Supporting online material

Reference isoprene emission model

Following Arneth et al. (2007) and Niinemets et al. (1999), the monthly isoprene emission rate I_{REF} in the reference model is modelled in LPJ as:

$$I_{REF} = \sum_{i=1}^{pfts} \varepsilon_i \frac{c_{i,370\,ppm}}{87 c_i} GPP_i \exp\left[0.1\left(T_{CAN-PFT} - 30^o C\right)\right]$$

where summation is over the plant functional types (PFTs), ε_i is the fraction of photosynthetic electrons used for isoprene synthesis under standard conditions (constant through the seasonal cycle), $c_{i,}$ is the CO_2 concentration inside the leaf, $c_{i,\ 370ppm}$ is the leaf-internal concentration corresponding to an ambient CO_2 concentration of 370 ppm and GPP_i is the gross primary production. Both c_i and $c_{i,\ 370ppm}$ are for non-water stressed conditions.

We used Guenther et al.'s (2006) values of the ε_i factors, with modifications from Millet et al., 2008, to derive values of the fraction of photosynthetic electrons diverted to isoprene production at standard temperature, ε, as follows: ε = 0.0085 for tropical broadleaved trees, 0.1070 for other broadleaved trees, 0.0303 for needle-leaved evergreen trees, 0.0106 for needleleaved summergreen trees, and 0.0076 for C_3 and C_4 grasses and other non-woody plants. Note that the effect of these different (and approximate) values on the modelled seasonal cycle of isoprene emission was slight, working only through differences in the phenology of different PFTs. This version of LPJ does not separate diffuse and direct radiation fractions, whose share in the total radiation received by the canopy may have an impact on photosynthesis rates in dense canopies (Alton, 2008).

Error analysis

For each region, for a given month, we calculate the standard deviation for all of the variables both as regional means (to yield σ_{IAV} - the inter-annual variability of the seasonal mean for a region) and as individual grid cells (to yield $\sigma_{SPATIAL}$ - the spatial variability of the seasonal mean within a region). The temperature σ values are shown in ^oC in Table S2. The other variables' σ values are normalized to the maximum seasonal average such that their values are on the same scale as in Figure 3. Even though the averaging period included two El Niño (1998, 2003) and two La Niña (1999, 2000) years, the inter-annual standard deviation of the signals, σ_{IAV}, is generally quite small with means less than 17% and 1.2^{o}C. The spatial variability within a given region is, on average, more than about twice the inter-annual variability, suggesting that the analysis might benefit from the use of smaller, more coherent regions. On the other hand, smaller regions are likely to suffer from greater contamination from neighbouring areas. The boreal regions show the largest $T_{CAN-LPJ}$ variance, whereas the temperate regions have the largest variance in the other variables. This is encouraging because the r and P values for these regions are very high and low, respectively, providing confidence in our conclusions about the correlations in these regions of relatively high variance.

In only three regions was $r(T_{AIR})$ greater than $r(T_{CAN\text{-}OBS})$ or $r(T_{CAN\text{-}LPJ})$: Aus2, San2 and Nord. Of those regions, only in the Nord region were the air-temperature correlations more than 0.07 greater than the canopy-temperature correlations. Nord is also one of the few sites in which $r(T_{CAN\text{-}OBS}) < 0.78$. In the other two sites with $r(T_{CAN\text{-}OBS}) <0.78$ (Con1 and Con2), both canopy temperature measures are better predictors of formaldehyde concentration than I_{REF}, *GPP* and T_{AIR}. These regions have the smallest seasonal temperature range (1.5°C, not shown) of all the regions examined and yet the P value for formaldehyde versus $T_{CAN\text{-}OBS}$ is 0.03, implying that the formaldehyde signal is influenced by $T_{CAN\text{-}OBS}$ with better than 95% confidence. It is not clear why the predictive capability of canopy temperatures fails in Nordeste. This region does not have unusually high standard deviations (Table S2), it has a fairly large temperature range (11°C in MODIS), the offset of the surface temperatures from the satellite data shows no bias towards large or small values, and similar biomes (Ivory Coast) are well modelled by the canopy temperatures. Unlike in other regions, where the 1-2 month shift in the maxima and minima of the canopy temperatures relative to air temperature improves the correlation with formaldehyde, in Nordeste, this timing shift degrades the correlation. Yet both the observed and the modelled surface temperatures show the same shift.

Supporting Material Figure Captions

Figure S1. The seasonal cycle of the net chemical modulation of formaldhyde, A, calculated from one year of full-chemistry simulation with the TM5 chemistry-transport model (Williams et al., 2012, Huijnen et al., 2010). Results for three latitude bands for the year 2006 are shown: the tropics (30°S-30°N, black solid line), the northern mid-latitudes (30°N-60°N, blue dashed line) and the southern mid-latitudes (30°S-60°S, red dotted line).

Figure S2. Differences between temperature fields (°C) for January and July 2002. Calculated canopy temperature *minus* air temperature (left, $T_{CAN\text{-}LPJ}$-T_{AIR}); MODIS 2m surface temperature *minus* air temperature (right, $T_{CAN\text{-}OBS}$-T_{AIR}).

Figure S3.1 to S3.4. A series of 4 panels showing scatter plots of the mean seasonal cycle of formaldehyde data, labeled HCHO here, versus the mean seasonal cycle of the candidate driving variables. These plots follow the same structure as Figure 3 in the main text, with four panels for the four ecobands – boreal, temperate forests, tropical forests and tropical savannas. The first column shows formaldehyde data versus the reference model, I_{REF}, with a linear least squares fit as the solid line and from which the correlation statistic r is derived and given in Table 2. Below the scatter plot is shown the difference between the two mean seasonal cycles, Δ, as a way of illustrating potential magnitude biases between the two variables. The second column shows the formaldehyde data versus *GPP*, while the last three columns plot the formaldehyde data versus T_{AIR}, $T_{CAN\text{-}LPJ}$ and $T_{CAN\text{-}OBS}$ again with the difference plots below.

Supporting Material Tables

Table S1 Correlations and P-values of formaldehyde concentration with environmental predictor variables. The regions are colour-coded in the same way as Figure 4. Averages for each ecological zone are labelled BOREAL (boreal regions), TEMP (temperate forests), TROP-F (tropical forests) and TROP-G (tropical savannas).

	Correlation Coefficient, r							**p value**						
						Canopy temp							Canopy temp	
	I_{REF}	*GPP*	*fPAR*	*Prec*	T_{AIR}	T_{OBS}	T_{LPJ}	I_{REF}	*GPP*	*fPAR*	*Prec*	T_{AIR}	T_{OBS}	T_{LPJ}
NSib	0.93	0.95	0.88	0.53	0.68	0.93	0.78	0.00	0.00	0.00	1.00	0.00	0.00	0.00
SSib	0.95	0.93	0.95	0.59	0.76	0.92	0.85	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Can1	0.94	0.92	0.93	0.54	0.74	0.96	0.90	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Eur	0.83	0.89	0.91	0.26	0.72	0.98	0.97	0.00	0.00	0.00	0.98	0.00	0.00	0.00
BOREAL	0.91	0.92	0.92	0.48	0.73	0.95	0.87	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Can2	0.89	0.95	0.89	0.71	0.76	0.95	0.97	0.00	0.00	0.00	1.00	0.00	0.00	0.00
NAm	0.87	0.81	0.82	0.85	0.73	0.94	0.93	0.00	0.00	0.00	1.00	0.00	0.00	0.00
USA1	0.96	0.93	0.95	0.89	0.74	0.94	0.97	0.00	0.00	0.00	1.00	0.00	0.00	0.00
USA2	0.98	0.70	0.97	0.64	0.80	0.89	0.96	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Aus1	-0.10	-0.84	-0.70	-0.03	0.85	0.87	0.88	0.62	1.00	1.00	0.40	0.00	0.00	0.00
Aus2	0.76	-0.89	-0.87	0.50	0.93	0.86	0.88	0.00	1.00	1.00	1.00	0.00	0.00	0.00
Chi	0.84	0.92	0.84	0.80	0.76	0.91	0.90	0.00	0.00	0.00	1.00	0.00	0.00	0.00
TEMP	0.74	0.37	0.41	0.62	0.80	0.91	0.93	0.09	0.29	0.29	0.91	0.00	0.00	0.00
Amz1	0.77	0.26	0.15	-0.87	0.28	0.93	0.75	0.00	0.20	0.31	0.00	0.18	0.00	0.00
Amz2	0.70	0.51	-0.21	-0.91	0.07	0.91	0.55	0.00	0.03	0.75	0.00	0.41	0.00	0.02
Amz3	0.69	0.21	0.25	-0.76	-0.11	0.95	0.68	0.00	0.25	0.20	0.00	0.63	0.00	0.00
Amz4	0.67	-0.01	0.74	-0.90	0.07	0.91	0.85	0.00	0.51	0.00	0.00	0.41	0.00	0.00
San1	-0.40	-0.91	0.48	-0.89	0.85	0.90	0.88	0.92	1.00	0.04	0.00	0.00	0.00	0.00
San2	0.39	-0.63	0.61	-0.85	0.82	0.79	0.83	0.09	0.99	0.01	0.00	0.00	0.00	0.00
Con1	0.03	-0.62	-0.65	-0.65	0.11	0.52	0.44	0.46	0.99	1.00	0.00	0.36	0.03	0.06
Con2	-0.34	-0.79	-0.77	-0.68	0.04	0.50	0.51	0.87	1.00	1.00	0.00	0.45	0.03	0.03
TROP-F	0.31	-0.25	0.08	-0.81	0.27	0.80	0.68	0.29	0.62	0.41	0.00	0.31	0.01	0.01
Nord	0.50	0.32	0.06	0.25	0.88	0.31	0.30	0.03	0.14	0.43	1.00	0.00	0.15	0.16
IC1	-0.59	-0.84	-0.88	-0.86	0.40	0.97	0.97	0.99	1.00	1.00	0.00	0.08	0.00	0.00
IC2	-0.41	-0.77	-0.86	-0.87	0.45	0.95	0.95	0.92	1.00	1.00	0.00	0.06	0.00	0.00
TROP-G	-0.16	-0.43	-0.56	-0.50	0.58	0.75	0.74	0.65	0.71	0.81	0.33	0.05	0.05	0.05
ALL	0.49	0.14	0.25	-0.08	0.56	0.85	0.80	0.22	0.41	0.35	0.52	0.12	0.01	0.01

Table S2 Standard deviation of seasonal cycle of formaldehyde concentration (HCHO), and modelled variables (I_{REF}, *GPP and* $T_{CAN\text{-}LPJ}$). We compute the standard deviation, $\sigma_{SPATIAL}$, for each variable in a given month, across a given region and across the 10 years in the study. In this table we present the monthly value of sigma that is *largest* over the course of the year. We also give the inter-annual variation, σ_{IAV}, as the standard deviation of the monthly regional averages. The regions are colour-coded as in Figure 3 and averages for each grouping are given as the last line in each colour, labelled BOREAL, TEMPER (temperate forests), TROPFOR (tropical forests) and TROPGRA (tropical savannas). The values of σ for HCHO, I_{REF} and GPP are computed as fractions of the maximum mean seasonal cycle such that they are on the same scale as in Figure 3, whereas the σ values for $T_{CAN\text{-}LPJ}$ are in ˚C.

	σ_{IAV}				$\sigma_{SPATIAL}$			
	HCHO	I_{REF}	GPP	$T_{CAN\text{-}LPJ}$	HCHO	I_{REF}	GPP	$T_{CAN\text{-}LPJ}$
NSib	0.10	0.03	0.02	2.46	0.22	0.13	0.08	2.97
SSib	0.12	0.04	0.03	2.68	0.32	0.12	0.07	4.11
Can1	0.11	0.03	0.02	1.88	0.31	0.38	0.20	3.82
Eur	0.14	0.03	0.03	1.99	0.39	0.13	0.09	3.43
BOREAL	0.12	0.03	0.02	2.25	0.31	0.19	0.11	3.58
Can2	0.30	0.04	0.05	1.81	0.52	0.10	0.11	2.83
NAm	0.26	0.07	0.11	1.83	0.48	0.35	0.26	4.56
USA1	0.23	0.04	0.05	1.70	0.40	0.14	0.13	2.91
USA2	0.09	0.07	0.08	1.35	0.22	0.26	0.27	3.74
Aus1	0.37	0.28	0.21	1.34	0.67	0.93	0.89	3.19
Aus2	0.42	0.39	0.38	1.34	0.72	1.17	1.53	2.51
Chi	0.12	0.04	0.06	0.93	0.32	0.22	0.18	3.64
TEMPER	0.26	0.13	0.13	1.47	0.47	0.45	0.48	3.34
Amz1	0.18	0.05	0.04	0.64	0.34	0.11	0.11	1.44
Amz2	0.16	0.04	0.03	0.40	0.34	0.12	0.06	1.61
Amz3	0.13	0.04	0.04	0.54	0.34	0.16	0.11	1.70
Amz4	0.16	0.05	0.05	0.63	0.33	0.11	0.11	1.03
San1	0.14	0.07	0.07	0.72	0.36	0.14	0.12	1.24
San2	0.15	0.05	0.04	0.67	0.32	0.13	0.11	1.20
Con1	0.13	0.06	0.04	0.67	0.27	0.14	0.15	1.37
Con2	0.14	0.07	0.03	0.78	0.27	0.14	0.12	1.36
TROPFOR	0.15	0.05	0.04	0.63	0.32	0.13	0.11	1.37
Nord	0.08	0.12	0.15	0.78	0.37	0.43	0.43	2.58
IC1	0.10	0.06	0.04	0.53	0.25	0.25	0.19	1.67
IC2	0.11	0.06	0.05	0.56	0.28	0.22	0.19	1.73
TROGRA	0.10	0.08	0.08	0.62	0.30	0.30	0.27	1.99
ALL	0.17	0.08	0.07	1.19	0.37	0.27	0.25	2.48

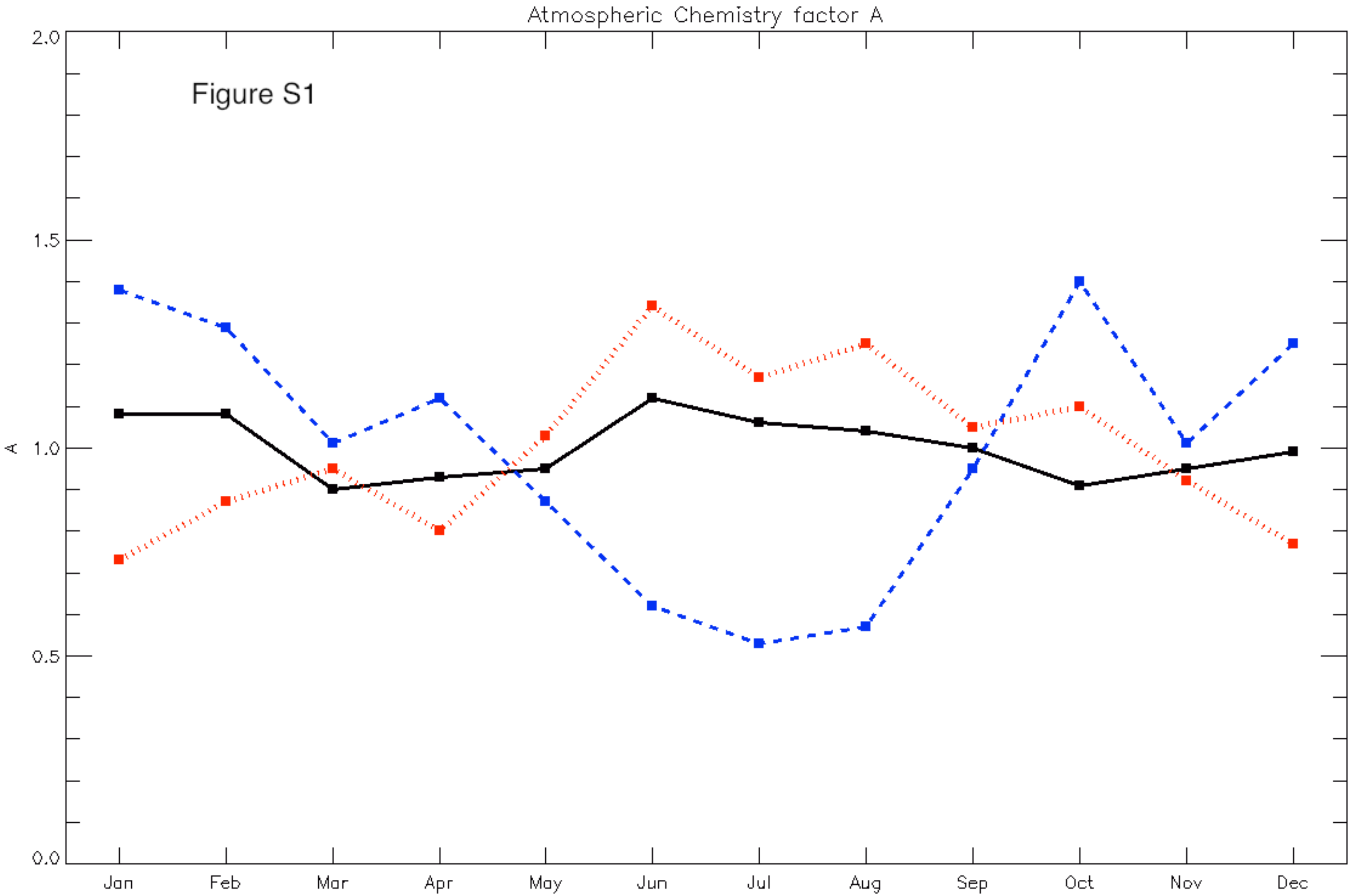
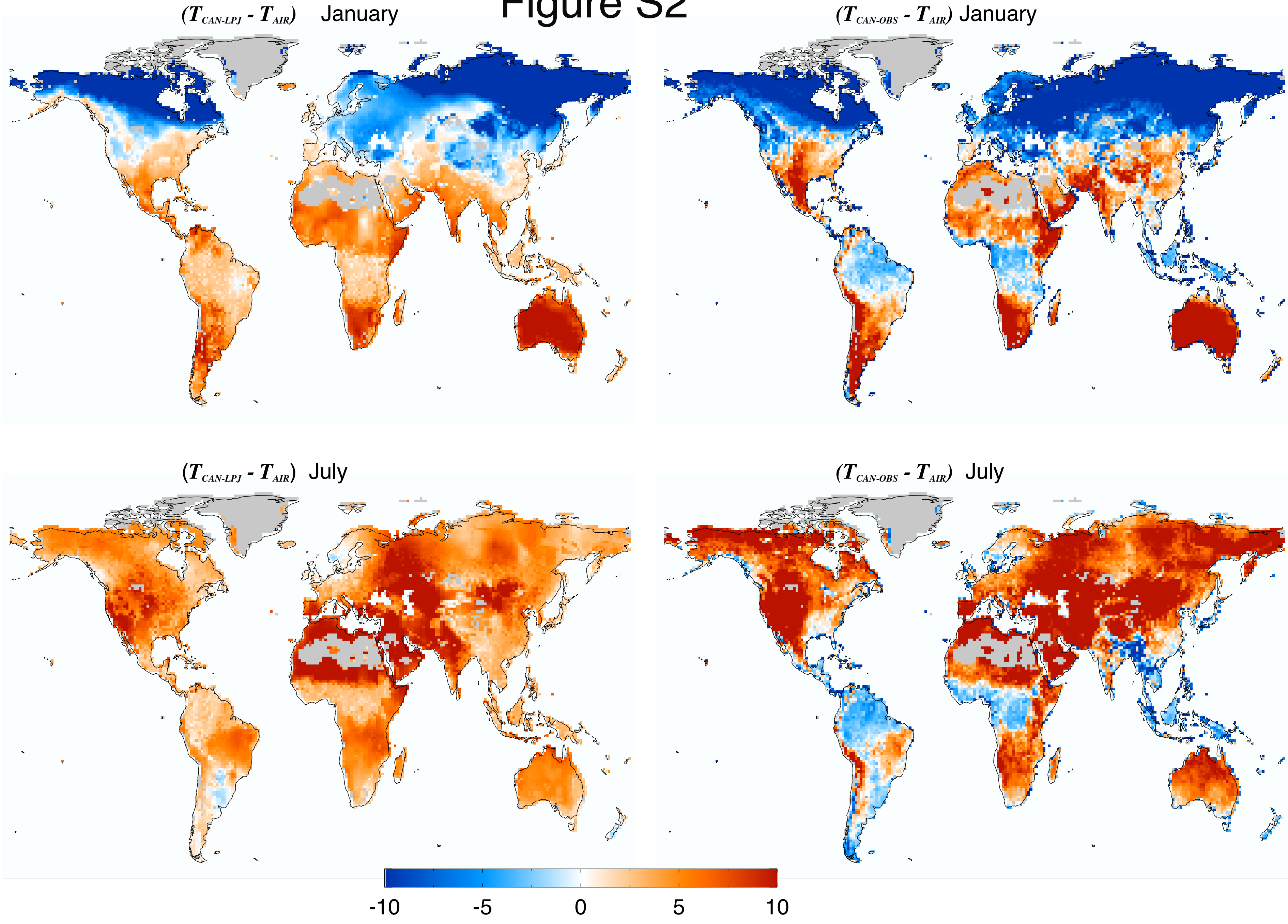
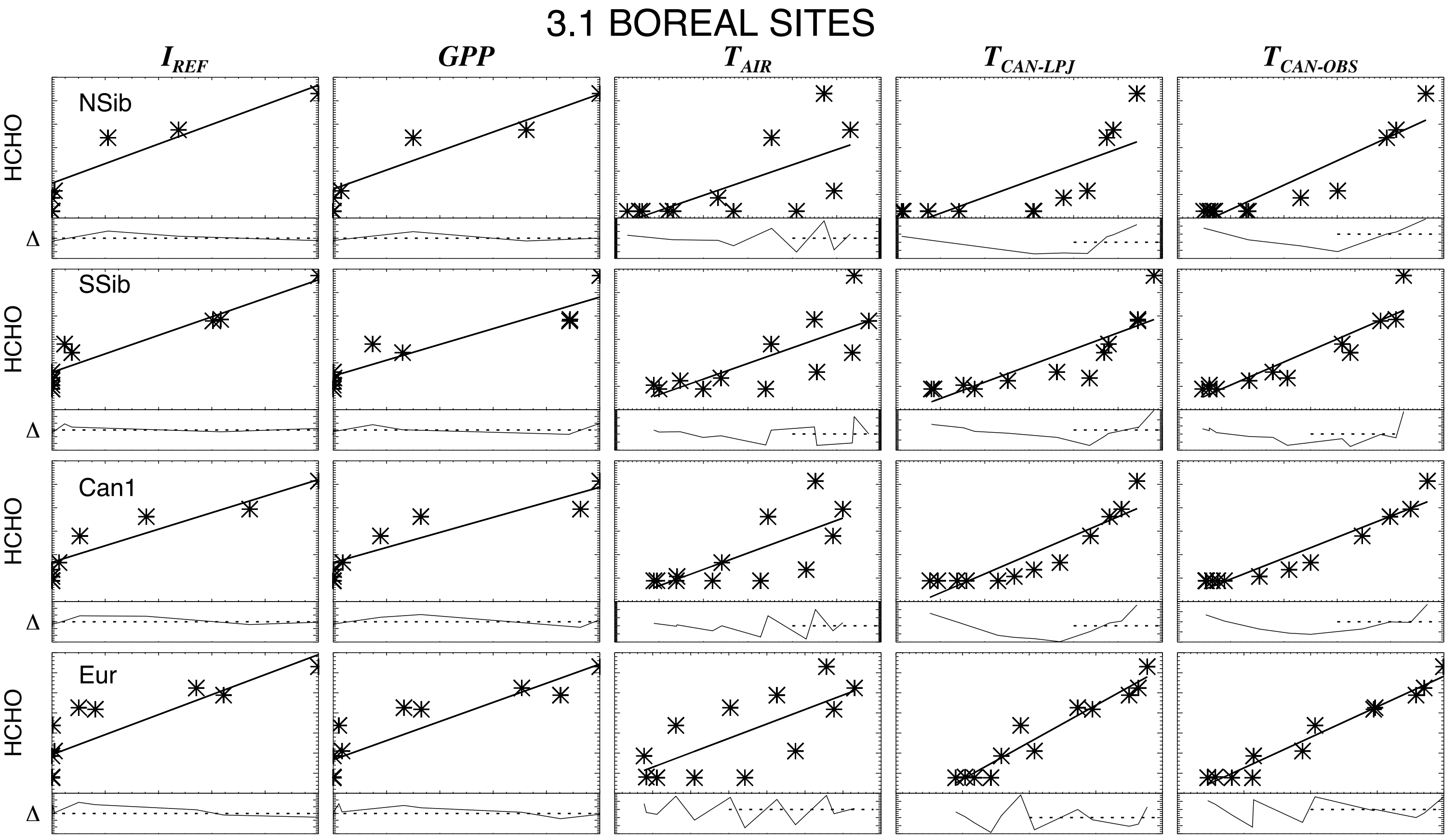

Atmospheric Chemistry factor A
Figure S1
2.0
1.5
1.0
0.5
0.0
A
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Figure S2

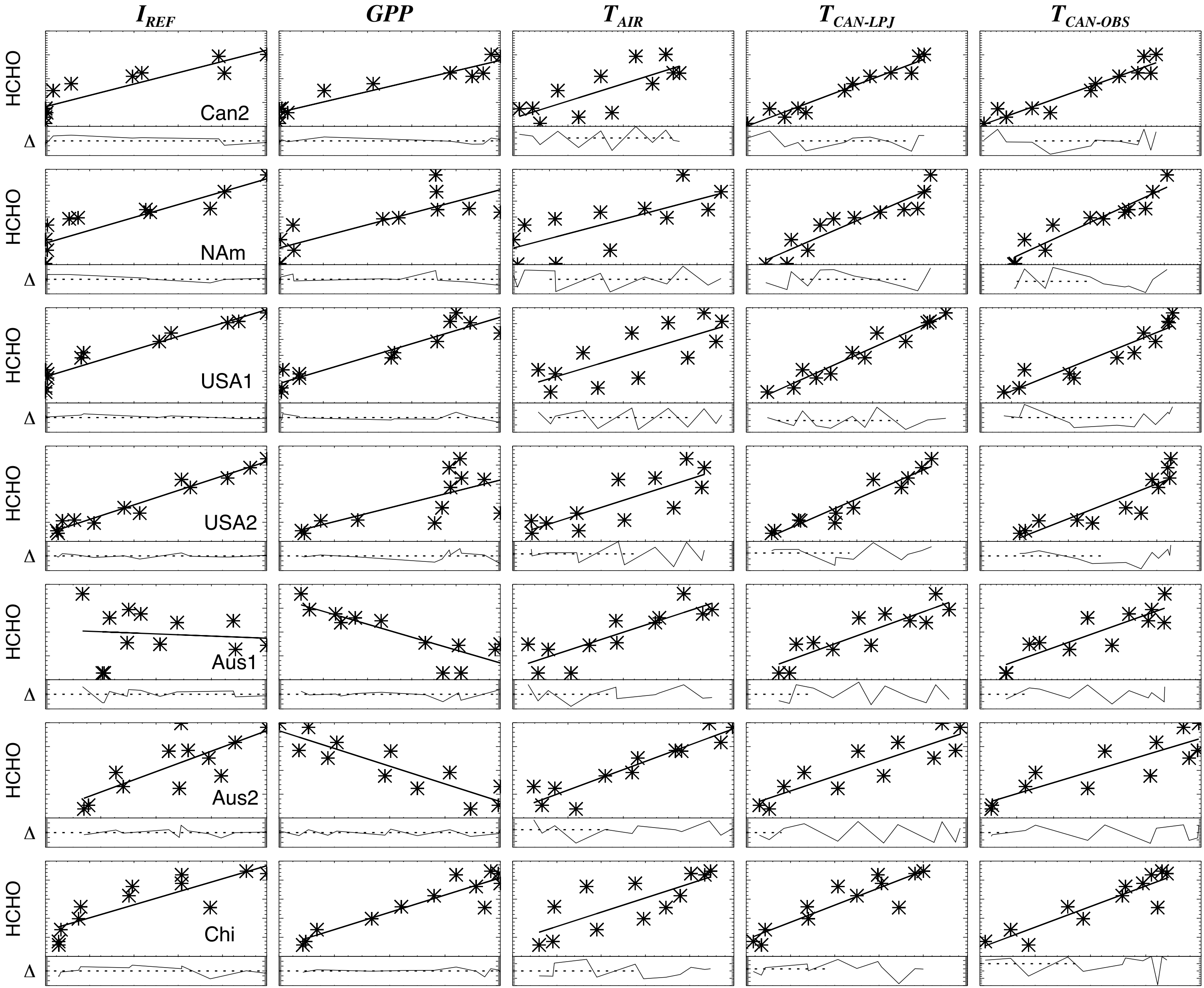


($T_{CAN-LPJ}$ - T_{AIR}) January
($T_{CAN-OBS}$ - T_{AIR}) January
($T_{CAN-LPJ}$ - T_{AIR}) July
($T_{CAN-OBS}$ - T_{AIR}) July
-10
-5
0
5
10

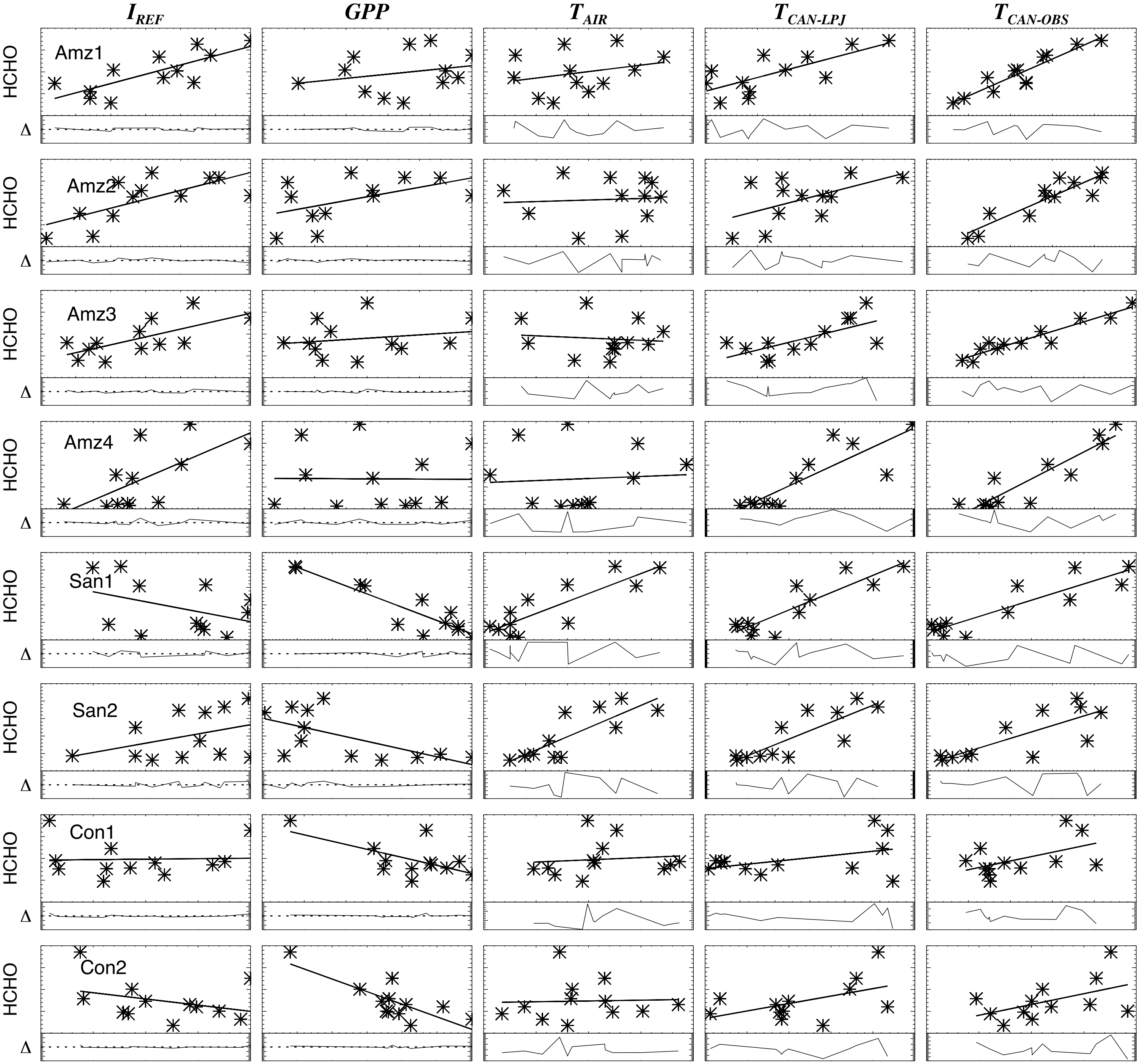


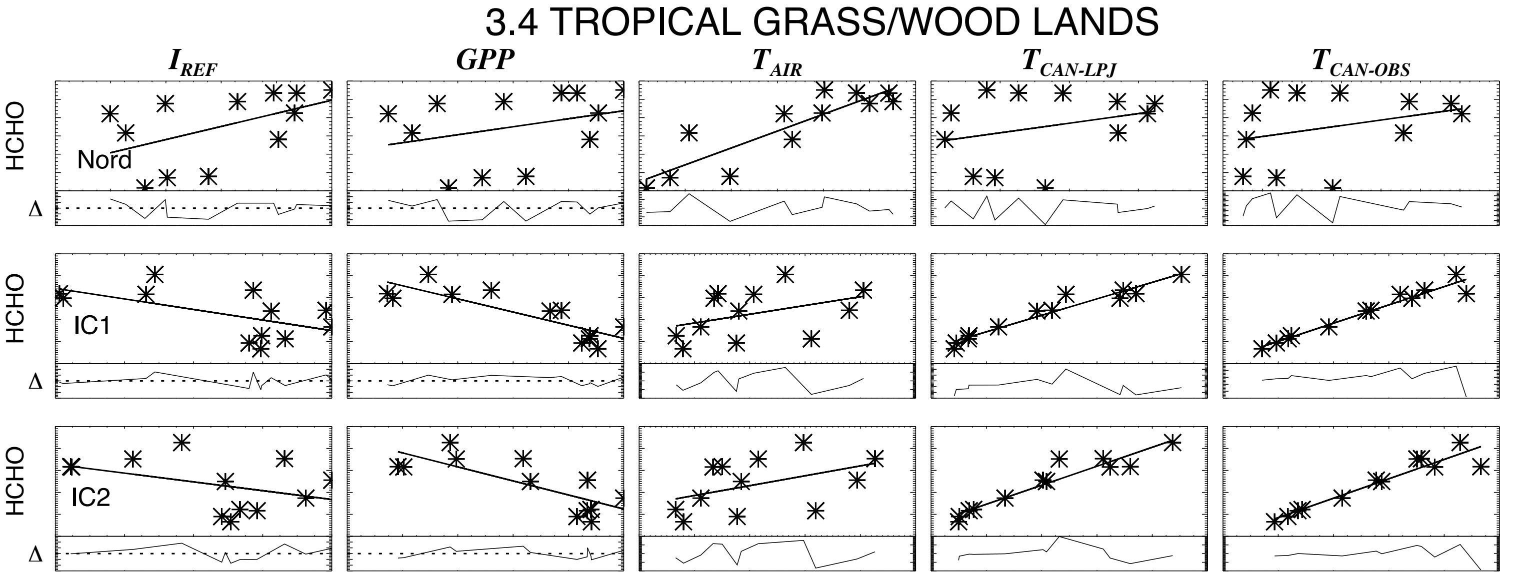
3.1 BOREAL SITES
I_REF
GPP
T_AIR
T_CAN-LPJ
T_CAN-OBS
NSib
SSib
Can1
Eur
HCHO
Δ

3.2 TEMPERATE SITES



3.3 TROPICAL FORESTS




3.4 TROPICAL GRASS/WOOD LANDS
I_{REF}
GPP
T_{AIR}
$T_{CAN-LPJ}$
$T_{CAN-OBS}$
HCHO
Nord
Δ
HCHO
IC1
Δ
HCHO
IC2
Δ