Supplement of Geosci. Model Dev., 14, 3633–3661, 2021
https://doi.org/10.5194/gmd-14-3633-2021-supplement

Geoscientific Model Development
Open Access

Supplement of

A model for urban biogenic CO_2 fluxes: Solar-Induced Fluorescence for Modeling Urban biogenic Fluxes (SMUrF v1)

Dien Wu et al.

Correspondence to: Dien Wu (dienwu@caltech.edu)

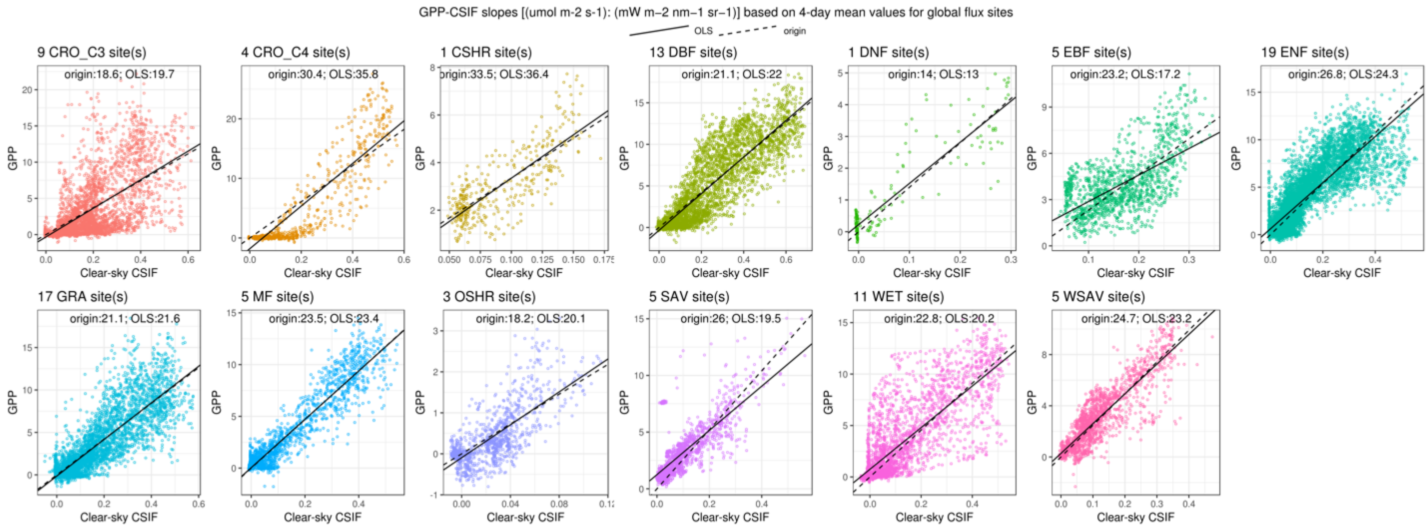

Figure S1. GPP-CSIF slopes [(μmol m^{-2} s^{-1}): (mW m^{-2} nm^{-1} sr^{-1})] based on 4-day mean CSIF and observed GPP from 98 global flux sites across 12 different natural biomes. Both linear fits across the origin (dashed lines) and fits using the Ordinary least squares (OLS) regressions have been plotted (solid lines) with slopes printed on the top of each panel.

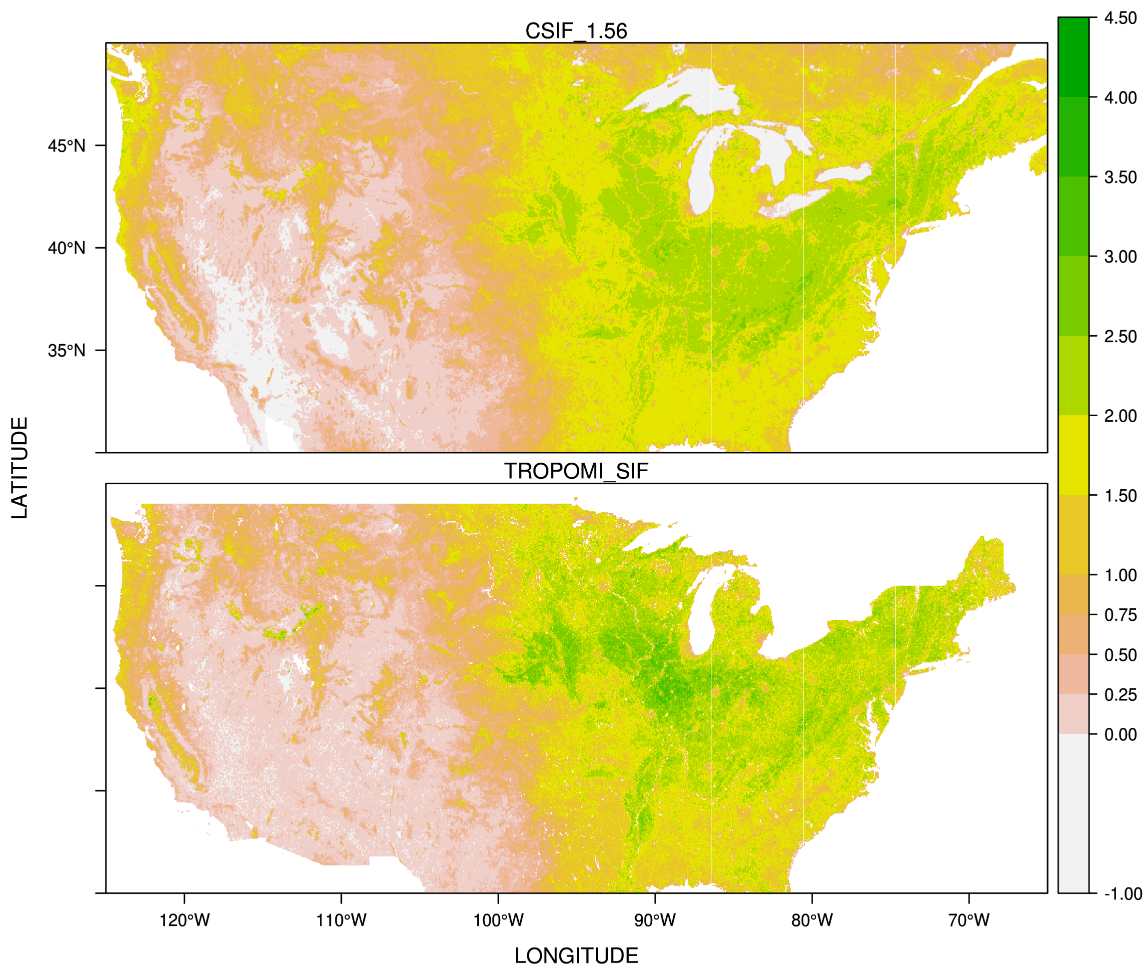

Figure S2. Comparison between instantaneous CSIF and TROPOMI SIF (mW m^{-2} nm^{-1} sr^{-1}, at the overpass time of ~1 pm vs. 1:30 pm), averaged over all days during JJA 2018 at a grid spacing of 0.05º. Downscaled TROPOMI SIF is initially available at 500 m horizontal grid spacing and then re-projected to CSIF grid and CSIF has been scaled up.

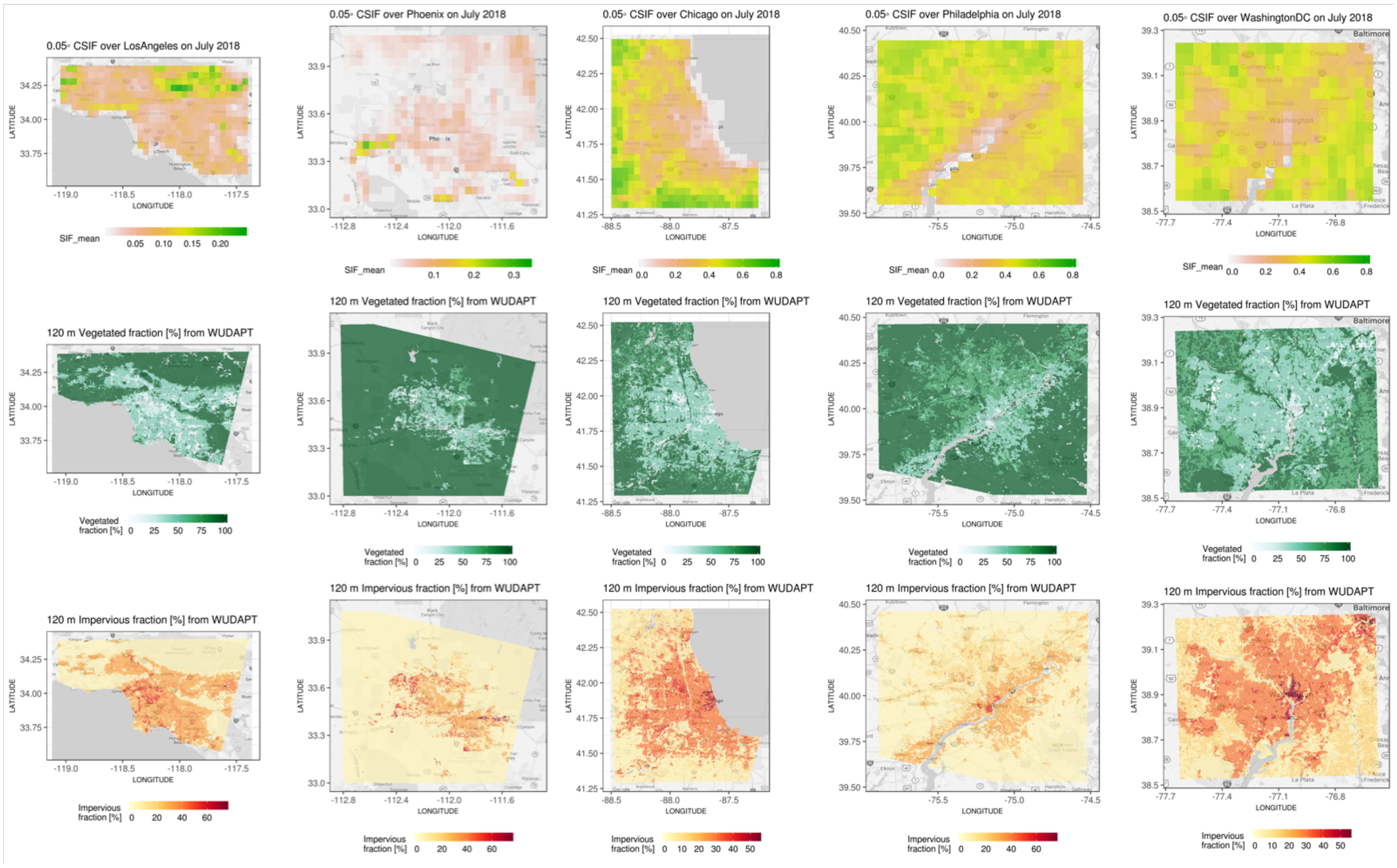

Figure S3. Monthly mean 0.05° CSIF over July 2018 (1st row) and 120 m vegetated fractions and impervious fractions from WUDAPT (2nd and 3rd rows) for Los Angeles, Phoenix, Chicago, Philadelphia, and Washington DC.

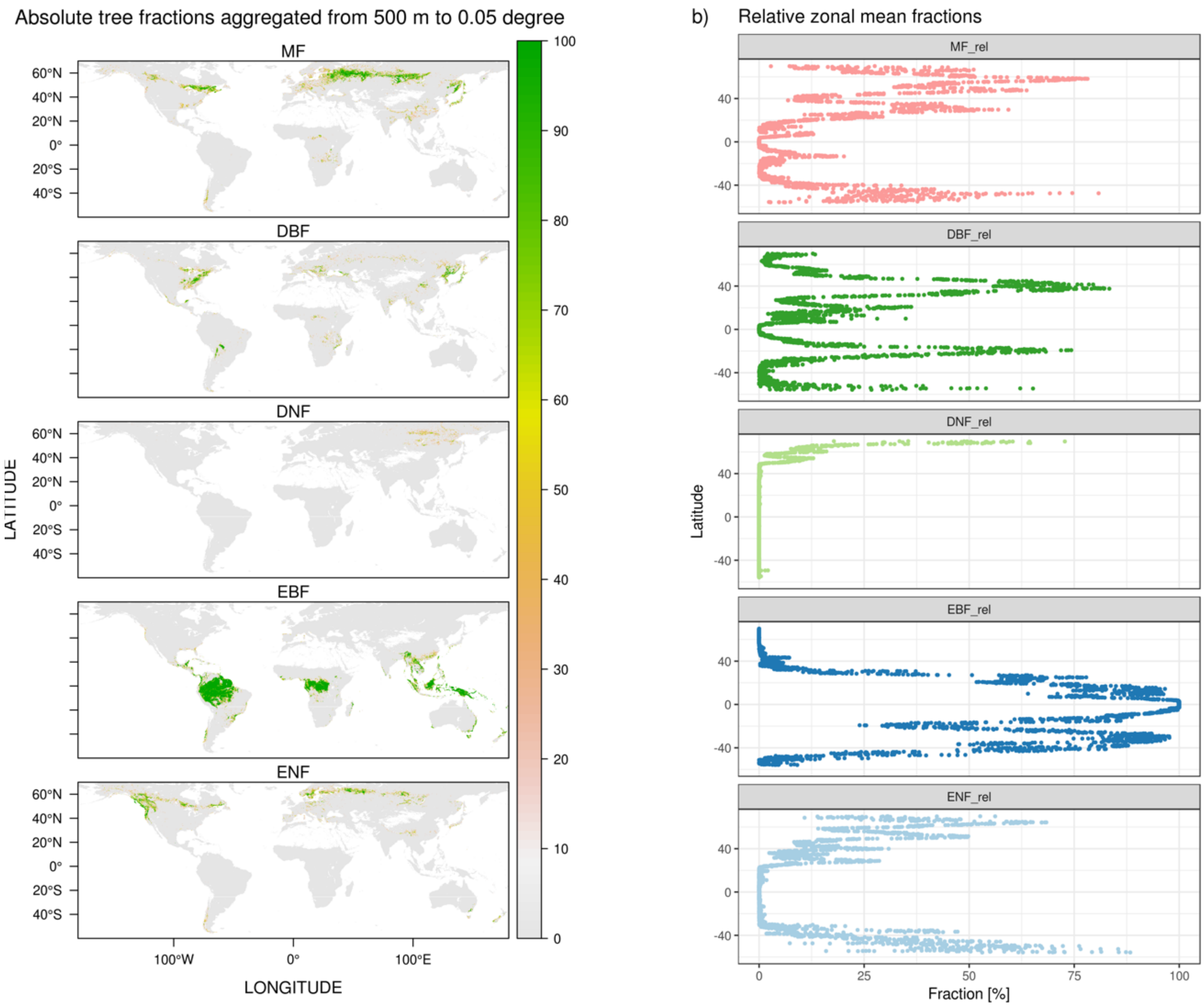

Figure S4. Aggregated fractions of mixed (MF), deciduous broadleaf (DBF), deciduous needleleaf (DNF), evergreen broadleaf (EBF), and evergreen needleleaf (ENF) forests at 0.05° according to 500 m MCD12Q1 (**a**) and the resultant relative fractions averaged over the longitude (zonal mean fractions in **b**). These fractions are used as a climatological latitude-dependence for approximating urban vegetation types and fractions.

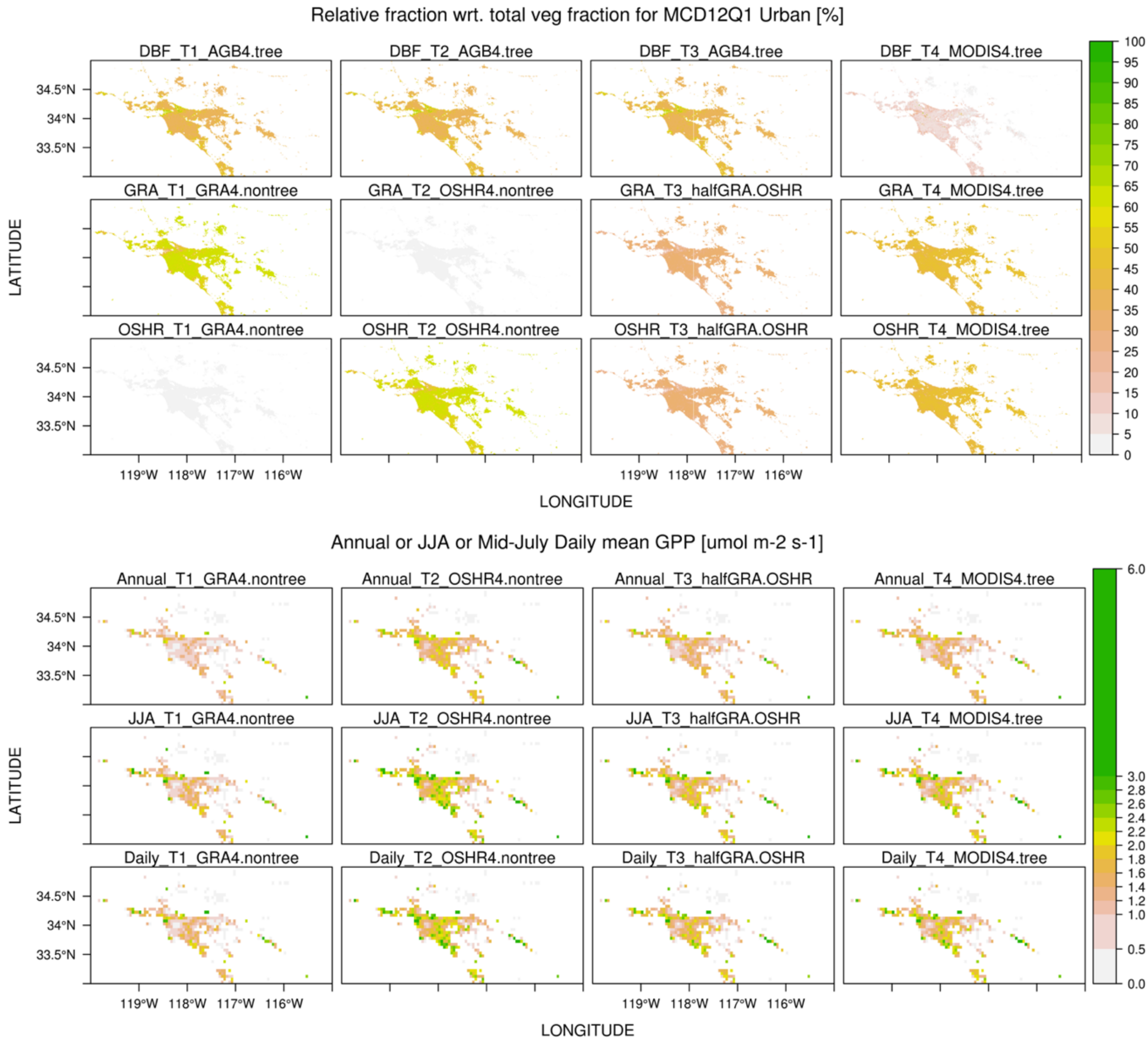

Figure S5. A sensitivity test of the impact of different assumptions for predicting vegetation fractions (upper panels) on final annual, summertime, and daily mean GPP fluxes (lower panels). Each column represents a separate test as follows:

1st column (T1): DBF fractions based on AGB-derived tree fractions, non-tree fractions as **GRA** (i.e., no OSHR);
2nd column (T2): DBF fractions based on AGB-derived tree fractions, non-tree fractions as **OSHR** (i.e., no GRA);
3rd column (T3): DBF fractions based on AGB-derived tree fractions, non-tree fractions as **half GRA & half OSHR**;
4th column (T4): DBF fractions based on MOD44B tree fractions, treat non-tree fractions as **half GRA & half OSHR**.

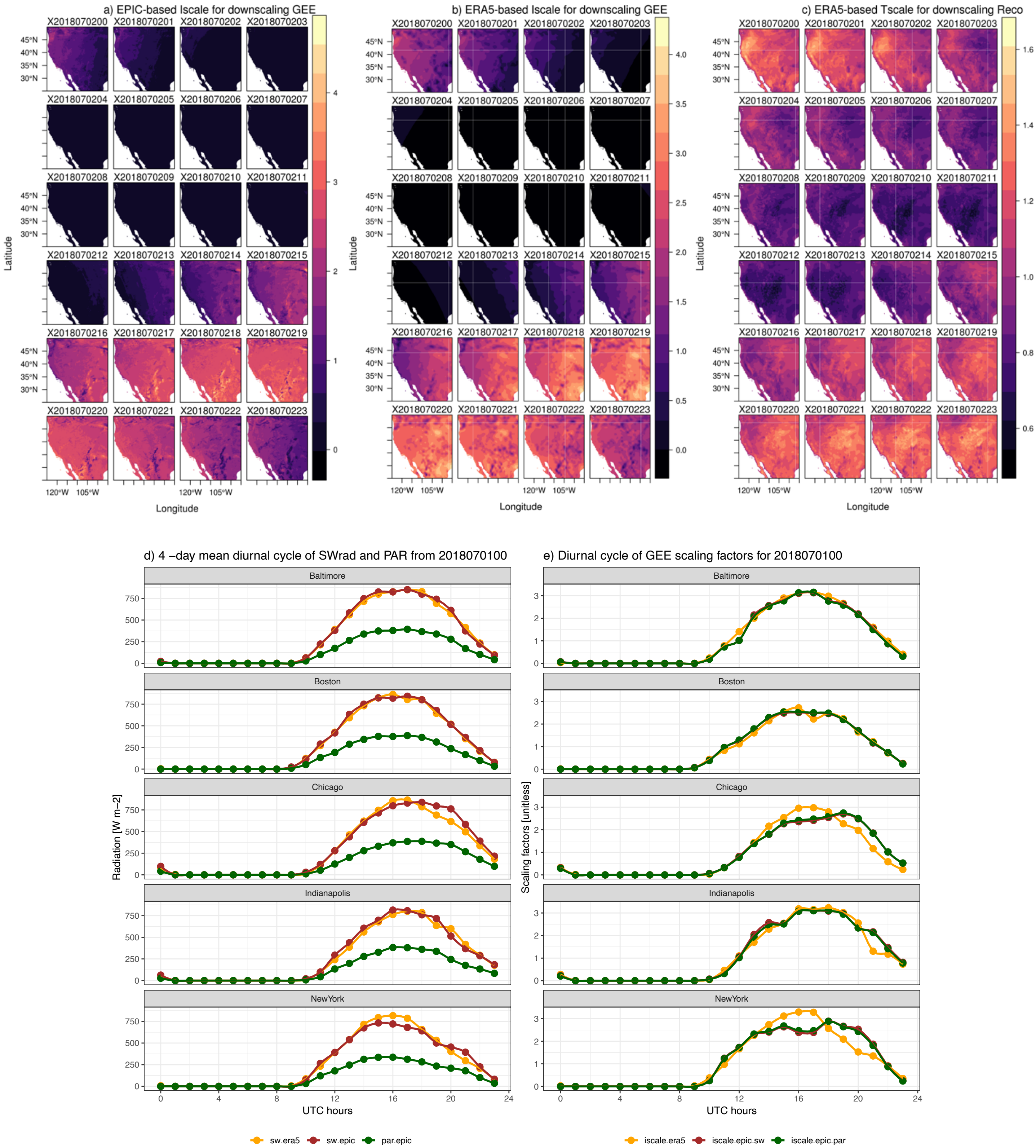

Figure S6. Examples of hourly scaling factors for GPP and R_{eco} over western CONUS and selected cities in eastern CONUS. I_{scale} and T_{scale} are calculated based on SW radiation received at the surface with account of cloud coverage using EPIC (panel a) or ERA5 (panel b) and the ERA5-based air temperature dependent Q10 functions (c) for the western CONUS on 07/02/2018. UTC times are denoted above each panel (e.g., 1200 UTC = 0800 ET; 16UTC = 1200 ET, etc) **d-e)** 4-day mean SW_{rad} (orange or red lines) or PAR (dark green lines) from two models (d) as well as hourly I_{scale} calculated using those radiation products (e), interpolated to 5 eastern US cities during July 1[st] – July 4[th], 2018. For example, I_{scale} is simply zero between 04:00 and 14:00 UTC for the western US while T_{scale} is usually smaller during nighttime than daytime. Consequently, plants may begin to photosynthesize at ~14:00 UTC (07:00 – 08:00 local time on this particular day), while NEE may remain positive due to opposing effects from R_{eco} and then become uptake.

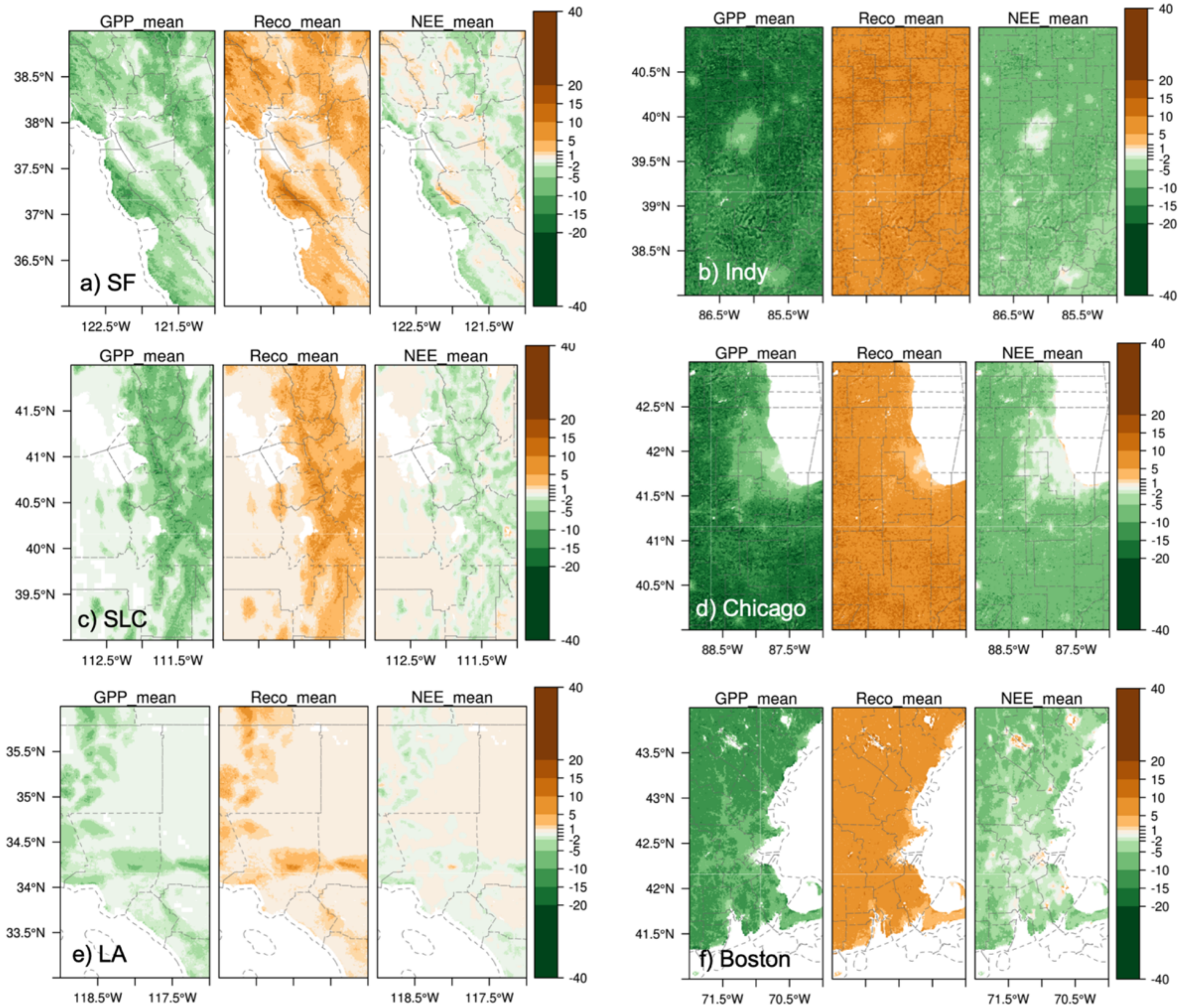

Figure S7. MOD44B VCF-downscaled fluxes at 1km grid spacing [µmol m^{-2} s^{-1}] zoomed into selected western vs. eastern CONUS cities, averaged over JJA 2018. Colour scales are consistent among 6 urban areas.

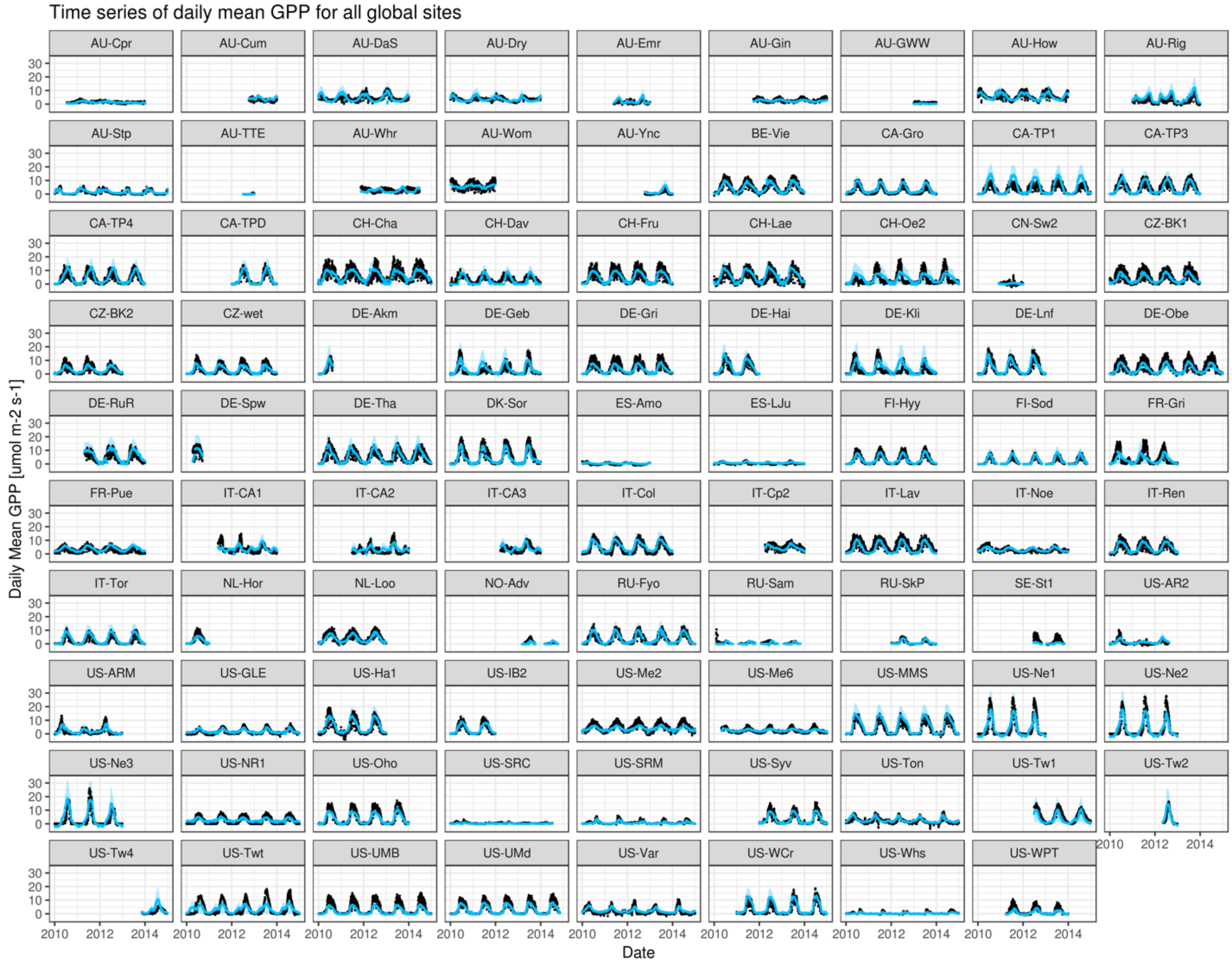

Figure S8. Time series of 4-day mean GPP between SMUrF (blue lines with uncertainties in blue ribbons) and FLUXNET (black dots) for 89 global EC sites from 2010 to 2014. Note that these modeled GPP are directly computed using biome-specific GPP-SIF slopes and SIF extracted at EC sites locations before aggregating to 0.05°. These directly computed fluxes differ from the values extracted from the 0.05° gridded fields (i.e., hourly NEE shown in **Fig. S10**) that integrate spatial variations in biome types in each 0.05° model grid.

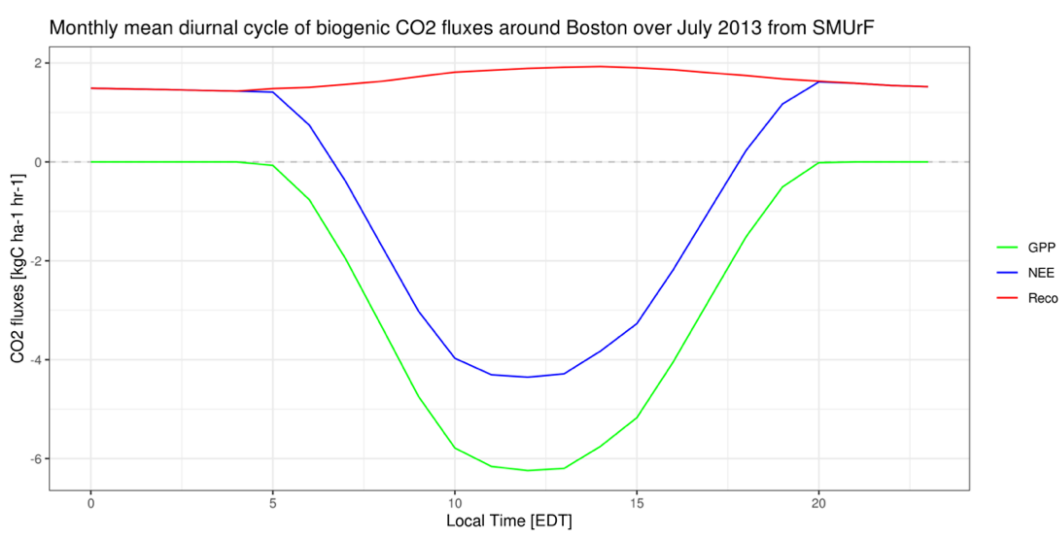

Figure S9. Monthly mean diurnal cycle of GEE, R_{eco} and NEE fluxes from SMUrF over the same Boston area considered in Hardiman et al. (2017) in July 2013. Fluxes units have been converted from μmol m^{-2} s^{-1} to kgC ha^{-1} hr^{-1} that was used in Hardiman et al. (2017).

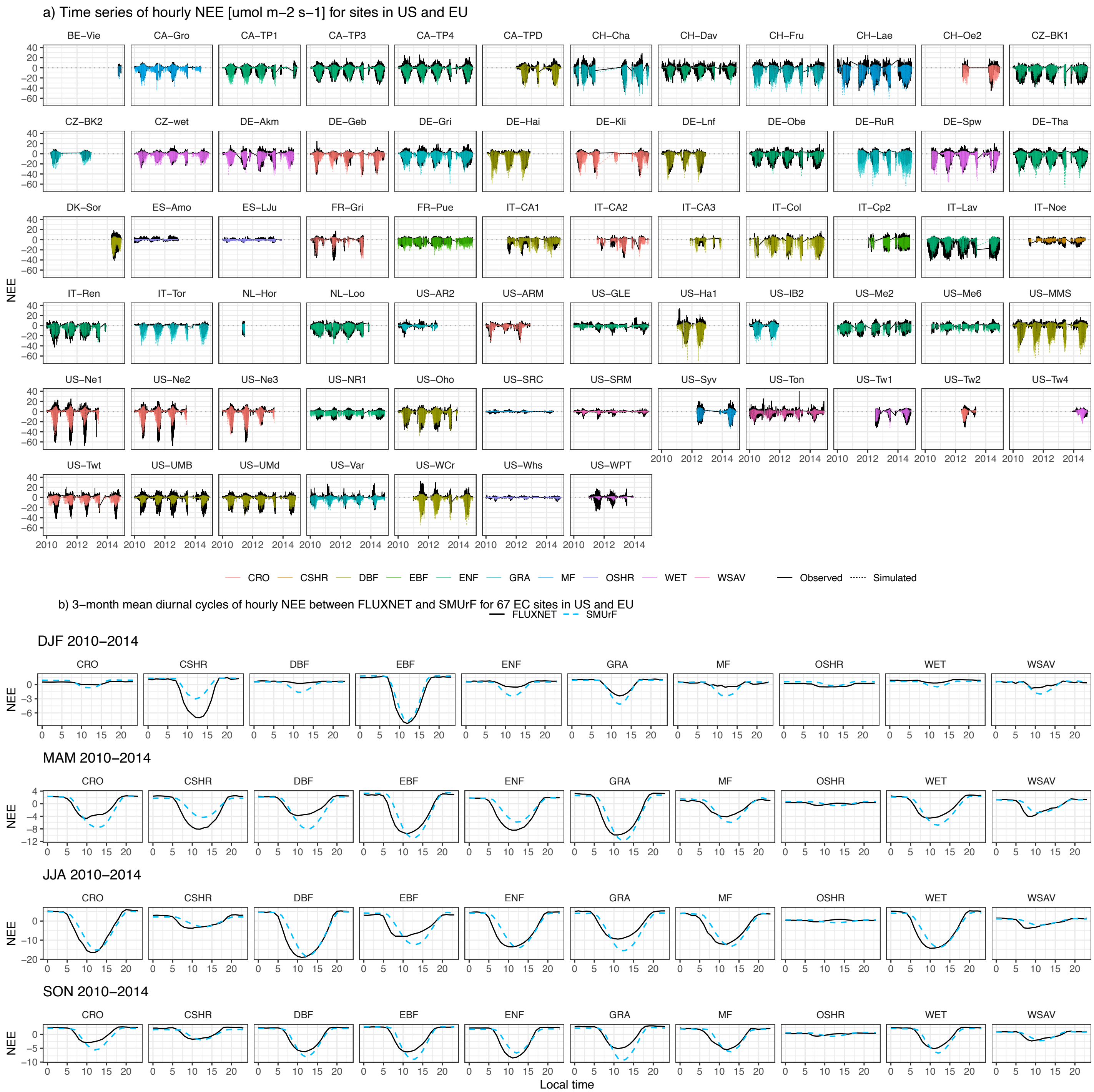

Figure S10. a) Time series of model-data comparisons of hourly NEE between modeled values extracted from the 0.05° SMUrF field (coloured dashed lines with colours differentiating biome types) and FLUXNET (black solid lines) during 2010-2014. Only non-gapfilled measured NEE data from 67 EC sites in US and Europe has been selected for the comparison. **b)** 3-month mean diurnal cycles of modeled (dashed blue lines) and measured NEE (black solid lines) per biome for the same sites.

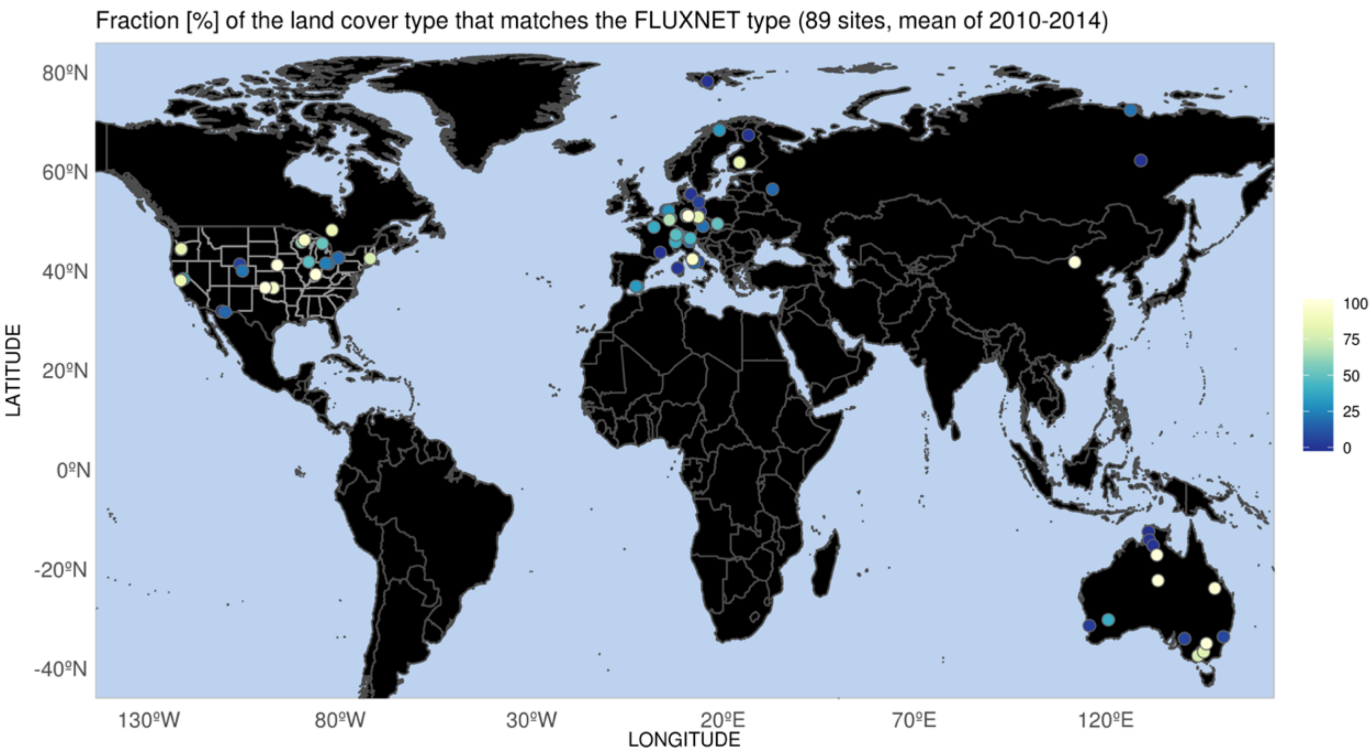

Figure S11. A simple measure of spatial heterogeneity in the 0.05° model grid cell, i.e., the percentage of the land cover types that matches those indicated by EC sites from FLUXNET2015 over all the land cover types at 500 m according to MCD12Q1.

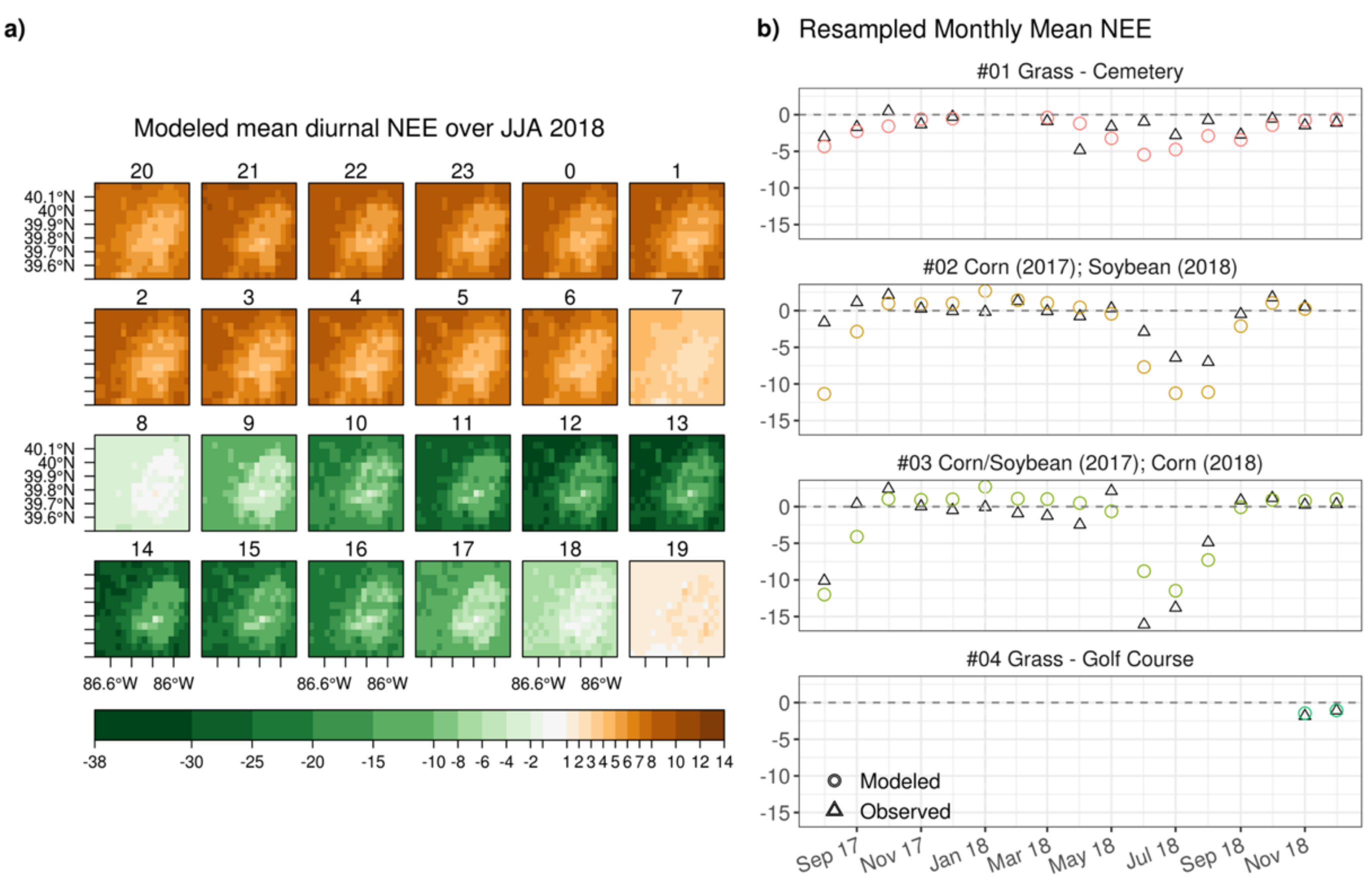

Figure S12. a) Modeled summertime mean diurnal cycles of NEE fluxes over JJA 2018 around Indianapolis. b) Monthly NEE comparisons [μmol m^{-2} s^{-1}] between model (coloured triangles) and observations (black circles) per INFLUX site from Aug 2017 to Dec 2018.

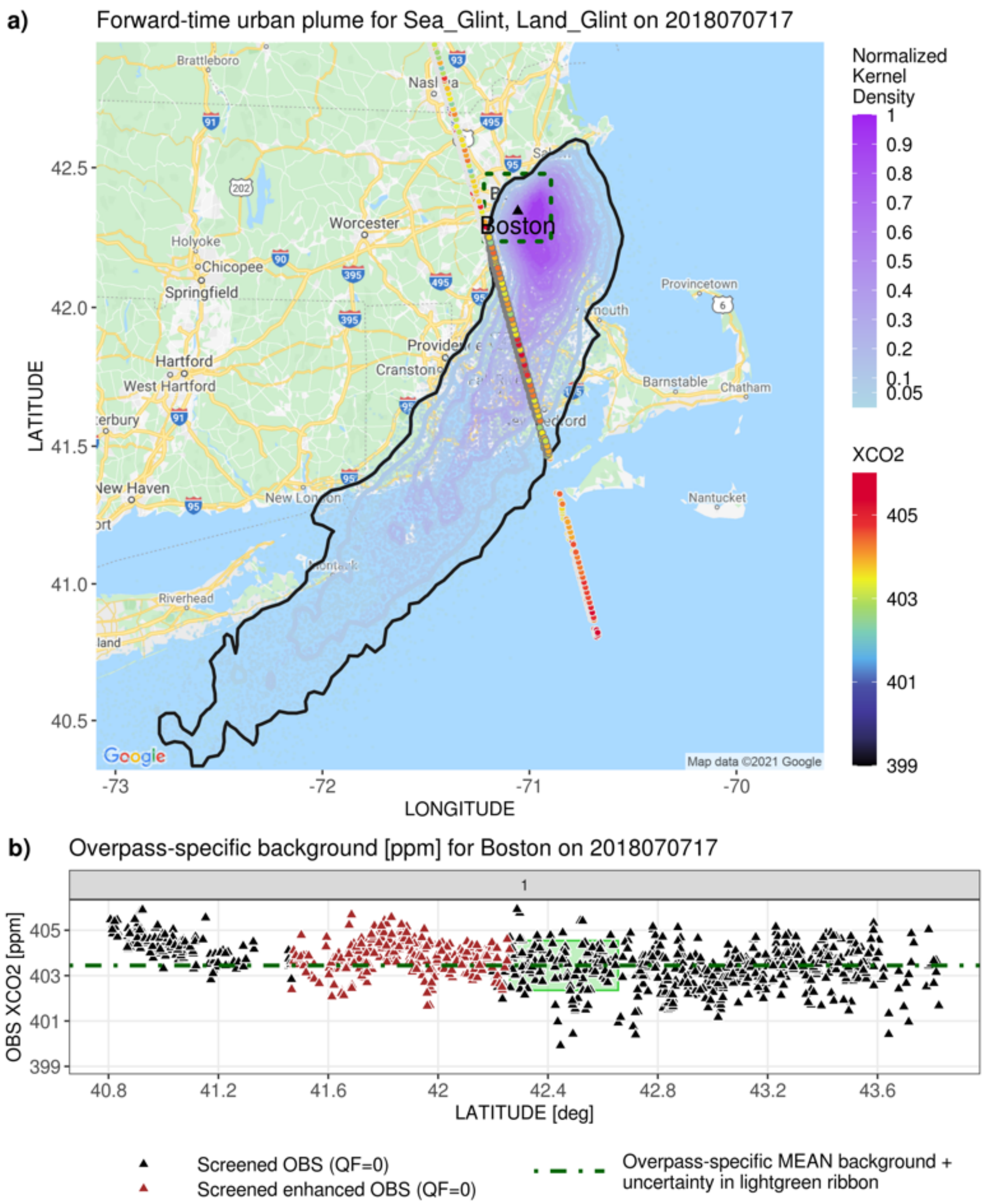

Figure S13. The same overpass-specific approach proposed in Wu et al. (2018) for calculating background on July 7th, 2018. **(a)** Forward particle distributions with random wind error of 2.4 m/s included (blue and purple dots) and their derived normalized 2-D kernel density (purple contours) during the OCO-2 overpass time (~3 mins) with observed XCO_2. Note that the random wind error is derived based on modeled u-/v- wind speed and observed wind fields from radiosonde stations around Boston. Urban plumes are defined based on 5% of the max 2-D kernel density estimated (black solid line). **(b)** Latitude-series of observed XCO_2 after the quality filter (QF = 0) with urban enhanced soundings highlighted in red. The constant background value is shown as dotted-dashed green line. The background uncertainty (green shaded area) includes both the spatial uncertainty and the retrieval uncertainty of observations over the background latitude range. Please refer to Wu et al. (2018) for more details.

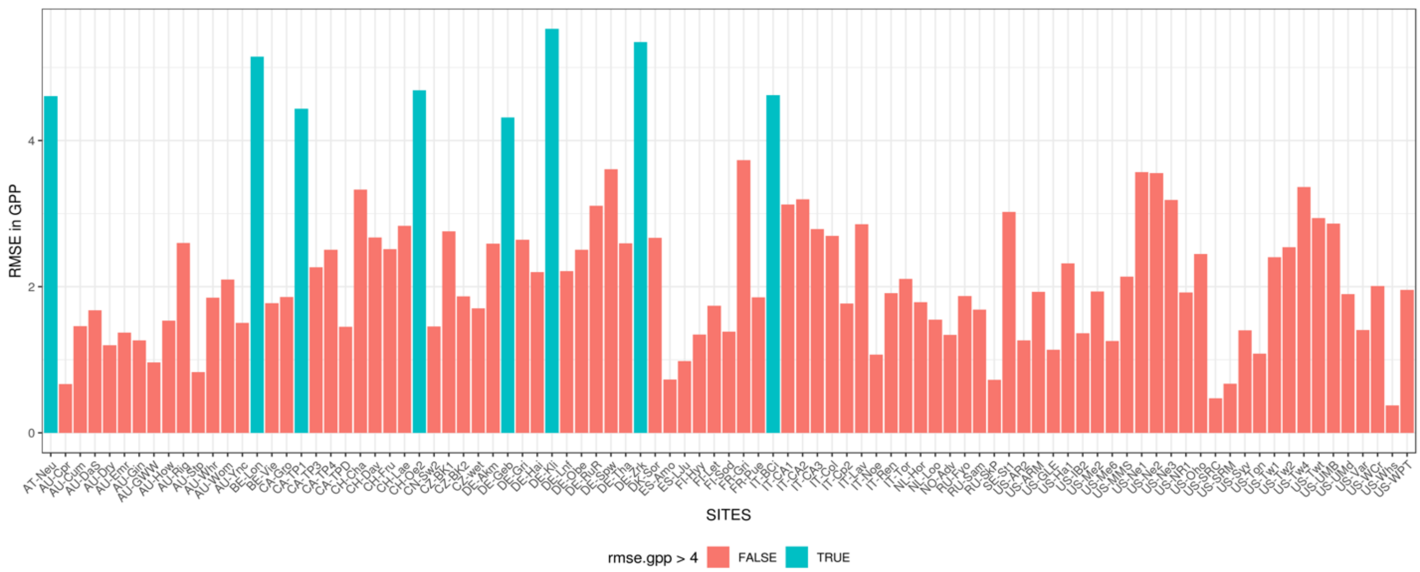

Figure S14. RMSE between modeled versus observed daily mean GPP for all EC sites. Sites with relatively large GPP errors have been removed before training R_{eco}.

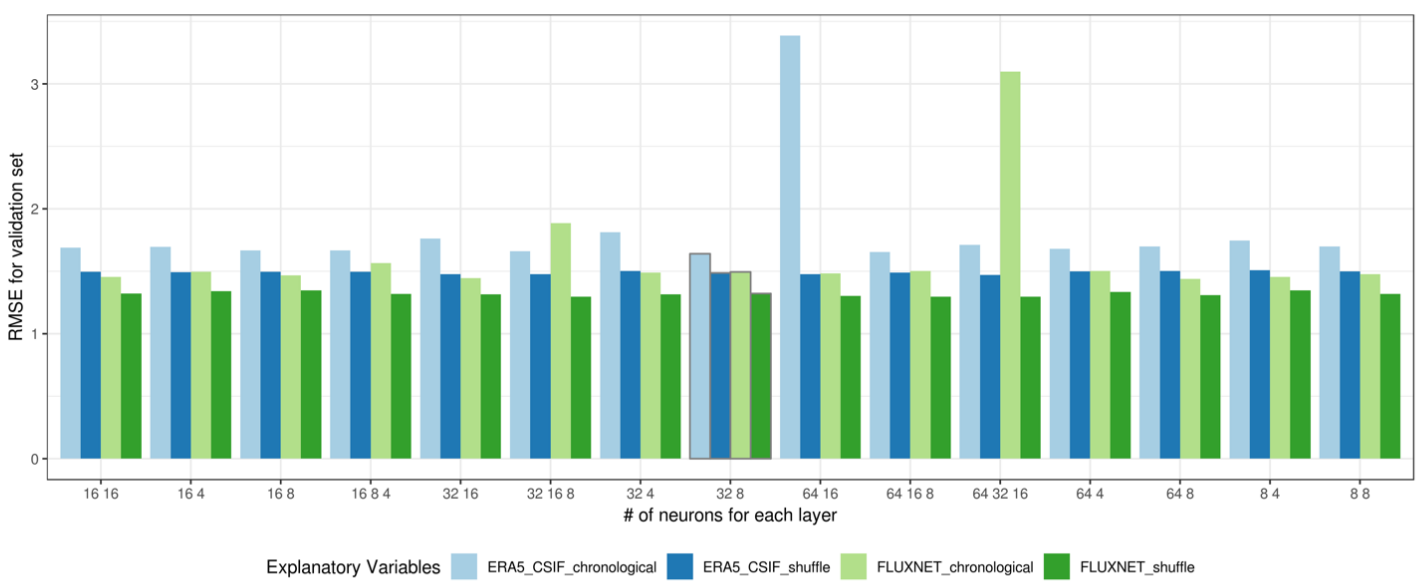

Figure S15. Overall RMSE between predicted and observed R_{eco} using different numbers of neurons per layers (labeled as x-axis) via a 5-fold cross validation following the M3 approach described in **Appendix B**. We tested four sets of neural network models using

1. ERA5-based temperatures and CSIF-based GPP **without** data shuffling before validation, i.e., data arranged in a chronological order from year 2010 to 2014 (‘*ERA5_CSIF_chronological*’);
2. ERA5-based temperatures and CSIF-based GPP **with** data shuffling before validation (‘*ERA5_CSIF_shuffle*’);
3. FLUXNET2015-based temperatures and GPP **without** data shuffling (‘*FLUXNET_chronological*’); and
4. FLUXNET-based temperatures and GPP **with** data shuffling (‘*FLUXNET_shuffle*’).

Because this is a 5-fold cross validation, 5 RMSE for 5 holdouts have been aggregated to an overall value as shown in one bar. If data had not been shuffled before training, it is arranged in a chronological order. We ended up using 32 neurons for the first hidden layer and 8 neurons for the second layer as outlined in gray.