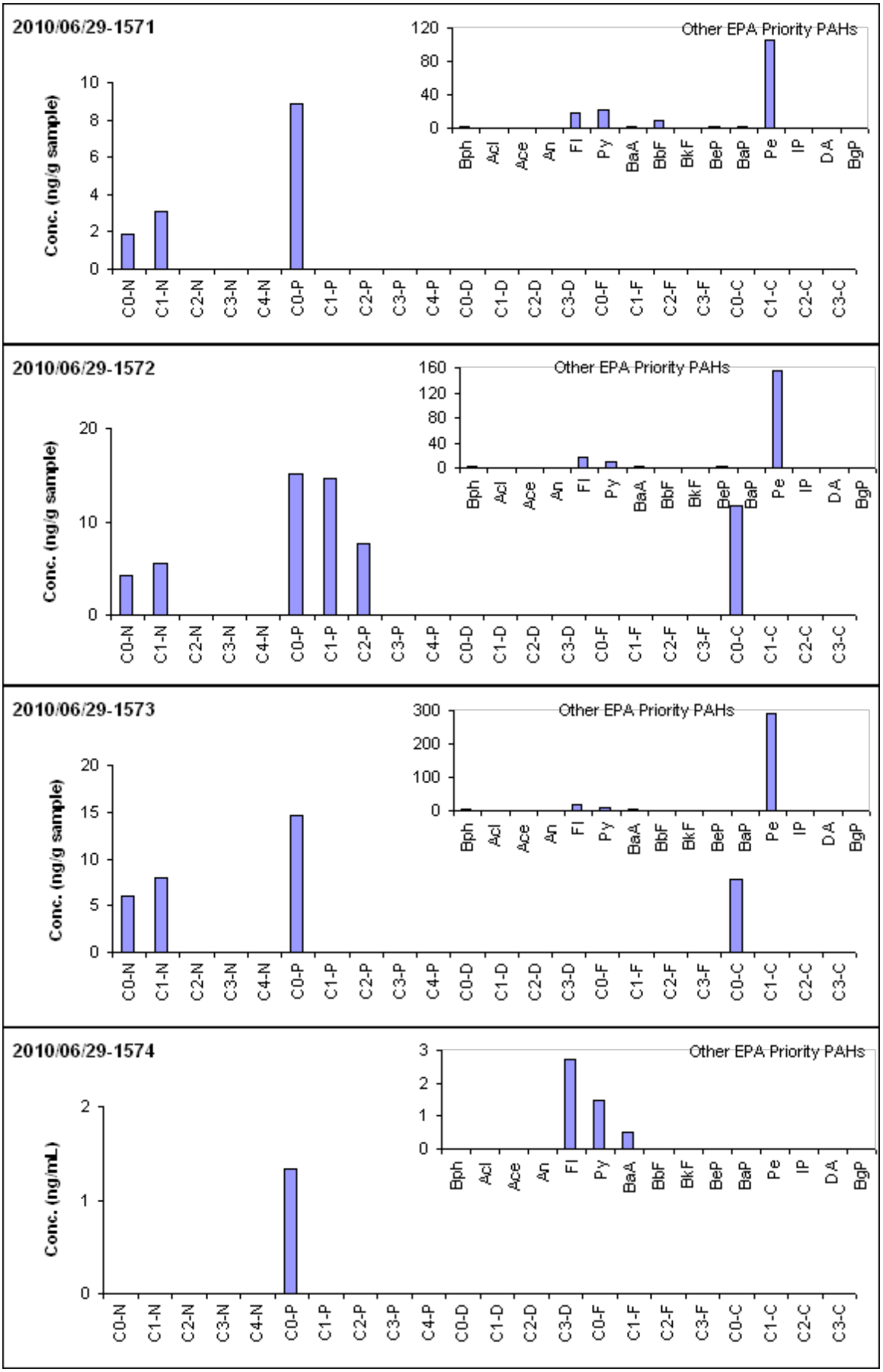


Figure S3 Distribution of target PAHs in samples

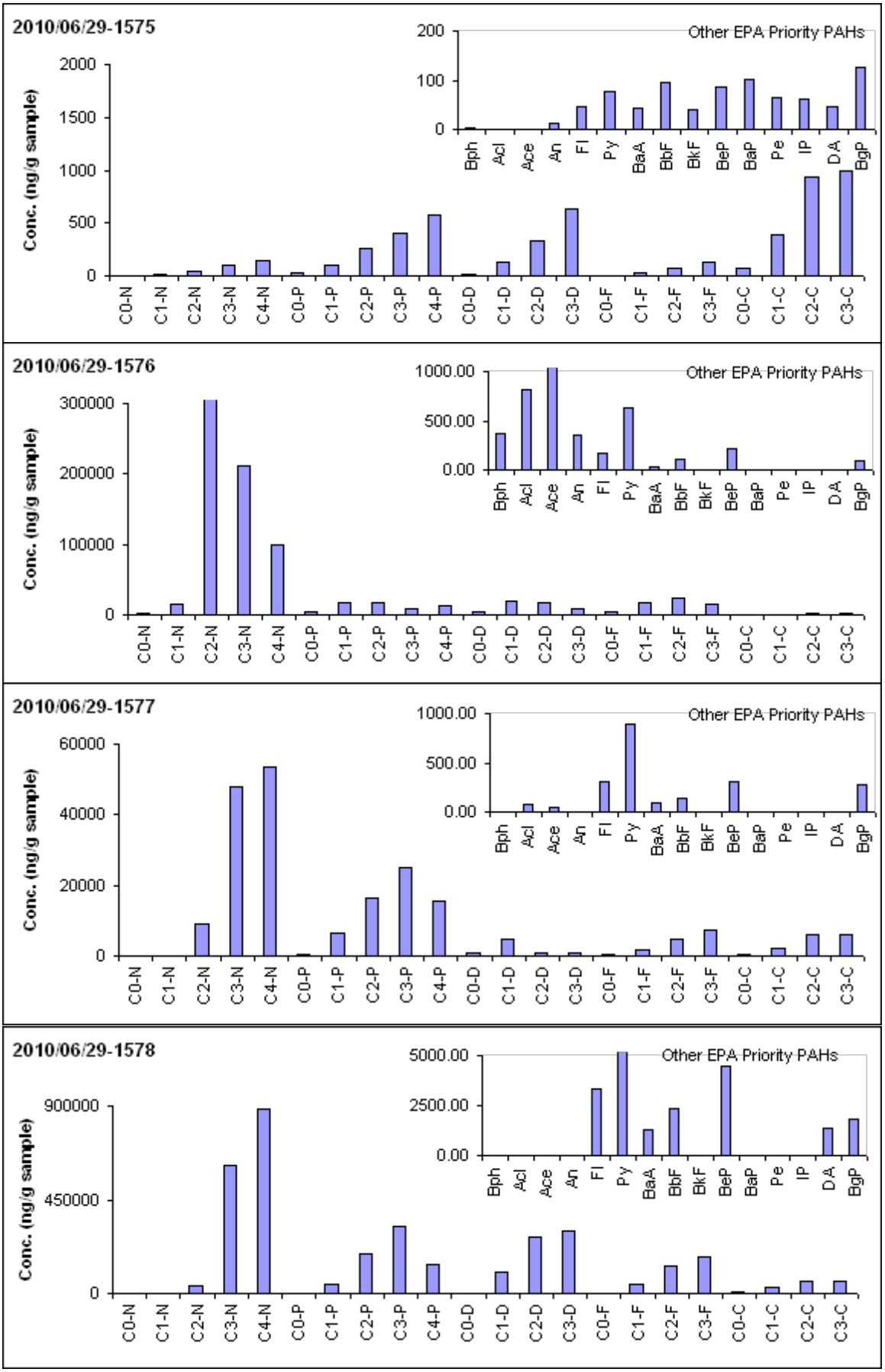


Figure S3 Distribution of target PAHs in samples

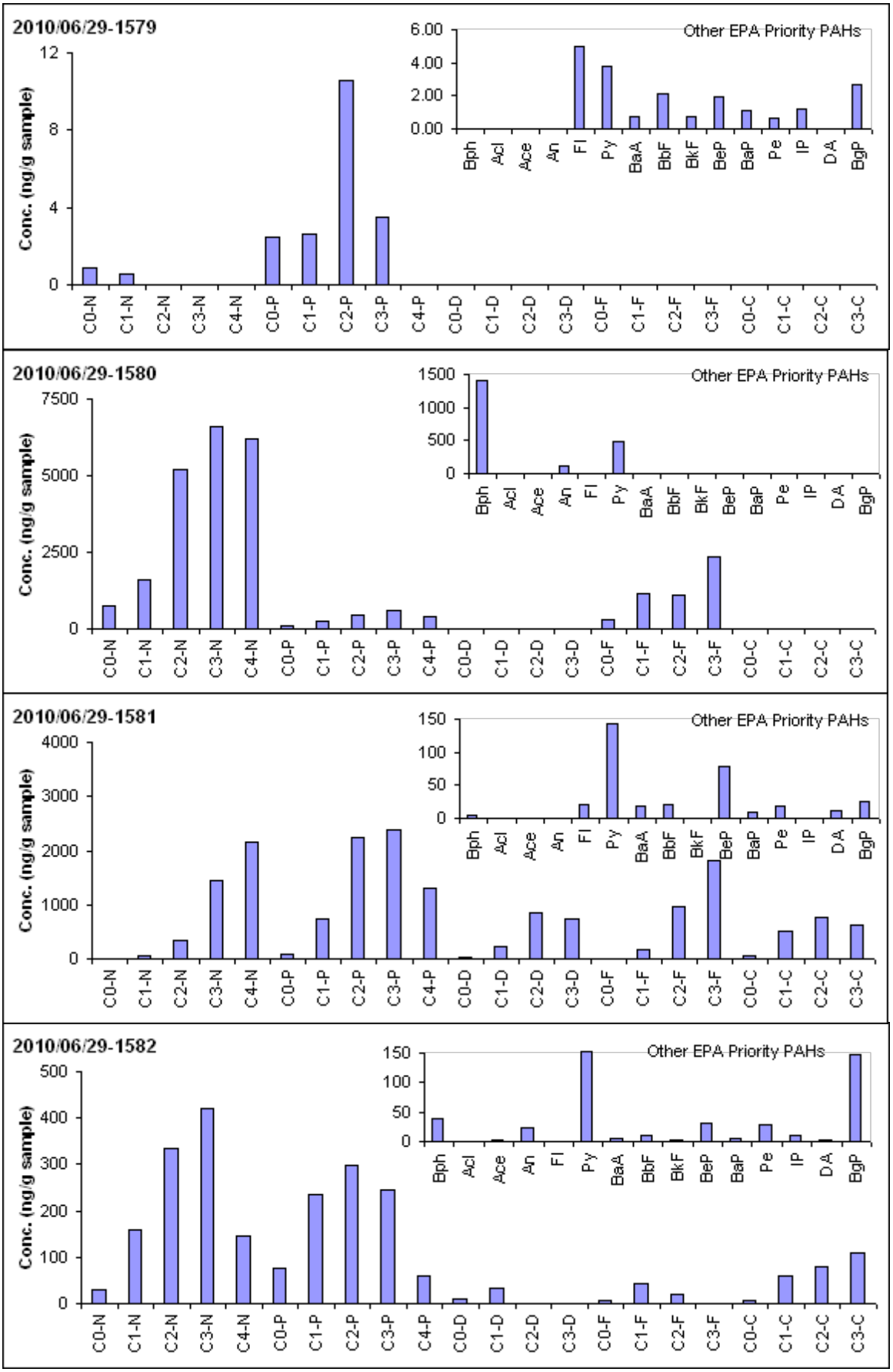


Figure S3 Distribution of target PAHs in samples

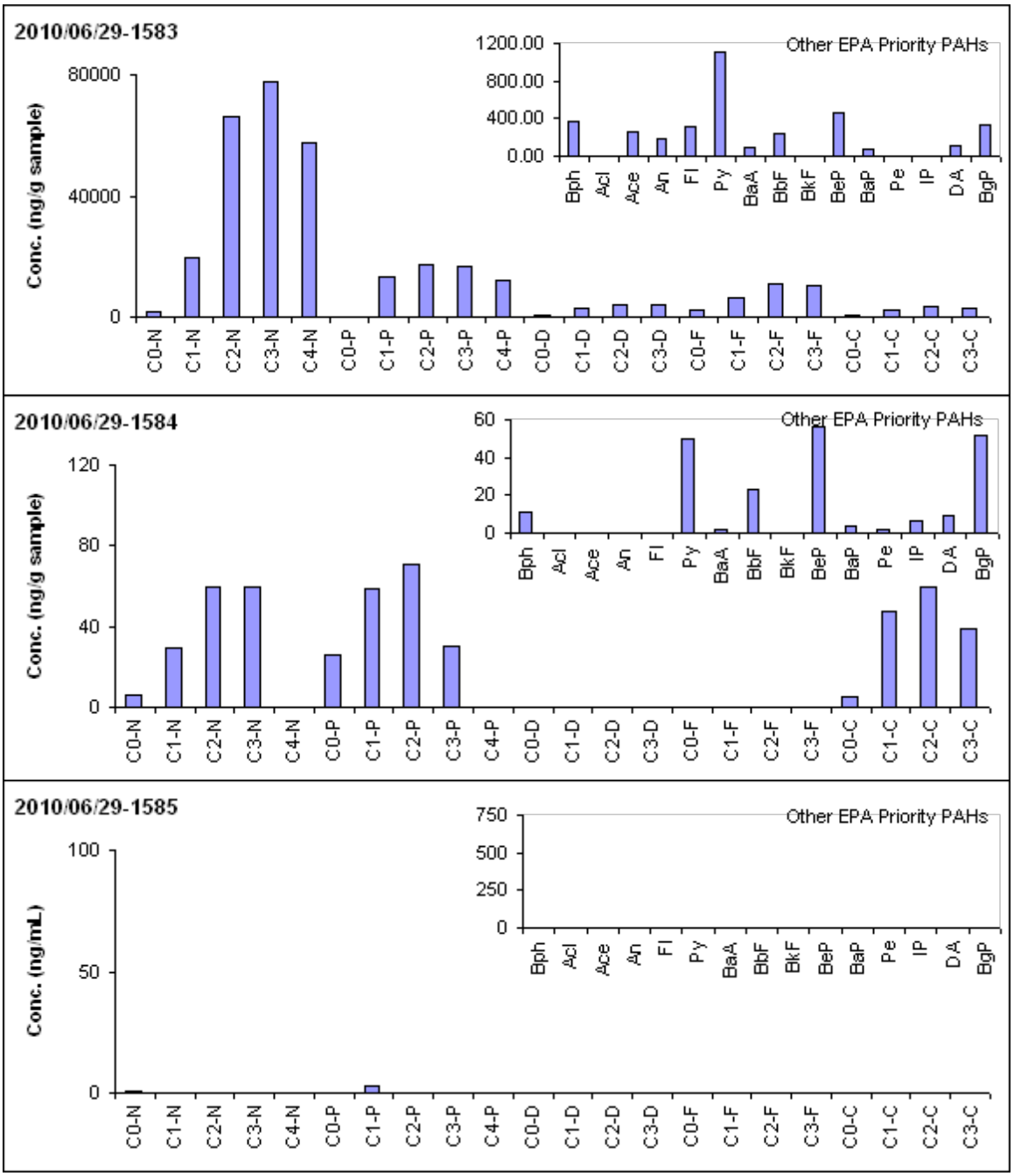


Figure S3 Distribution of target PAHs in samples

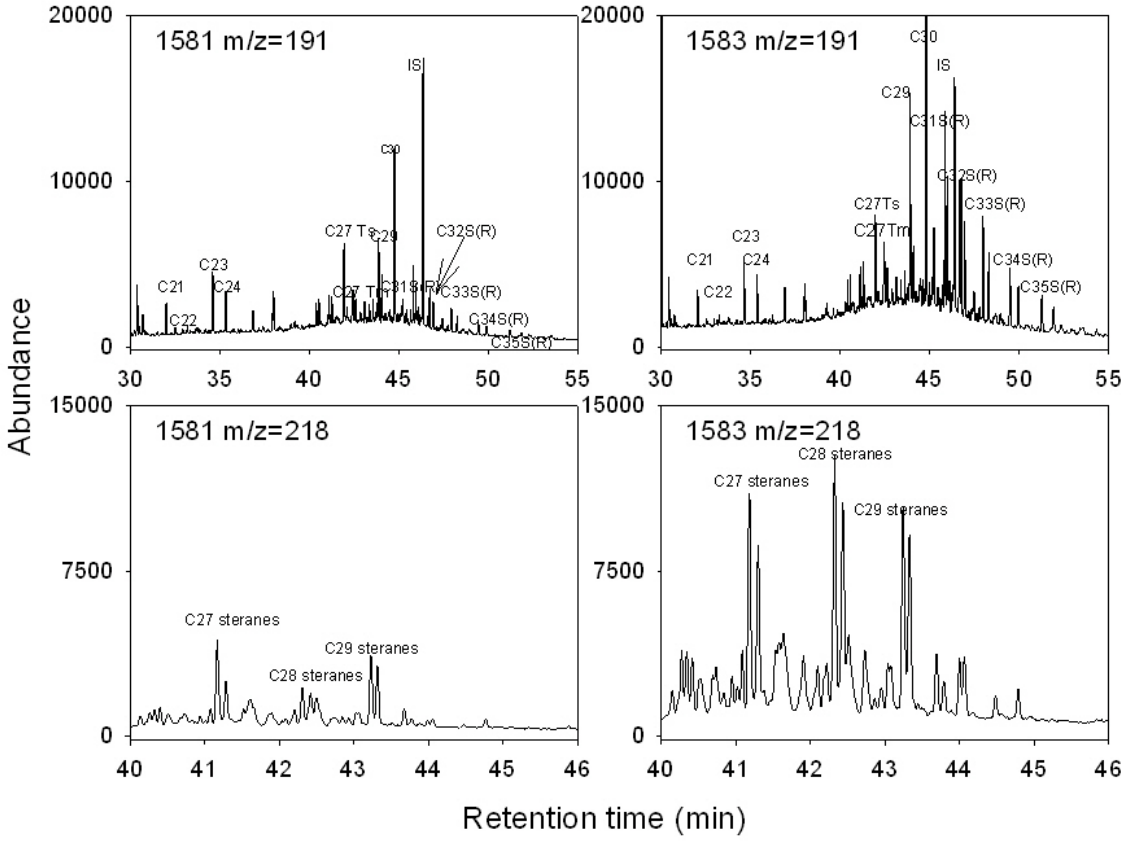


Figure S4 The GC-MS distribution of biomarker terpanes and steranes at m/z 191 and 218, respectively, in samples 1581 and 1583.

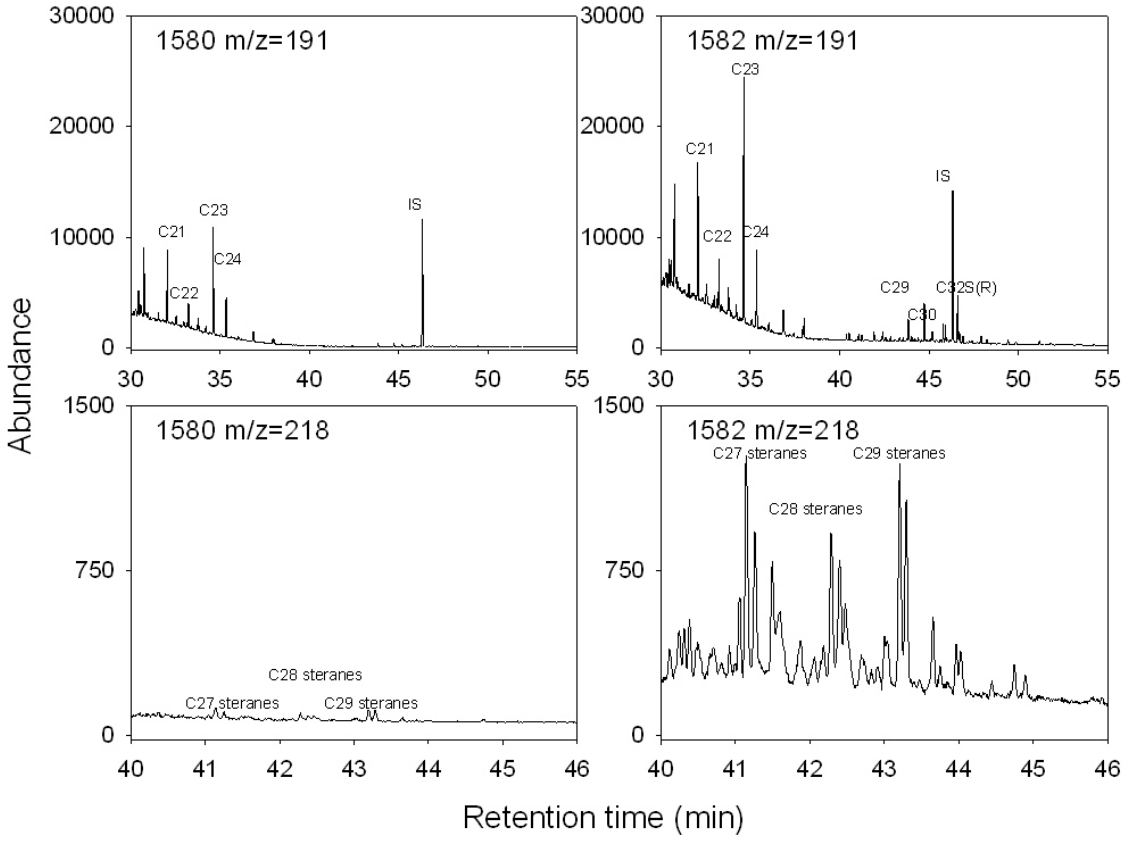


Figure S5 The GC-MS distribution of biomarker terpanes and steranes at m/z 191 and 218, respectively, in samples 1580 and 1582. The tricyclic C_{21} to C_{24} terpanes are significantly more abundant because these two sites were mainly contaminated by diesel fuels.

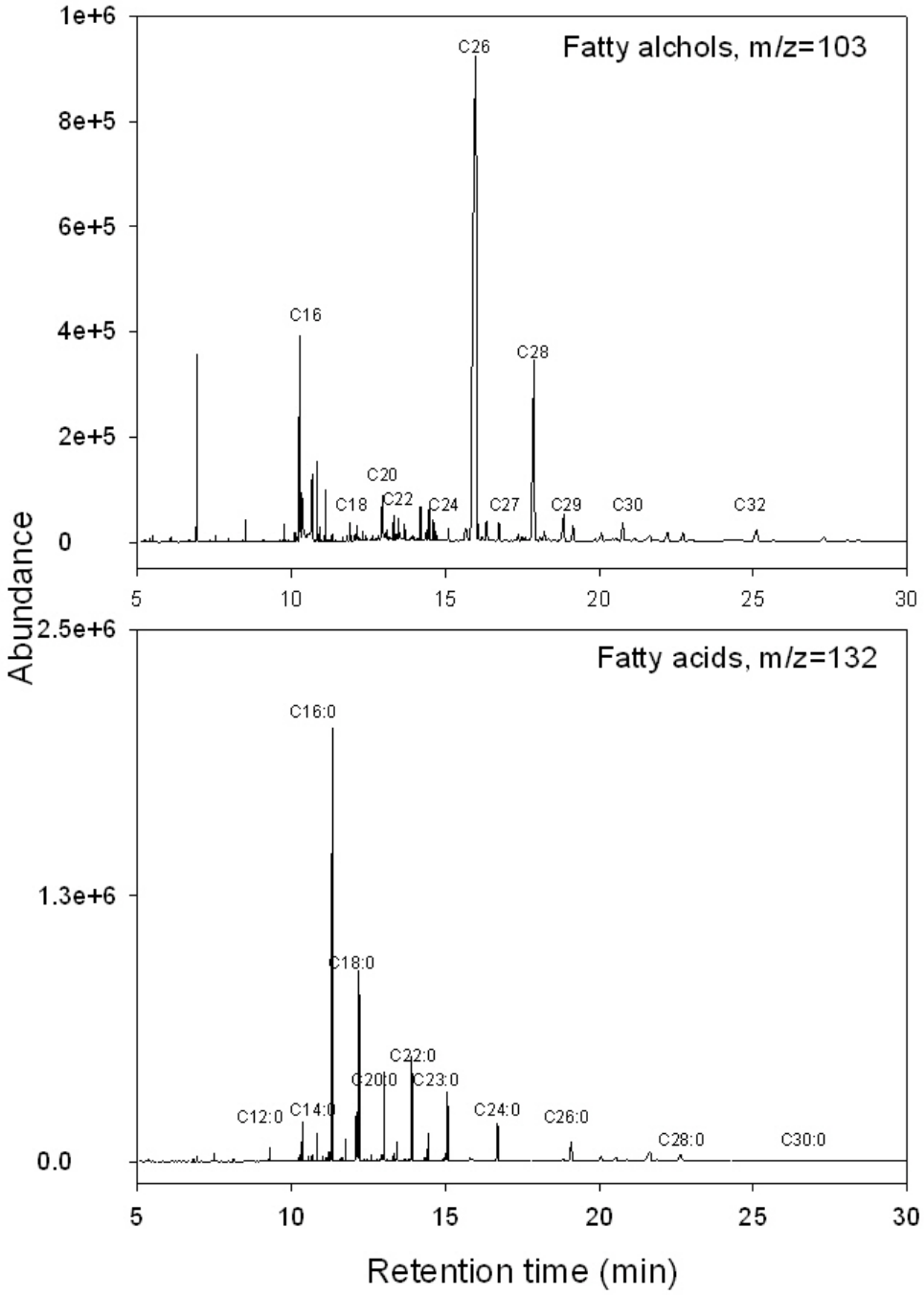


Figure S6 GC-MS-SIM chromatogram of the sample 1565 (top: at m/z 103, the key ion of silylated derivatives of fatty alcohols) for identification of fatty alcohols from C_{12} to C_{32} and for illustration of the distribution of these fatty alcohols in the sample; and the chromatogram of the same sample 1565 (bottom: at m/z 132, a key ion of silylated derivatives of fatty acids) for identification of fatty acids ranging from C_{10} to C_{30} and for illustration of the distribution of these fatty acids in the sample.

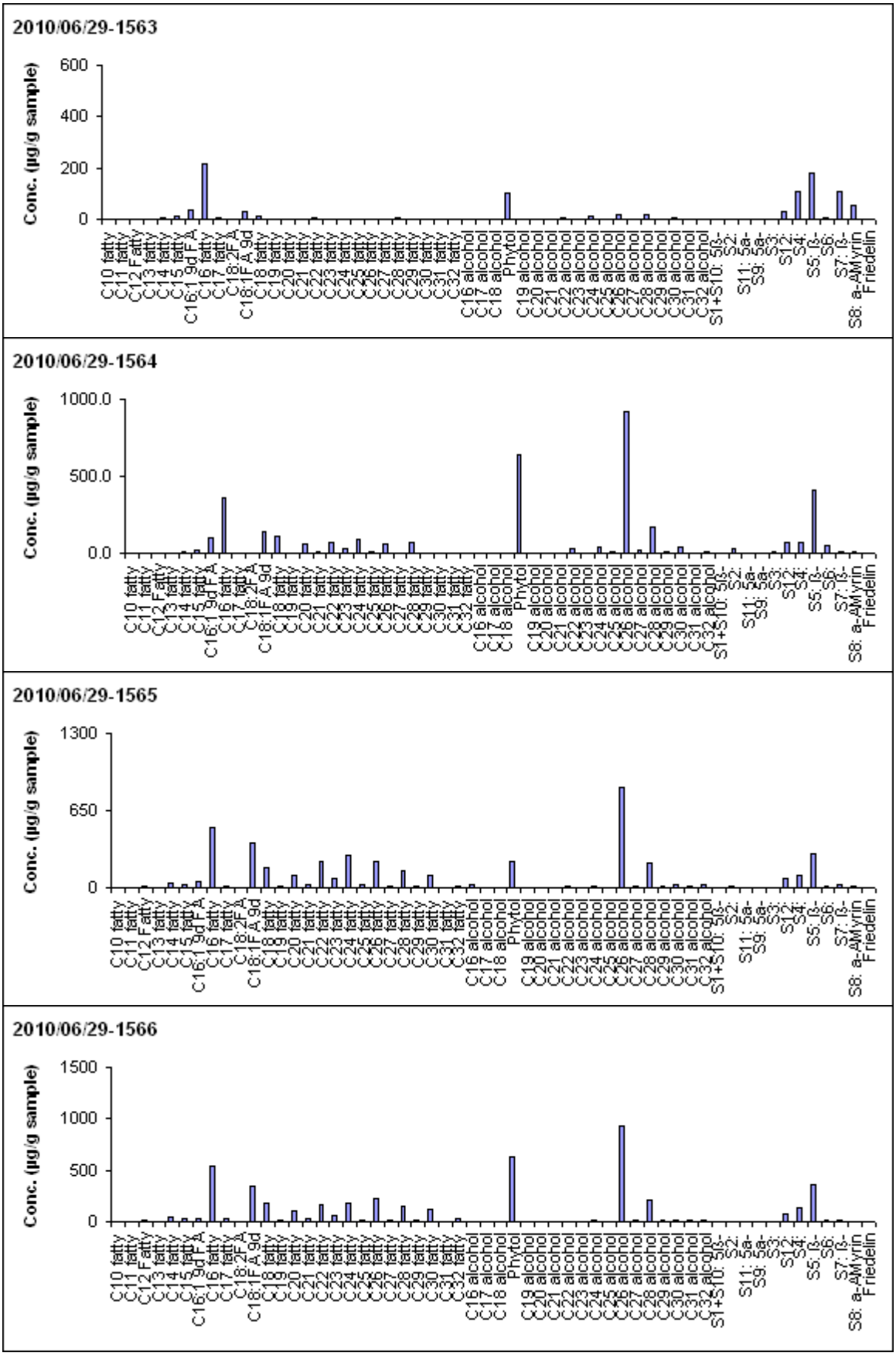


Figure S7 Distribution of target biogenic compounds in samples

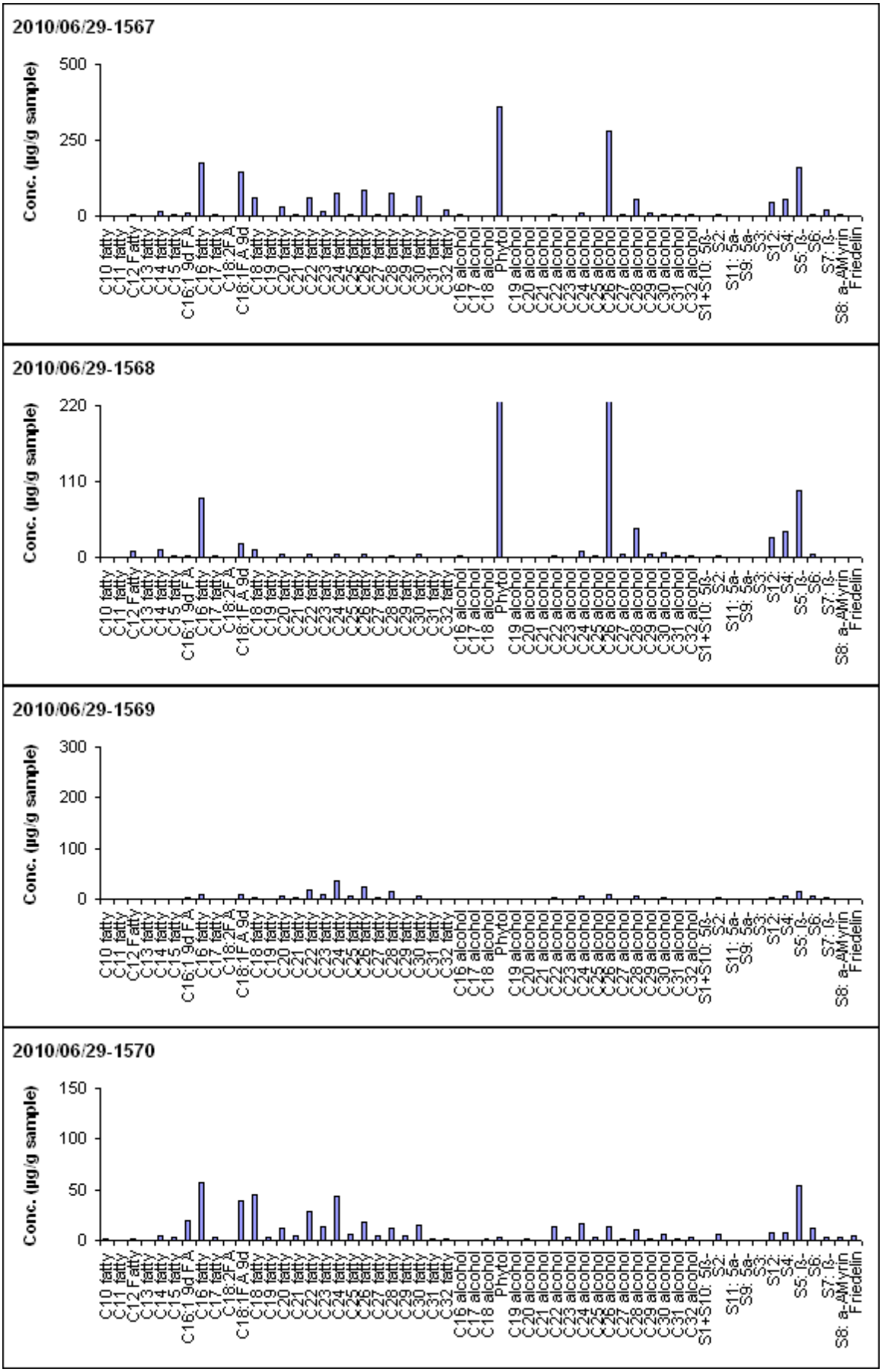


Figure S7 Distribution of target biogenic compounds in samples

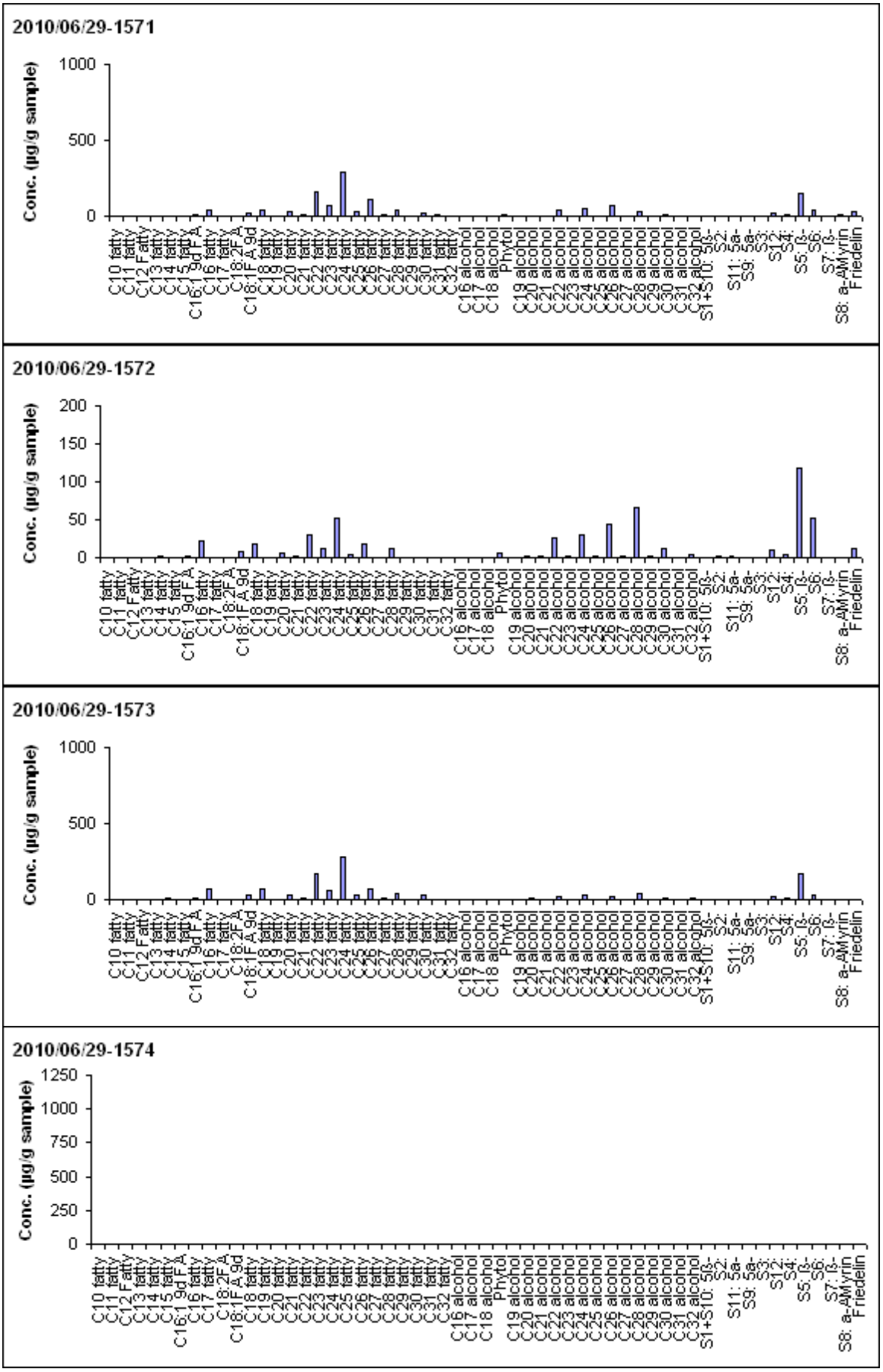


Figure S7 Distribution of target biogenic compounds in samples

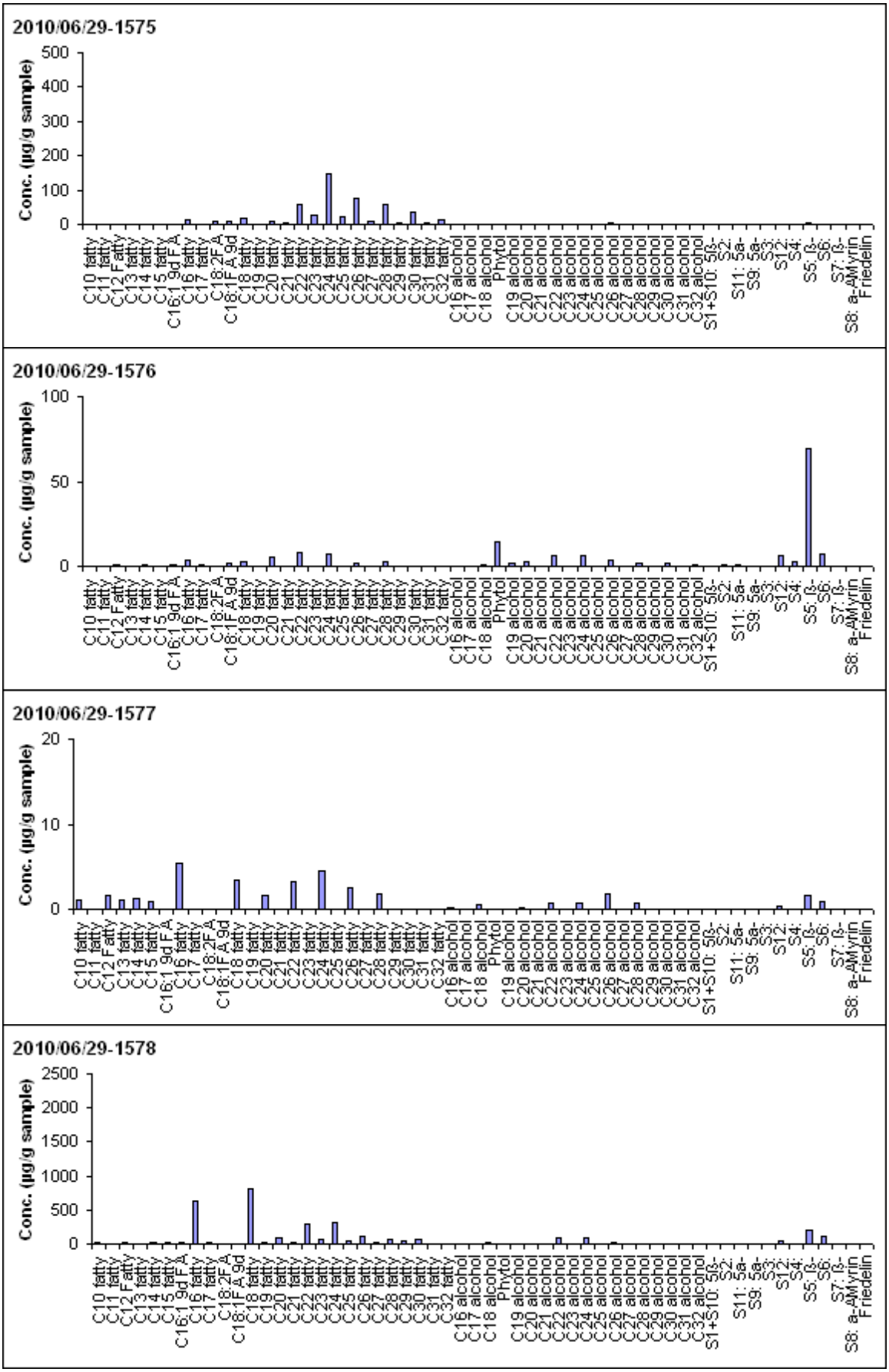


Figure S7 Distribution of target biogenic compounds in samples

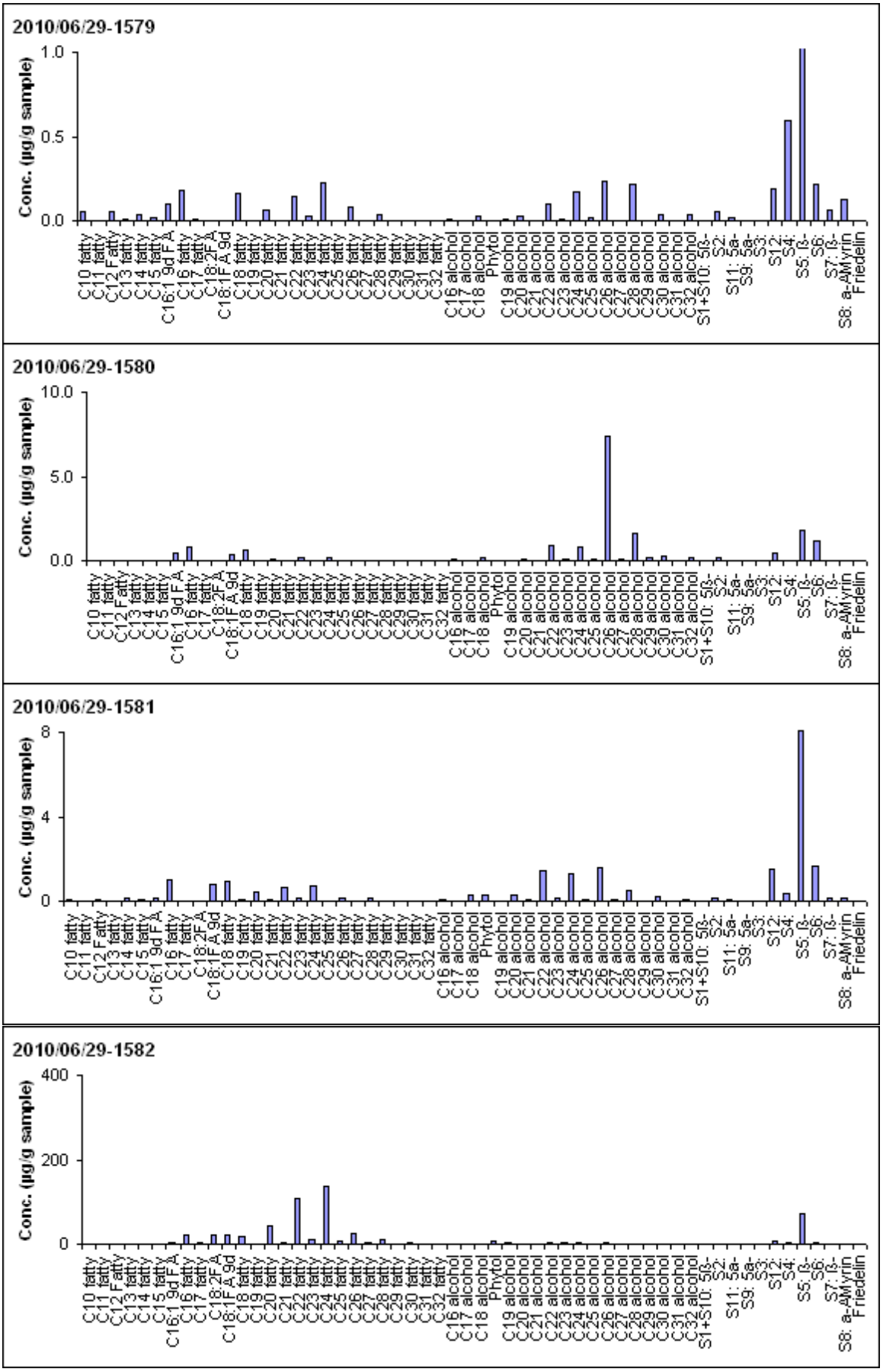


Figure S7 Distribution of target biogenic compounds in samples

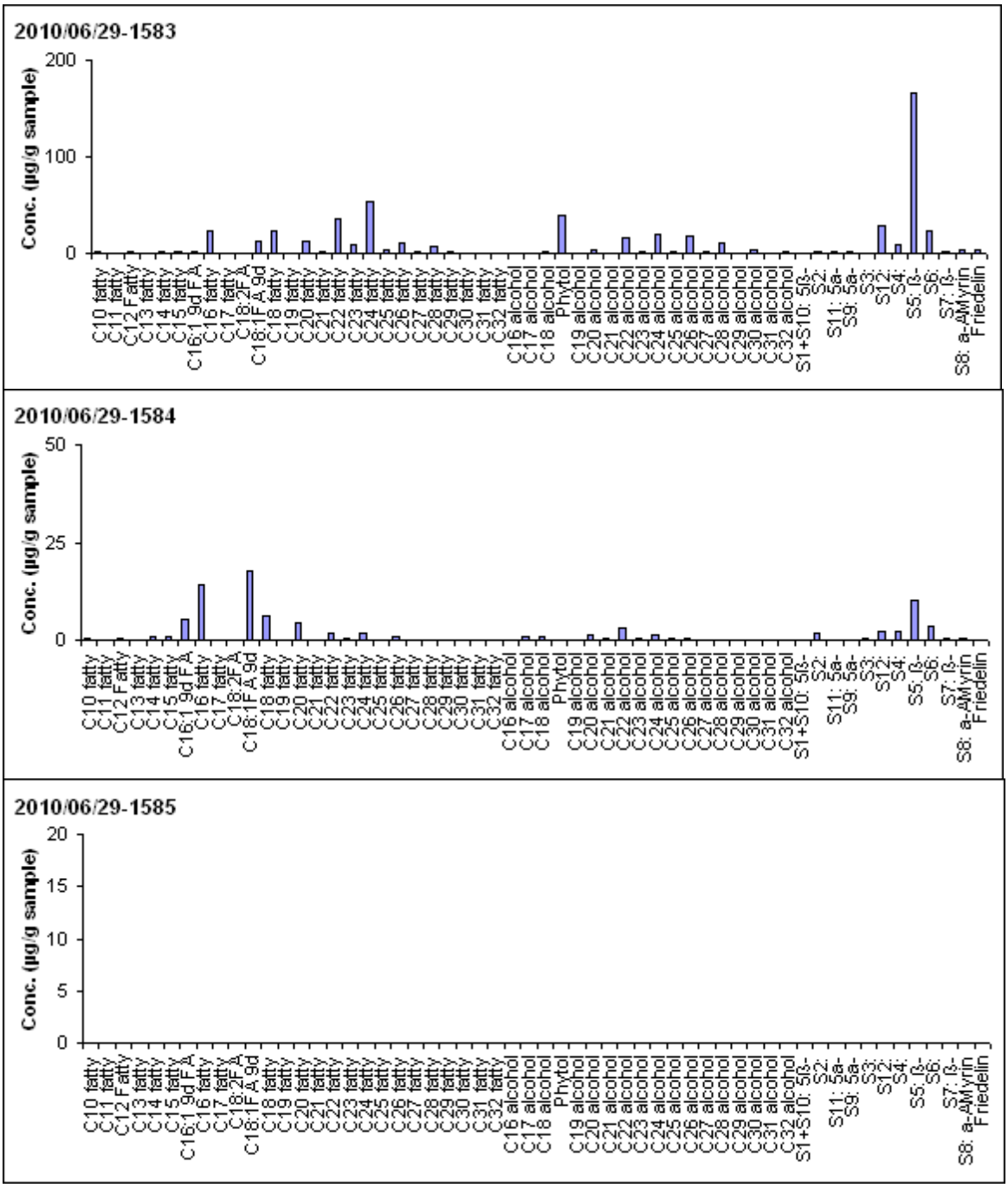


Figure S7 Distribution of target biogenic compounds in samples

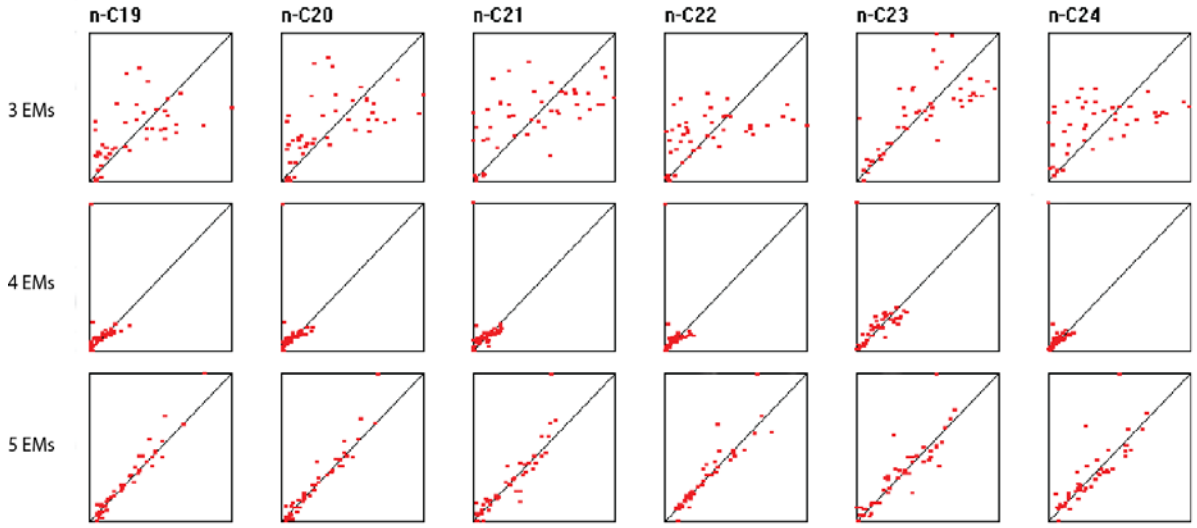


Figure S8 Improvement in the fit of the data to the PVA model as the number of end members increases for a selected group of mid-length alkanes. Going from three to four increases the predictability (y-axis) versus the actual data (x-axis) and going from four to five EMs makes a big difference. Increasing the number of EMs beyond five does not lead to any significant improvement in the fit.

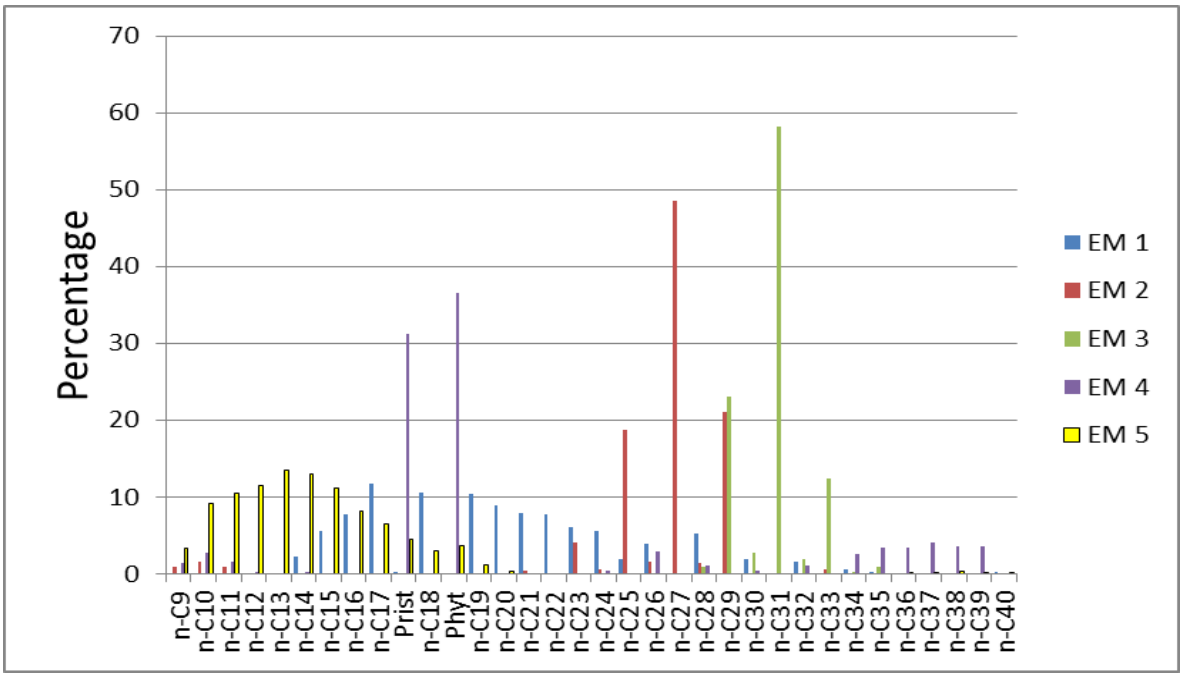


Figure S9 End member composition for the five EM solution. The EMs suggest the following source type: EM1, EM2, EM3, EM4, and EM5 (see Table 7).