Supplementary Material

A The effect of gap-filled data on the driver analysis

Analysis of canopy conductance relies on daily flux estimates. However, in Eddy Covariance measurements, portions of the daily data are frequently filtered due to the occurrence of stable boundary layer conditions. To obtain unbiased daily values, these gaps must be filled using interpolation, regression, or machine learning approaches. Such gap-filling procedures, however, inherently impose a relationship between the predictors and the target variable.

This issue can be partially mitigated by performing gap-filling at the half-hourly time scale and incorporating a broader set of predictors than those used in the subsequent driver analysis, followed by aggregation to daily values. This approach aims to minimize the influence of modeled values, as well as the impact of individual drivers within the gap-filling models. Nevertheless, the effect of gap-filling cannot be entirely disregarded.

Quantifying the influence of gap-filling is challenging for several reasons. Restricting analyses to observed data alone is impractical at the daily scale, as most days contain filtered measurements due to turbulence-based quality control. In our dataset, only 60 days were completely free of gap-filling (Fig. A1). Importantly, only a small fraction of daily aggregates required more than 50% gap-filling, indicating that observed values largely dominate the daily totals.

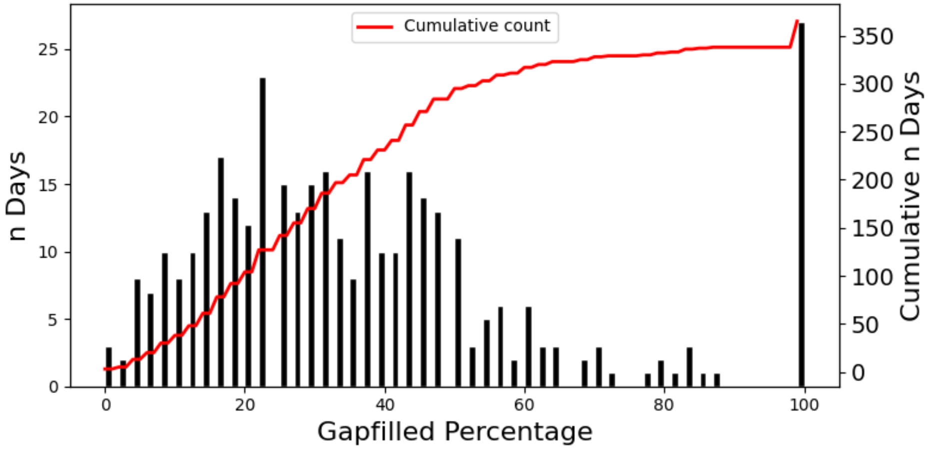

Figure A1: Histogram showing the amount of days with percentages of data filled with models. The red curve shows the cumulative number of days of the year when successively including days with more gap-filled data

One approach to limit the effect of modeled datapoints is to impose a threshold of maximum daily gap-fills to filter out data. However, the distribution of missing data is not necessarily evenly distributed throughout the year. In our case most missing values are in February, July, August and September and the mean of daily missing data percentage is between 30 and 67%.

Therefore, it is likely that a fixed threshold would remove disproportionally many datapoints between July and September, potentially skewing the driver analysis towards important predictors in winter and spring (Fig. A2).

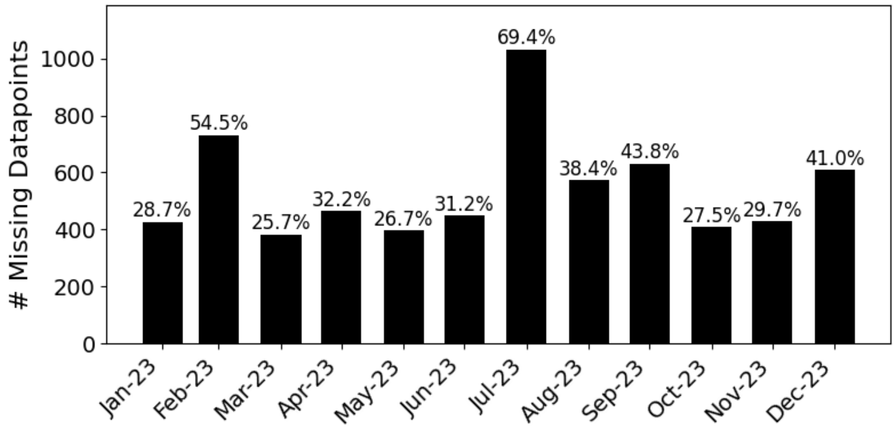

Figure A2: Amount of missing half-hourly datapoints per month with the average amount of missing data per day on top of the bars.

To test the influence of modeled datapoints on our results we repeated the analysis with four different thresholds of 30%, 50%, 75% and 100% of maximum daily gap-filled datapoints allowed. During each run all other datapoints were removed from the dataset.

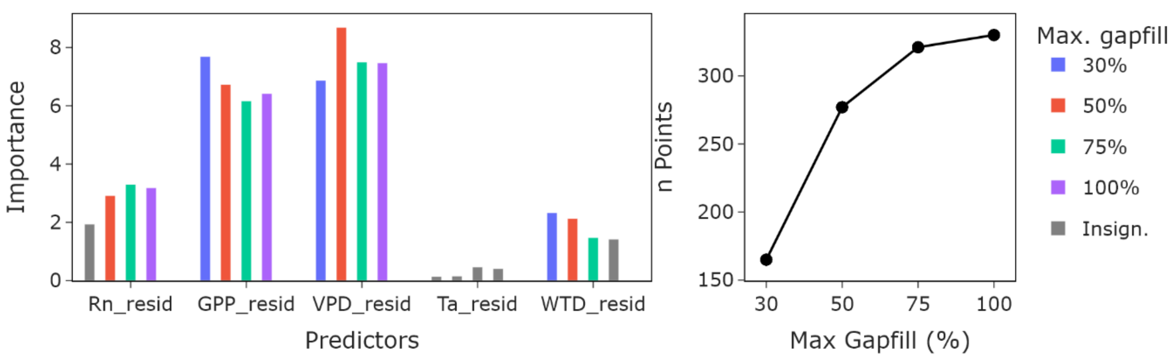

Figure A3: Left: Predictor importances for different imposed maximum allowed percentages of filled data per day. Grey bars indicate insignificance of the predictor. Right: Number of datapoints included when imposing the thresholds of maximum modeled data per day

Between thresholds of 30 to 100% allowed maximum daily gap-filled data we find that the driver importances are very similar from 100 to 50% (Fig. A3). We see an increase in the importance of VPD from ca. 7.5 to 8.5 and a small increase in the importance of WTD. The relative rankings of the driver's importance however remain the same. The strongest shift is observable between the 50% and 30% thresholds. At 30% net radiation (Rn) becomes insignificant, instead the importance of

GPP rises. Since Rn is the main driver of GPP this could be due to increased collinearity in the remaining data. Due to this rise in GPP it overtakes VPD as the most important predictor. Further a small increase in the importance of WTD is observable, however in all cases its weight remains lower than that of the other significant drivers. Temperature (Ta) remains insignificant across all imposed thresholds. The jump from a 50% threshold to 30% threshold results in the strongest filtering of the data (277 to 125 days). Further, at this jump the amount of data removed in late summer rises disproportionately, leading to around 20 to 25 removed days in July to September as well as large parts of December and February, which is likely to result in a skewed driver analysis.

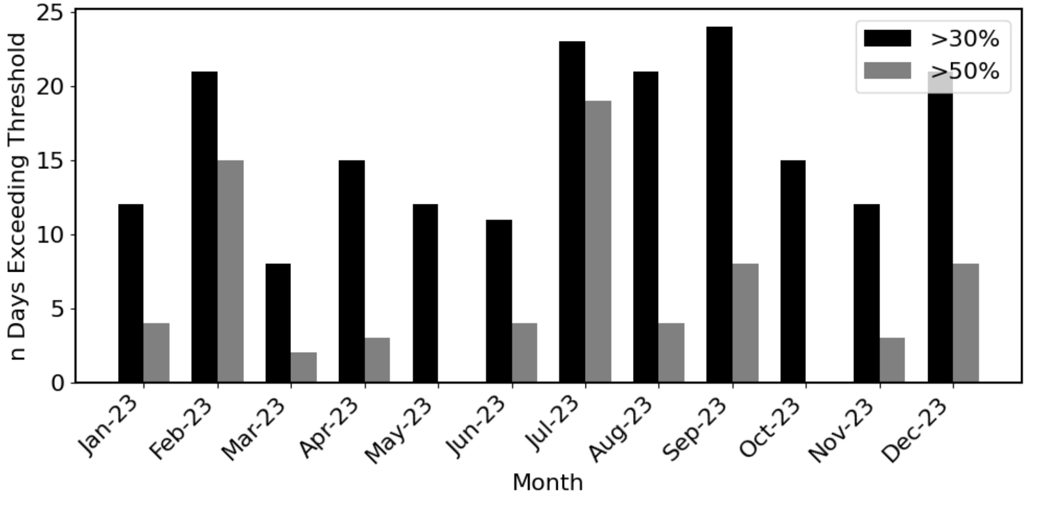

Figure A4: Number of filtered datapoints per month when imposing a threshold of 50% or 30% of maximum gap-filled data per day

Summarizing, we show that while there is some variation in the driver's importance when limiting the amount of filled data, the overall rankings and magnitudes remain largely stable until the threshold is so strict that an un-skewed sampling within the year cannot be guaranteed.