Supplementary Materials to ***Parsimonious gap-filling models for sub-daily actual evapo-transpiration observations from eddy-covariance systems*** **by Guo et al.**

Table S1.1. Existing approaches to infill gaps in latent heat flux, carbon flux or directly for ET_a. Orange cells highlight models that rely on additional input variable other than the variable to infill. Green cells highlight the only two existing parsimonious gap-filling models, the mean diurnal variation (MDV) and the analogue period (AP).

Ref.	Variable to infill	Approach	Input Data required	Case study used
[1]	Net ecosystem CO2 exchange (NEE) flux	Mean diurnal variation (MDV)	**This method only uses the variable to infill itself.** Gaps are filled by averaging valid values at the same time of measurement in m days before and m days ahead of the day with gaps, within a (at most) 2-week window around the gap.	18 sites from EUROFLUX and AmeriFlux over crops, grasslands and conifer and broad-leaved forests
		Lookup table (multi-variate)	NEE data are binned by **meteorological conditions** and missing values are filled with available records under similar meteorological conditions.	
[2]	Latent heat flux (LE)	Analog period (AP)	**This method only uses the variable to infill itself.** The full data record is scanned to identify similar periods similar to the periods just preceding and following the gaps. These periods are then used in simple or multiple regressions to fill the missing information. The simplest form uses only LE data itself.	Six AmeriFlux stations with 3 in forests (1 Mediterranean, 2 * humid subtropical), 1 in cropland (humid continental) and 1 in shrubland (Mediterranean)
[3]	Net ecosystem CO2 exchange (NEE) flux	Non-linear regressions	The infilling is based on NEE = GP – ER, where **meteorological data** are used to predict each component.	Six forested European sites that are representative of European forests and climates, including Mediterranean, deciduous broadleaf, and evergreen coniferous sites over a 20° latitudinal range.
		Unscented Kalman filter (UKF)	**Meteorological data** are used to predict NEE and serial correlation is used to continuously improve the prediction.	
		ANN	This is a black-box non-linear regression model to predict NEE from **meteorological data.**	
		Lookup table (LUT)	NEE data are binned by meteorological conditions and missing values filled with available records under similar **meteorological conditions.**	
		Marginal distribution sampling (MDS)	This is an improvement of standard LUT, where similar **meteorological conditions** (of a fixed margin) are sampled in the temporal vicinity of the gap to be filled.	
		Semi-parametric model (SPM)	This method uses a 3-dimensional, non-linear look-up table sorted with **meteorological variables** and time as a continuous representation of NEE.	
		Mean diurnal variation (MDV)	**This method only uses the variable to infill itself.** Missing NEE value for a certain 0.5-h is replaced with the averaged value of the adjacent days at exactly that time of day	
		Multiple imputation (MI) method	The method uses **multivariate correlation** to replace the missing NEE data with several simulated (imputed) values as a distribution, and taking the mean of the distribution	

		Terrestrial biosphere model	This is a process-based model to predict NEE with **meteorological data and LAI, soil type, texture, depth, canopy height and tower height**	
[4]	NEE flux	Look up table (LUT)	The fluxes are binned by similar **meteorological conditions** within a certain time window. The missing value of the flux is then calculated as the average value of the binned records.	25 sites from the LaThuile FLUXNET dataset which covers cropland, various broadleaf/needleleaf forest, grassland, shrubland, wetland and wood savanna.
		mean diurnal course (MDC) – equivalent to MDV	**This method only uses the variable to infill itself.** Autocorrelation of the fluxes is exploited by taking the average value at the same time of day within a moving time window of adjacent days	
		Sampling from marginal distribution of NEE and climate variables	The infilling is based on the covariation of the fluxes with the **meteorological variables** and their temporal autocorrelation	
[5]	Latent heat flux (LE)	Simple linear regression	A linear function is developed between LE and Rn thus **needing solar radiation**	Native forest in Middle Rio Grande in New Mexico
[6]	Latent heat flux (LE)	ERddyProc - Mean diurnal variation (MDV) and Lookup table (LUT)	The method fills missing LE data with those collected under similar **meteorological conditions** or with averaged values over adjacent days.	Three wheat crop fields – two sloping ones and one flat one in NE Tunisia (subhumid Mediterranean)
		Multiple linear regression	A regression is developed between LE **and other energy fluxes.**	
		Evaporative Fraction (EF)	The infilling is based on EF = LE/(Rn-G). Assuming EF at midday is statistically representative of daily EF, and infill missing LE based on EF (**requiring other energy fluxes**).	
[7]	Latent heat flux (LE)	Mean diurnal variation (MDV)	**This method only uses the variable to infill itself.** A missing observation is replaced by the mean for that time of day based on observations from the previous and subsequent days.	18ha winter wheat in NW of Guelph, Ontario, Canada.
		Multiple regression	A regression was developed between LE and **available energy flux (Rn-G) and VPD**	
		Two-week average Priestley and Taylor coefficient	Missing measurements of half hourly latent heat flux were estimated using the product of equilibrium evaporation for the half hour and the 2-week average Priestley and Taylor coefficient (**needing meteorological variables and energy fluxes**).	
		Multiple imputation (MI) method	The method used the distributions of **meteorological variables** and LE to impute LE distribution.	
		Kalman filter applied to a dynamic linear regression	A parametric model was developed to relate temporal variations of LE with **net radiation and VPD.**	

[8]	Latent heat flux (LE)	Non-linear interpolation of LE with either Multiple regressions (MRS) and K-nearest neighbours (KNNs) informed by a principal component analysis (PCA)	The infilling used input environmental variables including **meteorological variables, soil water deficit and LAI.**	Forest with mixed evergreens and hardwoods in central Taiwan (humid subtropical)
[9]	Carbon, vapour and sensible heat fluxes	Self-organizing linear output (SOLO) artificial neural network (ANN)	A self-organising feature map (SOFM) is constructed from the **meteorological variables** to identify dependencies between predictor variables. Then the responses of measured fluxes can be predicted with the SOFM.	A savanna woodland within tropical arid zone in central Australia
[10]	Daily ET data	Feed-forward (FF) artificial neural networks (ANN) with different climate inputs	An ANN is developed between ET and **meteorological data**.	Saltcedar forest in New Mexico, SW of United States

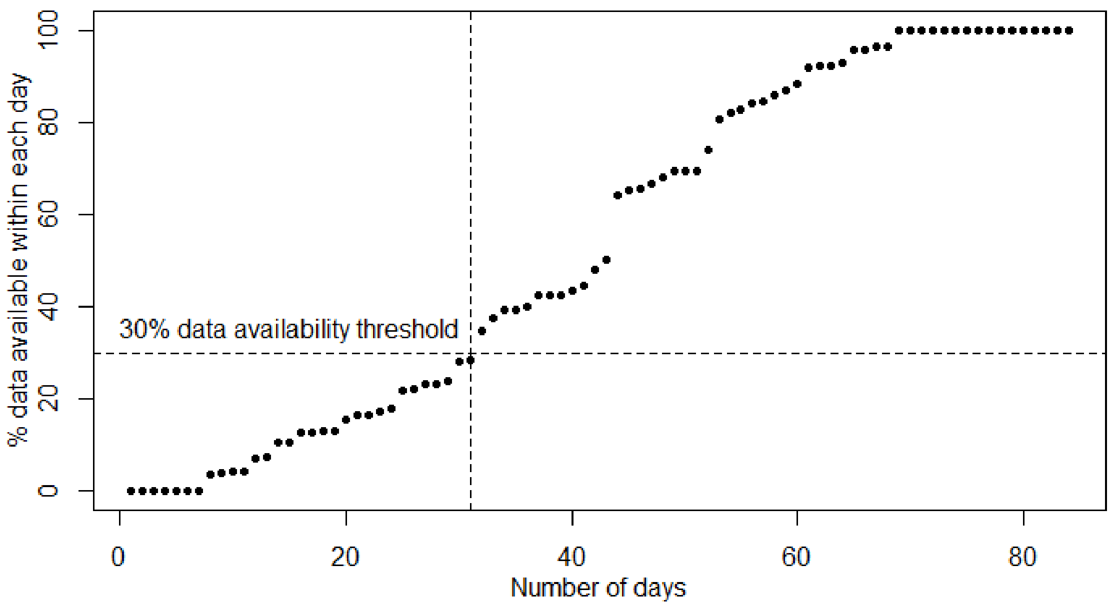

Figure S1.1. Percentage 30-min ETa data availability within each day, sorted from the lowest to highest across the full monitoring dataset.

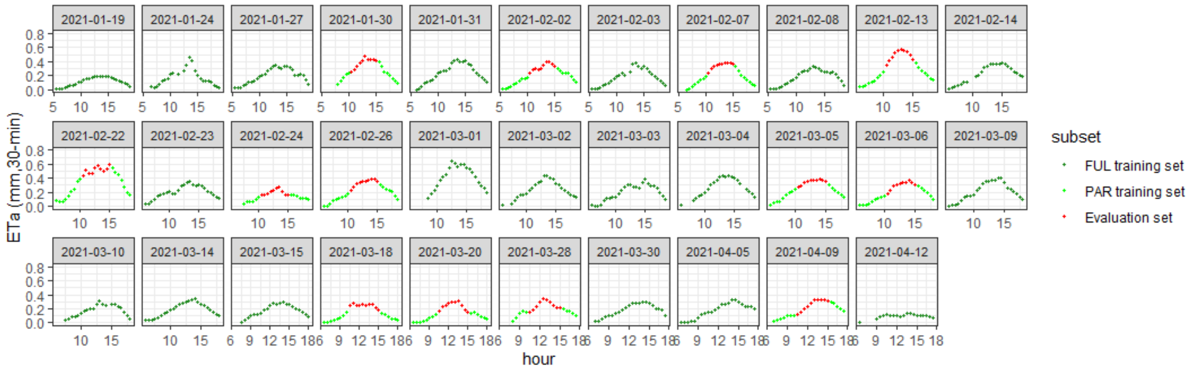

Figure S1.2. Split of the training and evaluation subsets to represent missing data Types B i.e., missing mid-day.

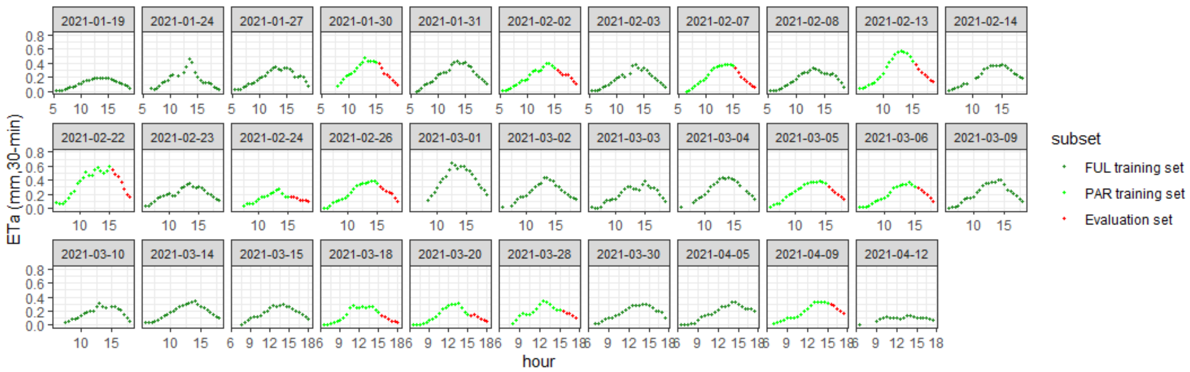

Figure S1.3. Split of the training and evaluation subsets to represent missing data Types C i.e., missing afternoon.

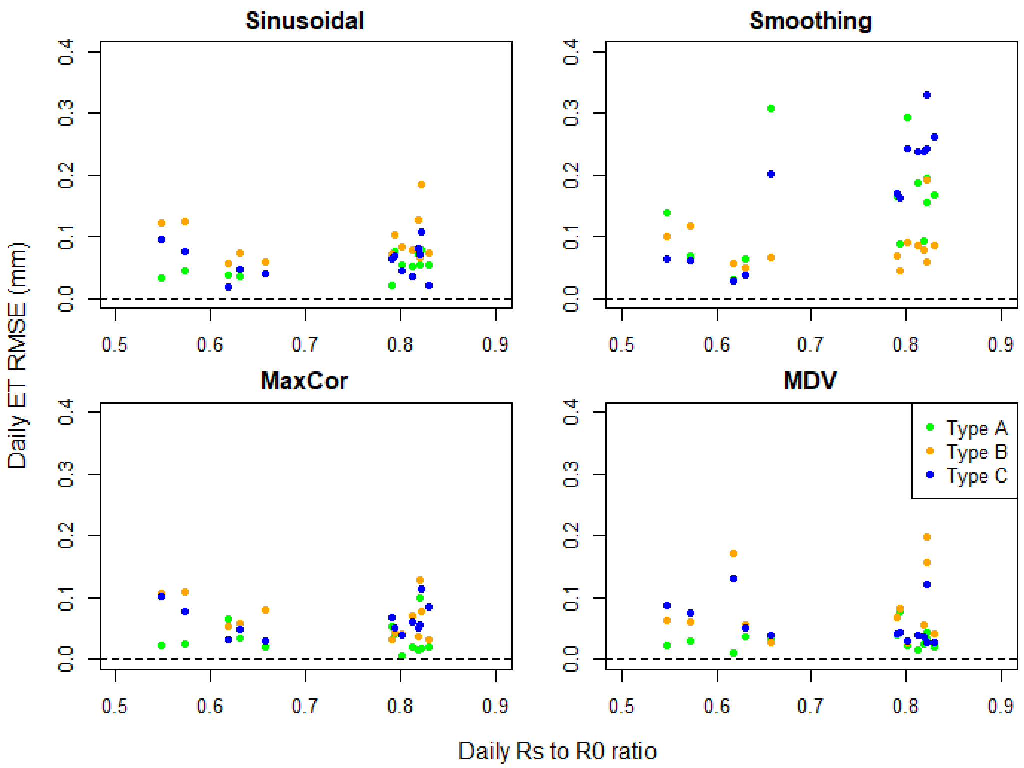

Figure S1.4. Daily RMSE of the four gap-filling models under the three typical patterns of missing data (A – missing morning; B – missing mid-day; and C – missing afternoon), plotted against the daily ratio of actual solar radiation to clear-sky solar radiation. Each panel shows one gap-filling model where the three missing data types are differentiated by colours.

References:

1. Falge, E.; Baldocchi, D.; Olson, R.; Anthoni, P.; Aubinet, M.; Bernhofer, C.; Burba, G.; Ceulemans, R.J.; Clement, R.; Dolman, A.; et al. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* **2001**, *107*, 43–69, https://doi.org/10.1016/s0168-1923(00)00225-2.
2. Hoeltgebaum, L.E.B.; Dias, N.L.; Costa, M.A. An analog period method for gap-filling of latent heat flux measurements. *Hydrol. Process.* **2021**, *35*, e14105, https://doi.org/10.1002/hyp.14105.
3. Moffat, A.M.; Papale, D.; Reichstein, M.; Hollinger, D.Y.; Richardson, A.D.; Barr, A.G.; Beckstein, C.; Braswell, B.; Churkina, G.; Desai, A.R.; et al. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* **2007**, *147*, 209–232, https://doi.org/10.1016/j.agrformet.2007.08.011.
4. Wutzler, T.; Lucas-Moffat, A.; Migliavacca, M.; Knauer, J.; Sickel, K.; Šigut, L.; Menzer, O.; Reichstein, M. Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences* **2018**, *15*, 5015–5030, https://doi.org/10.5194/bg-15-5015-2018.
5. Cleverly, J.; Dahm, C.N.; Thibault, J.R.; Gilroy, D.J.; Coonrod, J.E.A. Seasonal estimates of actual evapo-transpiration from Tamarix ramosissima stands using three-dimensional eddy covariance. *J. Arid Environ.* **2002**, *52*, 181–197, https://doi.org/10.1006/jare.2002.0972.
6. Boudhina, N.; Zitouna-Chebbi, R.; Mekki, I.; Jacob, F.; Ben Mechlia, N.; Masmoudi, M.; Prévot, L. Evaluating four gap-filling methods for eddy covariance measurements of evapotranspiration over hilly crop fields. *Geosci. Instrum. Methods Data Syst.* **2018**, *7*, 151–167, https://doi.org/10.5194/gi-7-151-2018.
7. Alavi, N.; Warland, J.S.; Berg, A. Filling gaps in evapotranspiration measurements for water budget studies: Evaluation of a Kalman filtering approach. *Agric. For. Meteorol.* **2006**, *141*, 57–66, https://doi.org/10.1016/j.agrformet.2006.09.011.
8. Chen, Y.-Y.; Chu, C.-R.; Li, M.-H. A gap-filling model for eddy covariance latent heat flux: Estimating evapotranspiration of a subtropical seasonal evergreen broad-leaved forest as an example. *J. Hydrol.* **2012**, *468–469*, 101–110, https://doi.org/10.1016/j.jhydrol.2012.08.026.
9. Eamus, D.; Cleverly, J.; Boulain, N.; Grant, N.; Faux, R.; Villalobos-Vega, R. Carbon and water fluxes in an arid-zone Acacia savanna woodland: An analyses of seasonal patterns and responses to rainfall events. *Agric. For. Meteorol.* **2013**, 182, 225–238.
10. Abudu, S.; Bawazir, A.S.; King, J.P. Infilling Missing Daily Evapotranspiration Data Using Neural Networks. *J. Irrig. Drain. Eng.* **2010**, *136*, 317–325, https://doi.org/10.1061/(asce)ir.1943-4774.0000197.