Chapter 6
Data Gap Filling

Dario Papale

6.1 Introduction

The eddy covariance (EC) technique provides data at high temporal resolution, continuously, day and night and potentially for multiple years. Despite the recent developments in the EC technique and the availability of instruments with low power consumption, system failures are unavoidable and create gaps in the measurements. Common problems in the data acquisition are power breaks, in particular when the power system is based on solar panels; damages to instruments, for example, due to animals or lightning; incorrect system calibrations; maintenances; and also human actions like vandalism or robbery. In addition to these events related to the data acquisition phase, there are also gaps introduced by the data quality filtering, where measurements are discarded if acquired under non ideal conditions. Examples of these filters are the raw data tests described in Sects. 3.2.2, 4.3.3 and the nighttime filtering depicted in Sect. 5.3. Falge et al. (2001) found on average 35% of data missing due to system failures and data rejections across 19 EC sites while Papale et al. (2006) estimated that 20–60% of the data was rejected by the different quality filters applied.

Are these gaps a problem in our analyses? When should we fill these gaps in the measured fluxes and which are the methods available? In this chapter, the flux measurement gap filling will be discussed, focusing in particular on the differences between the methods available and providing indications about the best way to fill gaps in the data set on the basis of the data use and ecosystem characteristics.

D. Papale (✉)
DIBAF, University of Tuscia, Viterbo, Italy
e-mail: darpap@unitus.it

M. Aubinet et al. (eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, Springer Atmospheric Sciences, DOI 10.1007/978-94-007-2351-1_6,

6.2 Gap Filling: Why and When Is It Needed?

Do we need to fill the gaps in an EC time-series? It depends on the use of the data and the analysis that we plan to do. Thanks to its high temporal resolution, the EC technique provides a large amount of data that are often acquired under similar situation in terms of vegetation status and meteorological conditions. This “redundancy” of data is fundamental in the gap-filling methods; it is also sufficient to perform specific analyses, when no gap-free data sets are needed. Examples are the analysis of functional relationships between fluxes and drivers or models validation and parameterization when the model time resolution is the same of the EC measurements. In these cases, it is not needed to fill the gaps present in the time-series and only the measured and not-rejected data can be used.

Instead, whenever it is needed to calculate aggregated values, for example, sums to estimate annual budgets or daily averages needed in model evaluations, the completeness of the data set is required. If missing and rejected values in the half-hourly data set would be perfectly random distributed, the calculation of an integrated value could be easily performed by taking the average of all available data. Unfortunately, data gaps do not occur randomly. For example, $u*$ filtering removes mainly nighttime data or power failures occur principally in winter and night when the solar panels are used. This nonrandomness of the gaps in the data set leads to the need to apply more sophisticated gap-filling methods to reconstruct the missing periods.

6.3 Gap-Filling Methods

There are different gap-filling methods, in particular for carbon fluxes, that have been proposed in scientific literature. These can be classified according to different characteristics:

– *Principles*: All the gap-filling methods make use of the valid data to reconstruct the missing period. This reconstruction however can be based on completely empirical techniques or on the use of “functional models.” In the first case, there are no assumptions imposed in the shape of the relations between drivers and fluxes and the data are used to find this relation and parameterize it. In the “functional models,” the knowledge about the process under study is used to prescribe the way how drivers and fluxes are linked and the data are used only to parameterize these functions. In general, functional models are not recommended when the data are used in models evaluation activities because the same knowledge about the processes involved could be used in both the gap-filling method and the model to validate, leading to spurious correlations and circularities. However, if the empirical methods are in general good in the interpolation, they have led to high uncertainty in extrapolation, where the empirical relation found using the available data could be not valid (e.g., filling winter time data with a relation found and parameterized using summer time data). In these cases the functional models are more suitable since the knowledge about the dynamic of the system and the role of the different drivers in the different periods of the year is included in the method.

– *Drivers*: The drivers are variables, which can explain at least partially the variability of the flux measured with the EC technique that needs to be filled. Generally, the meteorological variables are used as drivers in most of the gap-filling methods since they influence the ecosystem responses in terms of carbon, water, energy, and other greenhouse gas fluxes. Incoming shortwave radiation, air and soil temperature, vapor pressure deficit, and soil water content are in general the most used drivers; however, other variables like precipitation, diffuse and reflected radiation, and wind speed can also be important in specific sites or conditions. The gap-filling method flexibility in the requested or accepted drivers could be an important criterion to select the most appropriate technique. Methods that have a fixed list of drivers are clearly less flexible and cannot be applied if one of the drivers is also missing or if a variable that is supposed to be relevant in the flux reconstruction is not included in the model. In general, empirical methods are fully flexible in this respect and, for this reason, preferable in these conditions. There are, however, conditions when all the meteorological data are also missing. In these situations, if it is not possible to reconstruct at least some of the drivers the only method that can be applied is the *Mean Diurnal Variation*.
– *Variables simulated*: The variables that need to be gap-filled could be different. In addition to the fluxes (CO_2, H_2O, Energy, CH_4, N_2O, volatile organic compounds (VOCs) and all the other species that can be measured by the EC technique), the meteorological variables can also be filled, to construct a complete driver data set that can be used as input in the fluxes gap-filling. The ability of the methods to simulate different variables and be available for this reason as a gap-filling tool for different fluxes and meteorological data set should be taken into consideration.
– *Noise conservation*: Fluxes measured with the EC technique are affected by random errors that introduce noise in the data. Most of the gap-filling methods are based on interpolations and for this reason tend to remove the noise signal from the data. There are few methods that conserve the noise in the data, for example, the *Kalman filters* (Gove and Hollinger 2006) and the *Multiple Imputation* (Hui et al. 2004) approaches.
– *Implementation*: The computer computational power available today is more than sufficient to run all the existing gap-filling techniques. However, the implementation of some of the existing methods could be complicated and would need a good knowledge of programming languages. In these cases, the centralized services offered by databases and portals could play an important role, implementing these methods and giving a complete and robust gap-filling tool available to the users (see Sect. 17.3).

6.3.1 Meteorological Data Gap Filling

The gap-filling techniques presented in this chapter have been proposed mainly for CO_2 data; however, part of them, in particular the empirical methods, can be easily adapted to be used for other fluxes. All of them (except *Mean Diurnal Variation*) however require as input meteorological variables that for this reason should be available as continuous and gap-free data set. Although data quality filtering applied to meteorological measurements has minor impact in terms of data points removed, gaps can occur, in particular, due to sensors malfunctioning or power breaks. In these cases, it is needed to first fill the gaps in the drivers and then use gap-filled meteorological data in the fluxes gap-filling. This is clearly a delicate step since errors and uncertainties introduced in the drivers will be reflected also in the fluxes. In addition, it is important to underline that gap-filled meteorological data should not be used in the flux gap-filling model parameterization.

The best way to fill gaps in meteorological data is to have a back-up meteorological station with main variables measured (incoming radiation, air temperature, relative humidity, precipitation, wind speed) close to the main EC tower but independent regarding the power supply system. When this back-up system is not available, the empirical methods described later in this chapter can be used when only part of the meteorological variables are missing using as driver the variables present and additional inputs like top of atmosphere incoming radiation or indicators of date and time.

In the unfortunate but also quite common cases where all the meteorological data are missing and no meteorological stations are available in the area, linear interpolation of variables with slow changes (like temperature) or the use of *Mean Diurnal Variation* method are simplest solutions to implement. Other more complex possibilities that, however, would give more trustable results involve the use of remote sensing data or meteorological reanalysis data. New generation meteorological satellites like the European Meteosat MSG (http://www.esa.int/SPECIALS/MSG/) provide high temporal resolution images (15–30 min) that can be used to derive variables like incoming radiation, surface temperature, or albedo (http://landsaf.meteo.pt/). Finding the regression between the site-level measurements and the data produced for the pixel where the tower is located it is then possible to apply such regression to rescale the remote sensing products at site level when the tower measurements are missing. The same approach can be applied using meteorological reanalysis data instead of remote sensing products. These data are also gridded data set produced integrating observations and models, generally with daily temporal resolution (see as example the ERA-Interim data set produced by ECMWF: http://www.ecmwf.int/research/era/do/get/era-interim) that can be downscaled at local level using site-specific relations parameterized using periods where the variables of interest are present.

6.3.2 General Rules and Strategies (Long Gaps)

Gap-filling method setting and parameterization are crucial steps and they are directly linked to the quality of the results. The relative abundance of data due to the high time resolution of the EC technique and the number of meteorological variables measured should not lead to an underestimation of the importance of this phase that must be carefully implemented. The drivers, for example, should be selected carefully finding the right compromise between the known biological importance of a specific meteorological variable to explain the flux to be reproduced, its presence and quality through the year, and its possible correlation with other drivers used that for some of the methods could lead to an over-parameterization.

In general, the dataset used in the model parameterization should be as much as possible representative of the different conditions with an even distribution of samples measured in the diverse situations. This means, for example, that there should be equilibrium between data acquired during daytime and nighttime or in the different seasons. In addition, also the length of the time-windows used to parameterize and apply the model has an important role. In practice, a model could be parameterized using data from the whole year and then applied to all the present gaps. This, however, must imply that the model is able to distinguish the different "ecosystem states," for example, phenological phases or different agricultural periods (see next section). In fact, fluxes acquired under similar meteorological conditions but during completely different "ecosystem state" could be completely different. As an alternative, the model could be parameterized and applied on the basis of shorter time-windows, in the order of weeks or months, where it is assumed that certain conditions (e.g., phenology or biomass) are stable and the fluxes are explained only by the meteorological conditions. These time windows could be fixed in terms of length and position (e.g., a different model parameterization for each month) or, more sophisticatedly and correctly, be centered on each single gap to fill and have an increasing length, up to the minimum window size that provides a sufficient number of data points to parameterize the model (see as example Reichstein et al. 2005).

One problem, in particular when short time-windows are used, is the presence of long gaps where no data are available for the parameterization. In these cases, when multiple years are available and the ecosystem state did not change in the period, the model can be parameterized using data acquired in the same period of a year (e.g., season) but in different years. The basic assumption that justifies this approach is that the fluxes, in the same period but different years, are mainly function of the meteorological conditions. This is probably true, for example, in mature or old forests that did not experience substantial disturbances or management events or during the growing season of crops when the species and season are the same.

6.3.2.1 Sites with Management and Disturbances

Managed or disturbed ecosystems experience rapid changes of their conditions that drastically change the fluxes and their relations with the meteorological drivers due to changes of the ecosystem dynamics between the period just before and just after the perturbation. This is typical in cropland and managed grassland, in particular, after tillage or after harvesting and grazing (Hammerle et al. 2008; Wohlfahrt et al. 2008) when the green biomass is removed in a few days and rapid vegetation regrowth could start.

These are conditions that should be considered when the gap-filling method is selected and implemented. In theory, the method should to be able to recognize that the system status changed. This could be possible using as driver a variable related to the ecosystem characteristic affected by the disturbance or management practices. In case of harvesting, for example, a spectral reflectance measurement in the spectral bands linked to the green vegetation (e.g., the normalized difference vegetation index (NDVI) bands) or a below-canopy radiation sensor could help identify the abrupt change of ecosystem status, but the gap-filling method must be flexible enough to take this information as input.

An alternative method to take into consideration management and disturbances during gap filling is to change the parameterization strategy. Parameters of the gap-filling method are set using valid data measured in a time window that could have different sizes from multiple years to few days (see Sect. 6.3.2), and clearly in ecosystems where management and disturbances occurs, it is important to keep this window as short as possible to avoid the use of data acquired in periods with similar drivers values (e.g., meteorological conditions) but completely different fluxes due to the change of status in the same parameterization step. However, even a small window could include data before and after the disturbance event, in particular, when, in croplands or grassland, the EC system needs to be removed during harvesting and the gap is long and centered around the critical period.

The best way to solve this problem is probably the use of disturbance or management indicators (DI) to split the data set in subsets that do not include abrupt ecosystem status changes. In practice, registering the date and time of management practices or disturbances events that are supposed to have a direct and immediate effect on the fluxes it is possible to identify periods where the fluxes are function only of time (e.g., regrowth) and meteorological conditions. The gap-filling method can be then parameterized using only data acquired during the homogeneous subperiod (Fig. 6.1). In addition, in case of similar management across the years, for example, in grassland where generally the 3–4 cutting events per year happen in the same periods or in cropland if the same species is cultivated for different years, the gap-filling model can be parameterized using data from the same subperiod of previous years (if the others states conditions remained stable). In this way, the number of data points available increase making the parameterization more robust.

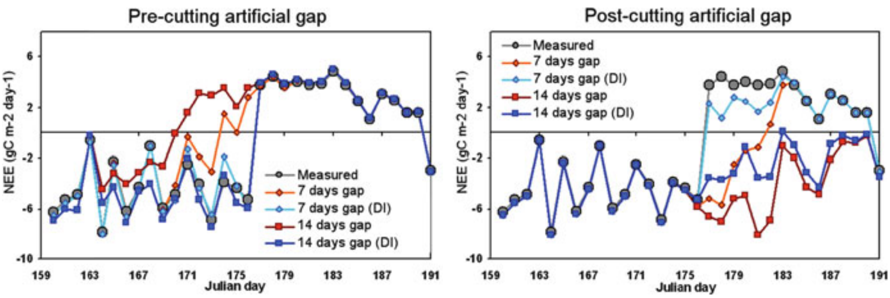

Fig. 6.1 Example of the Disturbance Indicator (DI) uses in gap filling NEE data measured in a managed grassland. Two artificial gaps of 7 days and 14 days have been added before (*left*) and after (*right*) the cutting dates that create a discontinuity. The artificial gaps have been filled using the MDS method (Sect. 6.3.3.2). It is possible to see how the gap-filling method performances improve when the DI is used (These data have been gently processed and provided by Arnaud Carrara)

6.3.3 Methods Description

6.3.3.1 Mean Diurnal Variation

The Mean Diurnal Variation method (MDV) is an interpolation technique that is based on the temporal auto-correlation of the fluxes (Falge et al. 2001). Missing observation is replaced by the mean of valid values measured on adjacent days at the same time (the same half-hour or with a buffer of +/− 1 h). The length and definition of the averaging period (window) can vary between different method implementations. In general, a window length not larger than 2 weeks is recommended since for longer periods nonlinear dependence on environmental variables could introduce large uncertainty and errors (Falge et al. 2001). Also the position of the window could be fixed or variable where in the first case the windows are predecided and fixed and all gaps occurring in each of the windows are replaced applying the MDV in the same period, while in the second case the windows are defined around each single gap. Clearly, the second method is preferable because the gap will always be centered in the window.

The MDV method does not require drivers and it is the only method applicable when all the meteorological data are missing; it is an empirical method and can be in theory applied to fill all the variables when temporal auto-correlation is expected. The method implementation is easy but the accuracy and performances are lower with respect to the others methods (see Sect. 6.4).

6.3.3.2 Look-Up Tables

The Look-up table (LUT) is an empirical method, easy to implement, where the missing values are replaced with the average of valid measurements occurring under similar meteorological conditions. In practice, a multidimensional table is created where the missing value can be "looked up" based on the values of the meteorological drivers.

For example, in Falge et al. (2001) four tables were created in the year (according to the different seasons) and the drivers used were photosynthetic photon flux density (PPFD) and air temperature. The valid NEE data were binned and averaged according with the drivers values (23 PPFD classes of 100 μmol m^{-2} s^{-1} and 35 Air Temperature classes of 2°C for a total of $35 \times 23 = 805$ classes per period) and each missing data point replaced with the NEE value in the table in same drivers combination class occurring during the gap. Gaps in the table, where no valid NEE data were present for a given combination of the two drivers, were filled with linear interpolation.

The drivers used in the table preparation should be selected according to the site characteristics, taking into consideration the environmental variables that are more important in the processes of interest (i.e., the flux to be gap-filled) without selecting too many variables that would lead to the impossibility of finding a sufficient number of valid data to calculate robust averages for each of the driver class combinations. In general, 3–4 variables selected among incoming and diffuse radiation, air and soil temperature, soil water content, and vapor pressure deficit are sufficient.

Also, the number of tables created in each year is an important aspect to consider. Monthly or biweekly LUTs are possible if the amount of valid data is sufficient, and in these cases the number of drivers classes can be smaller. In addition, the drivers considered could change according to the period of the phenological cycle and the daily course, for example, not using incoming radiation as a driver during nighttime.

Reichstein et al. (2005) proposed a method (Marginal Distribution Sampling – MDS) where they consider both the covariation of fluxes with meteorological variables and the temporal auto-correlation of the fluxes. In their approach, similar meteorological conditions are sampled in the temporal vicinity of the gap to be filled looking in a window around the gap as small as possible to include a sufficient number of valid data with similar meteorological conditions to calculate the average flux. In their method, the drivers used to evaluate the similarity in the meteorological conditions vary in order to find a compromise between number of drivers and window length. Incoming radiation, air temperature, and vapor pressure deficit are first considered; then, if the window needed exceeds predefined maximum length, only incoming radiation is considered, and finally the MDV method (described in Sect. 6.3.3.1). For a given gap, in the impossibility to find a sufficient number of valid data to calculate the average for certain window length and drivers set, the next step could be to increase the size of the window or to reduce the numbers of drivers considered; the strategy to decide which of the two options to follow in the different conditions is well explained in their paper and can be used as example. The MDS method has been implemented in the European database as one of the standard gap-filling methods available in the central processing (see Chap. 17).

6.3.3.3 Artificial Neural Networks

The Artificial Neural Networks (ANNs) are purely empirical nonlinear regression models with a medium level of implementation difficulties. The ANN consists in a set of nodes, often organized in layers and connected by weights that are equivalent to the regression parameters (Bishop 1995; Rojas 1996). The first step to use an ANN is the network parameterization process called "training." The ANN is trained by presenting it with sets of input data (drivers) and associated output data that, in the case of a gap-filling application, are valid fluxes. Once the ANN is trained, the underlying dependencies of the output on the driver variables are mapped onto the weights and the ANN can be then used to predict the missing values.

There are different algorithms to train the ANNs and one of the most used is the back-propagation algorithm, where the training of the ANN is performed by propagating the input data through the nodes via the weighted connections and then back-propagating the error calculated as difference between the predicted and real output and adjusting the weights to minimize this error (Papale and Valentini 2003; Braswell et al. 2005).

Similar to the LUT method, also in the ANN, it is important to select as input the appropriate and relevant environmental variables that drive the flux variability. These could be a large set (e.g., all the meteorological variables measured at the site) or just a preselected subset. In the first case, the ANN has the possibility to use (i.e., assign high weights) variables commonly not considered as drivers that would be probably excluded in the second case, but it is also important to keep in mind that increasing the number of input variables leads to an increase of the degree of freedom (number of weights) and requires the use of a larger training data set to avoid model over-fitting and consequent loss of generalization ability.

The quality and the representativeness of the training data set play also an important role. The ANNs, as all the purely empirical models, can only map and extract information present in the data set used in the parameterization; for this reason the data set must be accurate and cover as much and as homogenously as possible the different ecosystem conditions (e.g., seasons, phenological phases, daily courses). Presampling of the training data set to ensure an equal coverage of the different conditions and the use of fuzzy values to represent additional information such as time have been tested and used showing good results (Papale and Valentini 2003; Moffat et al. 2007). Also, the training of different ANNs for daytime and nighttime (using different drivers) or the training of different ANNs for different periods and using data from adjacent years, as explained in Sect. 6.3.2.1, can improve the method performances.

The performances of the ANN method in carbon flux gap-filling are good (see Sect. 6.4) and for this reason this technique is used as standard in the European database and in FLUXNET, together with the MDS method explained above. The ANNs require a gap-free driver dataset and for this reason it is needed to first gap-fill the meteorological variables (Sect. 6.3.1) or, when this is impossible to implement, a second method (e.g., the MDV, Sect. 6.3.3.1) to be used when one or more drivers are missing.

6.3.3.4 Nonlinear Regressions

The nonlinear regressions method is based on parameterized non-linear equations which express semi-empirical relationships between the flux and environmental variables, often temperature and light for CO_2 fluxes. There are different versions and implementations that have been proposed (Falge et al. 2001; Hollinger et al. 2004; Barr et al. 2004; Desai et al. 2005; Richardson et al. 2006; Noormets et al. 2007) but in general two different equations are used, one for nighttime data often estimated as function of temperature and one for daytime data using a light response function.

The response of fluxes to the photosynthetic photon flux density PPFD is commonly modeled using the rectangular hyperbola function like the *Michaelis and Menten* equation (Eq. 9.6, Sect. 9.3.3.4) or an exponential function like *Mitscherlich* equation (Eq. 9.8, Sect. 9.3.3.4) (Falge et al. 2001). For nighttime data the most used functions are the *Lloyd and Taylor* and the *Arrhenius* (Eq. 9.5, Sect. 9.3.2.2) (Lloyd and Taylor 1994; Falge et al. 2001; Moffat et al. 2007). Both the equations have temperature as driver and can use either air or soil temperatures.

The parameters estimation for all these functions is done using measured valid data. Also in this case it is important to carefully check the data before and use only accurate measurements. In addition, the regression parameters can be kept constant only for a certain period of time to accommodate the variation over the year of all the other drivers not considered in the equations (i.e., season, water availability etc.).

The method is semi-empirical because, although the parameters are estimated using the measurements, the shape of functions between drivers and fluxes are imposed. This is an important aspect to consider when the gap-filled data are needed in modeling activities because the model to validate or parameterize could have the same function, linking for example, temperature and respiration and leading to the risk of spurious correlations (Sect. 6.2). In addition, it is applicable only when the functions linking meteorological variables and fluxes are well known and consolidated.

6.3.3.5 Process Models

In the process models, we can include all the models that have been developed to estimate and predict fluxes, simulating all the processes occurring and using generally as input not only the meteorological variables but also state variables like soil and vegetation characteristics and others quantities like leaf area index (LAI) and biomass. These models are generally not developed specifically to be used as gap-filling techniques.

In these models, which make full use of our knowledge of the processes involved in the ecosystem functioning, the data are used to constrain some of the model parameters. The advantage is that, assuming that the processes are well represented in the model, it is possible to apply it to reconstruct also long periods of gaps or even the fluxes under different climate, for example, in different years with respect to the one with EC measurements. The disadvantages are related to the uncertainty in the reproduction processes in the model and the risk that some important process could be completely missing or not correctly reproduced.

The implementation of this method is quite complex and require knowledge of the model and parameters optimization techniques. The results can be used in the site-level analysis but not for model validation and parameterization if the two models have similar routines or functions.

6.4 Uncertainty and Quality Flags

Uncertainty estimation is an important information that should be always included when data are gap filled. There are two main different uncertainty sources in the gap-filled values: One is represented by the diverse estimations that different gap-filling methods give for the same missing data point and the other is due to the uncertainty in the selected gap-filling model parameterization that is, for example, larger when the gaps are longer.

Moffat et al. (2007) showed in a comparison of gap-filling techniques for carbon fluxes that most of the methods implemented give good results, often with errors with magnitudes similar to the noise component in the data (see also Sect. 7.3.3.3) but with a slightly higher performances for empirical methods like ANN and MDS. Based on these results, it is possible to conclude that the uncertainty related to the selection of the gap-filling method is relatively small when one of the high-performance methods is used if the gap length is not too long and if the data set available to set the gap-filling model parameters is sufficiently large and with of good quality.

The best way to assess the uncertainty due to the parameterization is linked to the method selected. The quality of the parameterization is a function of the number of data points, the data quality, and the number of the variables used to constrain the model. In general, long gaps, during which the general ecosystem conditions can change (e.g., growing season phase, ground water table, nutrients availability), are more difficult to fill and the uncertainty associated with the gap-filled values will be in general higher with respect to short gaps, with highest uncertainty values in the middle of the gap due to the distance (in time and for this reason also in terms of ecological conditions) from the measured data used to estimate the parameters values.

It is important to assess an uncertainty or a “confidence level” to associate with each gap-filled value; this information is essential for a correct data analysis and interpretation. For some of the methods introduced in this chapter, an estimate of the uncertainty level is relatively simple. In the LUT method (Sect. 6.3.3.2) for example, the standard deviation of the flux values in the same drivers class gives an indication of the variability inside a group of data that the method assumes to be similar. The same is valid for the MDV method (Sect. 6.3.3.1) where the standard deviation or percentile distribution of the measurements at the same time in adjacent days gives information about the uncertainty in the gap-filled values. In others cases, like when ANN or NRL are used (Sects. 6.3.3.3 and 6.3.3.4), the uncertainty can be estimated using subsets of the available data to parameterize different versions of the same model that then can be all applied obtaining different values for the same gap.

Independently of the estimation of an uncertainty value to associate to each gap-filled value, it is important to create additional information about the method applied that can be included in the data set. This information can include the distance of each single half-hourly missing from the first valid value, indication of the drivers used to fill the gap, length of the window needed to find sufficient data to parameterize the model, and number of the data points used. In addition, quality flags to summarize the expected quality of each gap-filled value can be defined and added; an example of these quality flags is presented in the appendix of the Reichstein et al. (2005) paper.

6.5 Final Remarks

Gap-filling is a process that is sometimes unavoidable, in particular when daily to annual integrals are needed, and different methods exist. The compherensive analysis conducted by Moffat et al. (2007) showed that all the gap-filling techniques give on average good results when the gaps are shorter than 10 days and the relevant meteorological drivers available. In addition it has been also shown that including information about discontinuity (6.3.2.1) can improve the results in case of sites with management.

The decision about which method to select should be then based on different considerations. First, the availability of drivers: if no meteorological data are available, the MDV method is often the only one available and the uncertainty associated to the simulated values will be large. Another important aspect to consider is the possibility of spurious or circular correlation between data and model results when the gap-filled measurements are used in process model validation. In these cases, it is important to use a purely empirical method.

Also the difficulties in the implementation could preclude the use of some of the methods. In these cases, however, the use of centralized gap-filling services often provided by the databases could help to use the best methods without the need to implement them locally.

Finally, due to the strong link between the gap-filling quality and uncertainty (always important to estimate), the availability of meteorological data, and information about management and disturbance events, it is fundamental to carefully register all the ancillary data about the site and to install a back-up meteorological station close to the EC tower and independent in terms of energy.

References

Barr AG, Black TA, Hogg EH, Kljun N, Morgenstern K, Nesic Z (2004) Inter-annual variability in the leaf area index of a boreal Aspen-Hazelnut forest in relation to net ecosystem production. Agric For Meteorol 126:237–255

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Braswell BH, Sacks B, Linder E, Schimel DS (2005) Estimating ecosystem process parameters by assimilation of eddy flux observations of NEE. Glob Change Biol 11:335–355

Desai AR, Bolstad P, Cook BD, Davis KJ, Carey EV (2005) Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA. Agric For Meteorol 128(1–2):33–55

Falge E, Baldocchi D, Olson RJ, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grünwald T, Hollinger D, Jensen N-O, Katul G, Keronen P, Kowalski A, Ta Lai C, Law BE, Meyers T, Moncrieff J, Moors E, Munger JW, Pilegaard K, Rannik U, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, Wofsy S (2001) Gap filling strategies for defensible annual sums of net ecosystem exchange. J Agric For Meteorol 107:43–69

Gove JH, Hollinger DY (2006) Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. J Geophys Res 111:D08S07. doi:10.1029/2005JD006021

Hammerle A, Haslwanter A, Tappeiner U et al (2008) Leaf area controls on energy partitioning of a temperate mountain grassland. Biogeosciences 5(421):431

Hollinger DY, Aber J, Dail B, Davidson EA, Goltz SM, Hughes H, Leclerc M, Lee JT, Richardson AD, Rodrigues C, Scott NA, Varier D, Walsh J (2004) Spatial and temporal variability in forest-atmosphere CO2 exchange. Glob Change Biol 10:1689–1706

Hui D, Wan S, Su B, Katul G, Monson R, Luo Y (2004) Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. Agric For Meteorol 121:93–111

Lloyd J, Taylor JA (1994) On the temperature dependence of soil respiration. Funct Ecol 8: 315–323

Moffat AM, Papale D, Reichstein M, Hollinger DY, Richardson AD, Barr AG, Beckstein C, Braswell BH, Churkina G, Desai AR, Falge E, Gove JH, Heimann M, Hui D, Jarvis AJ, Kattge J, Noormets A, Stauch VJ (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric For Meteorol 147:209–232. doi:10.1016/j.agrformet.2007.08.011, ISSN: 0168–1923

Noormets A, Chen J, Crow TR (2007) Age-dependent changes in ecosystem carbon fluxes in managed forests in northern Wisconsin, USA. Ecosystems 10:187–203

Papale D, Valentini R (2003) A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Glob Change Biol 9:525–535

Papale D, Reichstein M, Aubinet M, Canfora E, Bernhofer C, Longdoz B, Kutsch W, Rambal S, Valentini R, Vesala T, Yakir D (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences 3:571–583

Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, Bernhofer C, Buchmann N, Gilmanov T, Granier A, Grünwald T, Havrankova K, Ilvesniemi H, Janous D, Knohl A, Laurila T, Lohila A, Loustau D, Matteucci G, Meyers T, Miglietta F, Ourcival JM, Pumpanen J, Rambal S, Rotenberg E, Sanz M, Tenhunen J, Seufert G, Vaccari F, Vesala T, Yakir D, Valentini R (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob Change Biol 11:1424–1439

Richardson AD, Braswell BH, Hollinger DY, Burman P, Davidson EA, Evans RS, Flanagan LB, Munger JW, Savage K, Urbanski SP, Wofsy SC (2006) Comparing simple respiration models for eddy flux and dynamic chamber data. Agric For Meteorol 141:219–234

Rojas R (1996) Neural networks. Springer, Berlin

Wohlfahrt G, Hammerle A, Haslwanter A et al (2008) Seasonal and inter-annual variability of the net ecosystem CO2 exchange of a temperate mountain grassland: effects of weather and management. J Geophys Res 113:D08110. doi:10.1029/2007JD009286