



Upscaling of evapotranspiration fluxes from instantaneous to daytime scales for thermal remote sensing applications

C. Cammalleri^{1,*}, M. C. Anderson¹, and W. P. Kustas¹

¹US Department of Agriculture, Agricultural Research Service, Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA

* now at: European Commission, Joint Research Centre, 21027 Ispra (VA), Italy

Correspondence to: C. Cammalleri (cammillino@gmail.com)

Received: 6 May 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 7 June 2013

Revised: 21 March 2014 – Accepted: 2 April 2014 – Published: 22 May 2014

Abstract. Four upscaling methods for estimating daytime actual evapotranspiration (ET) from single time-of-day snapshots, as commonly retrieved using remote sensing, were compared. These methods assume self-preservation of the ratio between ET and a given reference variable over the daytime hours. The analysis was performed using eddy covariance data collected at 12 AmeriFlux towers, sampling a fairly wide range in climatic and land cover conditions. The choice of energy budget closure method significantly impacted performance using different scaling methodologies. Therefore, a statistical evaluation approach was adopted to better account for the inherent uncertainty in ET fluxes using eddy covariance technique. Overall, this approach suggested that at-surface solar radiation was the most robust reference variable amongst those tested, due to high accuracy of upscaled fluxes and absence of systematic biases. Top-of-atmosphere irradiance was also tested and proved to be reliable under near clear-sky conditions, but tended to overestimate the observed daytime ET during cloudy days. Use of reference ET as a scaling flux yielded higher bias than the solar radiation method, although resulting errors showed similar lack of seasonal dependence. Finally, the commonly used evaporative fraction method yielded satisfactory results only in summer months, July and August, and tended to underestimate the observations in the fall/winter seasons from November to January at the flux sites studied. In general, the proposed methodology clearly showed the added value of an intercomparison of different upscaling methods under scenarios that account for the uncertainty in eddy covariance flux measurements due to closure errors.

1 Introduction

Routine monitoring of actual evapotranspiration (ET) is widely seen as a key scientific issue benefiting practical applications in a variety of fields, including water management, water rights regulation, crop water use efficiency assessment and drought monitoring (e.g., Allen et al., 2005; Anderson et al., 2011, 2012; Mu et al., 2013). These applications usually require time-integrated ET from daily to monthly and seasonal scales. Thermal remote-sensing-based methods are often used to characterize the spatial variability of this component of the hydrological balance over the landscape at various spatial scales (Kalma et al., 2008); however, the applicability of these models is controlled by the availability of cloud-free land surface temperature (LST) acquisitions. Clear-sky LST maps are usually retrieved at a specific time of day, depending on satellite orbit configuration. As an example, the overpass time of the Landsat series, in sun-synchronous polar orbit, is around 1000 local solar time, while MODIS sensors on board the Terra and Aqua platforms have an equator crossing time of 1030 and 1330, respectively. Remote ET estimates acquired with these instruments, as a single snapshot during the day, have to be upscaled to longer timescales (i.e., daily total ET) in order to become useful for hydrologists and water managers.

Temporal upscaling is commonly performed by assuming conservation of some ET metrics over the course of the day, generally expressed as a ratio between instantaneous ET at a specific time of day and a reference variable that can be computed hourly. This hypothesis is generally known as self-preservation (Crago, 1996). Several studies have analyzed

the reliability of this hypothesis, especially when the available energy (the difference between net radiation, R_n , and soil heat flux, G_0) is assumed as the reference variable (e.g., Brutsaert and Chen, 1996; Lhomme and Elguero, 1999). Brutsaert and Sugita (1992) demonstrated that this ratio, commonly referred to as the evaporative fraction (EF), is relatively constant during the central daytime hours for days with clear skies. However, Gentine et al. (2007) observed a sensitivity of self-preservation to soil moisture and canopy coverage, and Crago and Brutsaert (1996) have shown that EF is significantly higher during early morning and late afternoon, causing a systematic underestimation of daytime average values by the midday values. Some studies have introduced a correction multiplicative factor of 1.1 to compensate for this well-known systematic error (e.g., Anderson et al., 1997). Other authors (e.g., Hoedjes et al., 2008; Delogu et al., 2012) have proposed and tested correction procedures to account for EF diurnal variations using hourly ancillary meteorological data.

Another common assumption in flux upscaling procedures is that clear-sky conditions persist throughout the day (see Brutsaert and Sugita, 1992; Delogu et al., 2012). However, as pointed out by Van Niel et al. (2012), the assumption of clear-sky conditions during the whole day is not always assured for remote sensing applications, for which only the specific time-of-day of the satellite overpass must be clear. Hence, analyses that do not assess errors associated with all-sky (clear and cloudy) conditions have limited applicability for operational remote sensing applications.

Other commonly employed upscaling methods use the incoming solar radiation, R_s (Jackson et al., 1983; Zhang and Lemeur, 1995), or even the top-of-atmosphere (or clear-sky) irradiance, R_{TOA} (Ryu et al., 2012), as reference variables. Both methods have demonstrated value for upscaling specific time-of-day ET estimates to daily, 8-day, and monthly scales (Ryu et al., 2012; Van Niel et al., 2012). In addition, specifically for applications over agricultural areas, Trezza (2002) introduced the use of standardized reference evapotranspiration (ET_0) as an upscaling variable, based on the assumption that ET_0 incorporates most of the main meteorological factors that influence the evaporative process. A variation of this method was tested by Delogu et al. (2012), using crop-specific potential evapotranspiration instead of reference ET. Over 11 Mediterranean sites they found no substantial difference in using this method instead of EF, although more auxiliary information is needed.

In many cases, previous analyses of upscaling methods reported in the literature have focused only on a few experimental sites and/or short time periods, and many were based on assumptions that may not hold in all cases (i.e., all-sky conditions vs. only clear-sky days, assumption of energy balance closure in ET observations). A substantial intrinsic limitation in such analyses has been the absence of unanimous consensus regarding the definition of time-integrated “daily” variables – the nominal representation of “truth” – particu-

larly when the eddy covariance technique is used to collect in situ fluxes. Eddy covariance (EC) measurements are known to be less reliable during nighttime hours when turbulence is weak (Falge et al., 2001; Fisher et al., 2007), and a question remains regarding proper treatment of the surface energy imbalance inherent in most EC measurement sets (Wilson et al., 2002). Some authors (e.g., Twine et al., 2000) suggested various methods to force energy budget closure by altering the observed latent and/or sensible heat fluxes, while others (e.g., Leuning et al., 2012) assert that it is possible to obtain the correct balance at the half-hourly scale by careful attention to the different sources of error. These uncertainties have resulted in a diversity of definitions of “daily” ET that can differ between studies, and can lead to different conclusions about an optimal upscaling approach. The quotation marks around “daily” refer to this ambiguity in absolute definition of daily ET.

Finally, in the context of remote sensing applications, requirements regarding auxiliary information may further limit operational utility of different upscaling methodologies. The methods proposed in the literature represent a range in requirements; from extensive modeling of available energy (which requires several surface-related variables that are not easily retrievable), to the almost null requirements of methods based on top-of-atmosphere (or clear-sky) irradiance. Even correction procedures proposed for accounting for well-known limitations in some of those methods require different levels of inputs and are variably complex and site specific. All of these constraints must be considered in selecting an optimal upscaling procedure for different regions and applications, and it must be remembered that accuracies obtained at local scales may not be indicative of results using only remote sensing and regionalized data.

In this paper, we evaluate some common upscaling methods using an approach that attempts to account for the uncertainty in surface energy balance closure, and considers the typical operational constraints of thermal remote-sensing-based applications. With this aim, an intercomparison of 4 different upscaling methods is conducted using surface energy fluxes collected by 12 stations from the AmeriFlux network (www.ameriflux.ornl.gov). The in situ flux observations are used to represent both the instantaneous specific time-of-day retrieval (i.e., assuming a perfect satellite retrieval model) and the “daily” upscaled ET. This is done in order to isolate the uncertainty of upscaling method from ET model-specific uncertainties. All-sky diurnal conditions are simulated, with the only constraint of clear skies at the sensor overpass time. The study evaluates upscaling error as a function of scaling flux, month of year, and time of satellite overpass (between 0900 and 1500 solar time).

2 Materials

The daytime total actual evapotranspiration (ET_d , from sunrise to sunset), upscaled using a generic reference variable X , can be computed using the following relationship:

$$ET_{d-X} = \beta \frac{1}{\lambda} \frac{\lambda ET_t}{X_t} X_d, \quad (1)$$

where λET_t is the instantaneous latent heat flux at the time-of-day t , λ is the latent heat of vaporization, X_t and X_d are the values of the reference variable at the “acquisition” time t and the daytime total, respectively, and β is a correction factor to account for potential systematic biases in the upscaling method (Van Niel et al., 2011). In this paper we use units of mm d^{-1} for daytime (subscript d) variables and W m^{-2} for “instantaneous” (subscript t) fluxes.

Four upscaling methods were tested: (1) the evaporative fraction (EF) method, where the reference variable is the available energy, $X = (R_n - G_0)$; (2) the solar radiation method (RS), where the reference variable is the incoming shortwave radiation at the land surface ($X = R_s$); (3) the top-of-atmosphere irradiance method (TOA, $X = R_{TOA}$); and (4) the reference evapotranspiration method (REF), where the reference variable is the standard crop reference evapotranspiration ($X = ET_o$), computed following the FAO-56 paper (Allen et al., 1998). To compensate for systematically high values of EF observed during early morning and late afternoon, β is generally assumed equal to 1.1 for the EF method (Anderson et al., 1997); the effects of this assumption are discussed. Since the literature has little information pertaining to systematic errors in the RS, TOA and REF methods, especially in the case of daytime fluxes, β is assumed equal to 1 for all these cases. Two notable exceptions are the analyses conducted by Van Niel et al. (2012) and Delogu et al. (2012), where correction factors for the retrieval of 24 h ET were proposed as a function of day of the year, time-of-day, cloud conditions and ancillary data.

Two variations on the EF method have also been tested, using reference variables that are simplified representations of the total available energy. The first neglects daytime time-integrated G_0 in the computation of EF ($X_d = R_{n,d}$ and $X_t = R_{n,t} - G_{0,t}$), using only daily net radiation as the scaling flux (referred to as the RN method). The second further abstracts EF using only the net shortwave component of R_n (RSW method). Analysis of these simplified methods allows for the assessment of values conveyed by the G_0 and net longwave components of EF – components that can be more difficult to retrieve accurately in remote sensing applications.

The data set used in this study includes half-hourly observations of surface energy fluxes collected at 12 AmeriFlux stations. These sites were selected in order to cover a wide range of both plant functional types and meteorological conditions (Table 1). Data recorded in 2 different years were used for each site, selected to minimize data gaps while providing significant variation in water stress conditions. Turbu-

lent fluxes of sensible (H) and latent heat were obtained from the Level 2 standardized AmeriFlux data set and observed G_0 values were corrected for heat storage using soil temperature and moisture measurements above the plates (Fuchs and Tanner, 1968). Data gaps were not filled, and only days with fully available half-hourly daytime data were used in the analysis.

Given the surface energy imbalance typical of EC data, three different “daily” ET data sets were used in the following analyses: (i) the “Unclosed” data set, where closure was not enforced; (ii) the “Residual” data set, where λET is obtained as a residual term of the surface energy budget ($\lambda ET = R_n - G_0 - H$); and (iii) the “Bowen” data set, where surface energy balance was forced by preserving the observed Bowen ratio $H / \lambda ET$ (Twine et al., 2000).

Daytime ET was derived as a sum of half-hourly latent heat flux data collected between local sunrise and sunset, computed separately for the three “daily” ET data sets. The choice of focusing on daytime fluxes instead of 24 h fluxes was motivated by the poor reliability of nighttime EC observations (Falge et al., 2001). Half-hourly λET were used as input to Eq. (1), while daytime-integrated ET fluxes were adopted as validation quantities. The observed reference variables, X in Eq. (1), at both half-hourly and daytime scales were used as proxies for remote estimates. The reliability of this hypothesis is discussed successively. In the case of the REF methodology, half-hourly ET_o was modeled using local meteorological data collected at flux stations (Allen et al., 1998), and daytime values were obtained through a simple summation.

3 Analysis methods

3.1 The effects of energy budget closure technique

As mentioned above, choice of closure correction technique applied to EC data used to evaluate flux upscaling approaches can significantly impact the conclusions, potentially resulting in different rankings in accuracy. To demonstrate this, the four upscaling methods were applied to the sites of 12 “daily” ET data sets as adjusted with each of the three closure methods (Bowen, Residual and Unclosed). The methods were applied over the whole year, but only on days that were predominantly clear at the nominal “acquisition” time t , as assessed by the threshold $R_s / R_{TOA} > 0.70$ (roughly corresponding to 90 % of clear-sky irradiance). This value ensures inclusion of data that were not significantly contaminated by clouds, while assuming that a slight reduction of incoming irradiance does not affect remote sensing acquisitions. Seven different possible satellite overpass times-of-day were considered (from 0900 to 1500 solar time, at 1 h time steps) to test dependence of upscaling errors on time of clear-sky acquisition. In this demonstration, method performance was assessed using metrics of (i) relative error, RE (%), computed as the ratio between mean absolute error (MAE; $E|ET_{d-X} -$

Table 1. Study site information. The ID represents the standard AmeriFlux identification code, and the column Years reports the 2 observation years analyzed for each site.

ID	Location	Lat	Lon	Plant type	Climate	Years
Ne-1	Mead, NE	41.2	-96.5	Cropland	Temperate	2003, 2005
Bo-1	Bondville, IL	40.0	-88.3	Cropland	Temperate	2004, 2007
Br1	Brooks Field, IA	41.7	-93.7	Cropland	Temperate	2006, 2008
IB-1	Fermi Agricultural, IL	41.8	-88.2	Cropland	Temperate	2006, 2008
ChR	Chestnut Ridge, TN	35.9	-84.3	Deciduous broadleaf forest	Temperate	2006, 2007
MOz	Missouri Ozark, MO	38.7	-92.2	Deciduous broadleaf forest	Temperate	2006, 2007
Me-2	Metolious, OR	44.5	-121.6	Evergreen needleleaf forest	Sub-tropical	2004, 2005
Aud	Audubon, AZ	31.6	-110.5	Grassland	Semi-arid	2004, 2008
Bkg	Brookings, SD	44.3	-96.8	Grassland	Temperate	2006, 2009
Var	Vaira, CA	38.4	-121.0	Grassland	Mediterranean	2002, 2007
FPe	Fort Peck, MT	48.3	-105.1	Grassland	Temperate	2004, 2006
Ton	Tonzi, CA	38.4	-121.0	Woody Savanna	Mediterranean	2002, 2007
SRM	S. Rita Mesquite, AZ	31.8	-110.9	Woody Savanna	Semi-arid	2004, 2008
FR-2	Freeman, TX	29.9	-98.0	Woody Savanna	Semi-arid	2005, 2006

ET_d) and the observed average daytime ET; and (ii) relative bias, RB (%) given by the ratio of the mean bias error (MBE; $E[ET_{d-X} - ET_d]$) and the observed average daytime ET.

The plots in Fig. 1 summarize the sensitivity of the RE and RB performance metrics to choice of closure technique as a function of assumed overpass time (allowed to vary over the midday period from 0900 to 1500). Figure 1a–c shows RE, averaged over all 12 sites, for the Bowen, Residual and Unclosed data sets, respectively, while Fig. 1d–f reports the corresponding average RB values. The strong dependence of apparent method performance on closure technique makes it difficult to conclude anything definitive about the overall accuracy of each method, as well as on the relative performance of one method with respect to the others. Even the overall accuracy seems to vary considerably for the different “daily” cases, and similarly the diurnal shape of both RE and RB curves.

3.2 Statistical analysis approach

As demonstrated in Sect. 3.1, the relative performance of the various methods is strongly connected to the degree of closure observed at the different sites, as well as to the diurnal variability of this imbalance. Such problems cannot be ignored when evaluating methods of upscaling using EC data. For this reason, an ensemble-based intercomparison method that explicitly accounts for the uncertainty in the assessment of “daily” ET has been adopted in this study.

Combining the three “daily” ET datastreams, a minimum (ET_{\min}) and maximum (ET_{\max}) observed daytime ET value is identified for each day and site. In most cases, the ET_{\min} value is associated with the unclosed datastream (no closure correction applied), while ET_{\max} is obtained from the residual closure method, while the Bowen ratio closure typically gives an intermediate value. The “true” state is assumed to

lie between ET_{\min} and ET_{\max} . Two additional thresholds, defined as $ET_{\min} - \Delta$ and $ET_{\max} + \Delta$, with $\Delta = 0.5(ET_{\max} - ET_{\min})$, are used to define 5 general classes of accuracy: discriminating upscaled estimates that have *acceptable accuracy* (ET_{\min} to ET_{\max}), those with *moderate errors* (ET_{\min} to $ET_{\min} - \Delta$ and ET_{\max} to $ET_{\max} + \Delta$), and those with *major errors* ($< ET_{\min} - \Delta$ or $> ET_{\max} + \Delta$). These accuracy categories replace the absolute accuracy metrics used in the demonstration exercise presented in Sect. 3.1, and in most prior upscaling studies.

Next, for each method and day an ensemble of up to 21 daily ET estimates are generated assuming a range of 7 possible overpass times t (0900 to 1500) and 3 possible simulated instantaneous λET_t “retrievals”, extracted from the observation time series under the different closure scenarios. On some days, 21 values may not be available since conditions were observed to be cloudy ($R_s/R_{TOA} < 0.70$) at a given nominal acquisition time. The ensemble estimates are pooled over a given time interval (e.g., month, year, full 2-year sample) and sorted into the five accuracy categories. Finally, frequency distributions are computed to characterize the accuracy of each method (reflected by the percentage of estimates between ET_{\min} and ET_{\max}), as well as systematic positive or negative biases (values $> ET_{\max}$ or $< ET_{\min}$, respectively).

Additional analyses were performed to test variability in results with respect to both acquisition time-of-day and season. In the first case, two complementary analyses are performed: the first using the seven local times between 0900 and 1500, and the second using seven normalized times, $t^* = (t - t_s)/N$, between 0.2 to 0.8, where t_s is the local sunrise and N is the daytime length. The first case addresses impacts of satellite acquisition time from a sun-synchronous orbit, while the second case may be more robust in minimizing impacts of seasonal changes in the diurnal course of the actual surface fluxes. Analysis on seasonality was performed

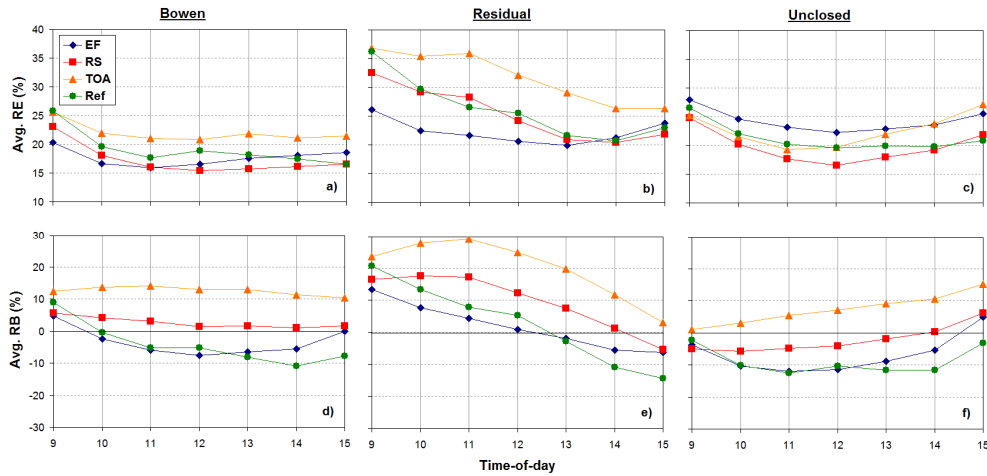


Figure 1. Statistical metrics computed using observed daytime-integrated “daily” ET (obtained using different closure constraints) and modeled values upscaled from half-hourly observations collected midday (from 0900 to 1500 solar time) time-of-day. Panels (a–c) show RE, averaged over all tower sites, for the Bowen, Residual and Unclosed data sets, respectively; panels (d–f) report the corresponding average RB values.

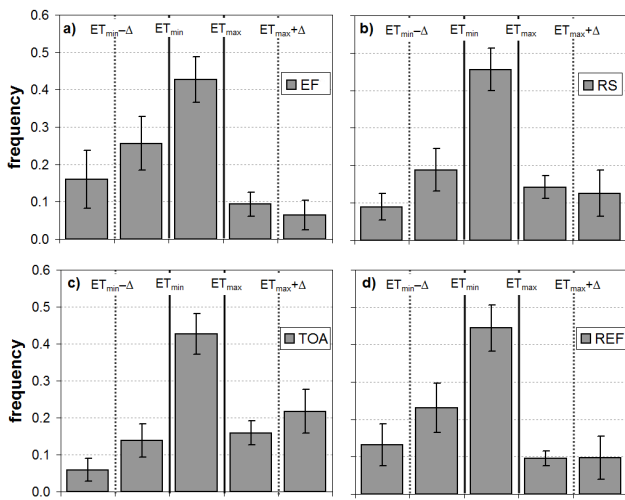


Figure 2. All-site average frequency distribution. Bars represent the average frequency of upscaled estimates from each method (combined over the 2 observation years) in the classes defined by vertical lines, while the error bars show the site-to-site standard deviation in frequency values.

by segregating the ensemble data by month, assuming that a reliable monthly frequency distribution was obtained when more than 15 days of data were available.

4 Results

The data reported in Fig. 2 summarize the results obtained following the methodology introduced in the previous section, showing the all-site average frequency values as well as the standard deviation between sites within each accu-

racy class. The histograms show that all the models have similar frequencies of *acceptable accuracy*, defined here as the percentage of upscaled values that matched the daytime-integrated “daily” values within the uncertainties of the observations (ET_{min} to ET_{max}). Of the tested methods, RS and REF were most accurate according to this criterion, yielding peak frequencies of 46 and 44 %, respectively, while EF and TOA give a somewhat lower peak value (43 %). The RS method results in a slightly lower site-to-site standard deviation in the peak frequency (5 %) compared to the other methods (6 %), potentially indicating more robust performance across varying surface and meteorological conditions. Comparing the less accurate methods, the EF method marginally outperforms TOA in terms of *moderate errors*, with a combined (positive and negative errors) frequency of 35 % vs. 30 %. The difference between positive ($> ET_{max}$) and negative ($> ET_{min}$) biases suggests that the RS method is practically unbiased (27 % for both), TOA tends to overestimate (in 38 % of the cases), while both REF and EF tend to underestimate (37 and 41 %, respectively). The use of a β correction factor (Eq. 1) in EF improves method performance, increasing accuracy from 41 % for $\beta = 1$ (not shown) to 43 % with $\beta = 1.1$ and reducing *major errors* from 24 to 22 %. Most notable is the reduction in systematic biases, where the underestimation frequency of 50 % was reduced to the above reported value of 41 %.

The relationship between satellite acquisition time-of-day and upscaling model accuracy is shown in Fig. 3. In these plots, each bar is analogous to a single plot in Fig. 2 but computed using only data collected at a specific time of day. These data show that the model accuracy (amplitude of the central black bar) varies only slightly over the daytime hours for all the models. On the other hand, only the RS method

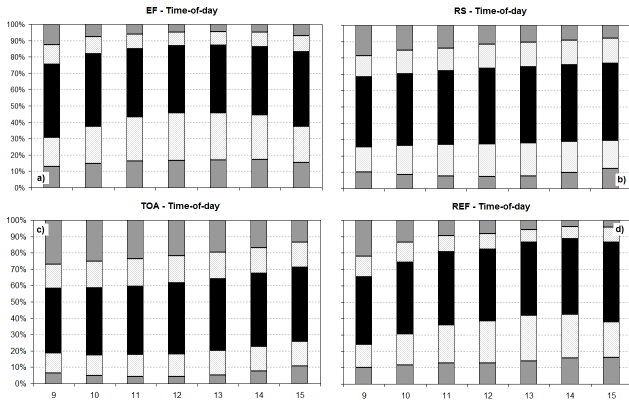


Figure 3. Variability of the accuracy of upscaling methods as a function of satellite acquisition time-of-day. The black central bar represents the frequency of data between ET_{max} and ET_{min} , the light-gray bars represent the “moderate” errors, while dark gray bars represent “major” errors. Frequencies of underestimation ($< ET_{min}$) and overestimation ($> ET_{max}$) are indicated by bars below and above the black bar, respectively. See the text for the definition of “moderate” and “major” errors.

yields relatively uniform bias for various choices of acquisition time. This characteristic benefits ET retrieval approaches that can use remotely sensed land-surface temperature data from a combination of thermal satellite sensors with varying overpass times. The EF approach shows less bias for morning acquisition times (0900 and 1000) or for late afternoon (1500), while TOA and REF show a linear trend in bias over the course of the day. TOA tends to be significantly positively biased early in the morning and almost unbiased late in the afternoon, while REF has the opposite behavior, with small bias during the morning and high negative bias during the afternoon. The accuracies obtained as a function of normalized time (scaled between sunrise and sunset) are almost identical to those reported in Fig. 3 (segregated into bins of absolute local time), with only a marginal decrease in time dependence (not shown).

To study seasonal variability in method accuracy, the plots in Fig. 4 report analyses of upscaled ET estimates segregated by month obtained with the EF (panel a), RS (b), TOA (c) and REF (d) methods. As with Fig. 3, each bar of these plots is analogous to the corresponding plot in Fig. 2, but for a specific month. The data reported in Fig. 4b demonstrate relatively small seasonal variability in the accuracy of the RS method in comparison with other upscaling techniques. The RS results are practically unbiased across the whole year, with a standard deviation (over time) in accuracy of only 3%. Similarly, the REF method (Fig. 4d) is characterized by a small seasonal variability in accuracy, although there is a systematic underestimation for all months. In contrast, the EF and TOA methods show a clear seasonality in both accuracy and biases. EF performs better during the summer months (June to August), with accuracy similar to that of RS,

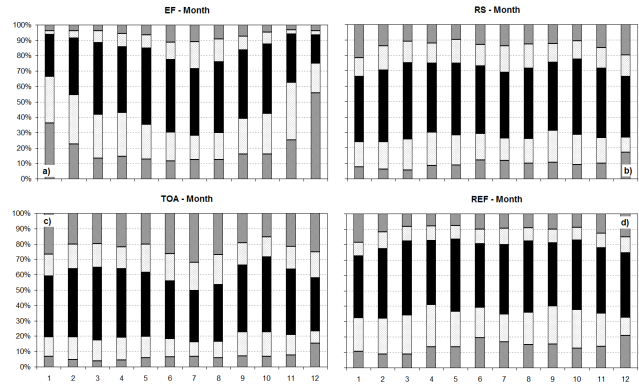


Figure 4. Monthly variability of the accuracy of upscaling methods. The black central bar represents the frequency of data between ET_{max} and ET_{min} , the light-gray bars represent the “moderate” errors, while dark gray bars represent “major” errors. Frequencies of underestimation ($< ET_{min}$) and overestimation ($> ET_{max}$) are indicated by bars below and above the black bar, respectively. See the text for the definition of “moderate” and “major” errors.

and very poorly from November to January (underestimation in up to 75% of the cases). TOA has the worst performance during July and August, when it clearly overestimates the observed daily fluxes (in about 50% of the cases). The frequency of underestimation by TOA is relatively constant over the course of the year.

The positive bias in daytime ET resulting from the TOA method can be in large part explained by the clear-sky fraction ($R_{s,d}/R_{TOA,d}$) computed for days when skies were clear at the nominal acquisition time (i.e., times/days where a clear-sky retrieval was theoretically possible). The monthly clear-sky fraction has a clear negative linear correlation with the difference between the overestimation frequency for TOA and RS methods, with a determination coefficient (R^2) of 0.74 (Fig. 5). This means that TOA methods perform similarly to RS when the sky is clear, while overestimation in TOA upscaled values increases under mixed cloud cover conditions.

The seasonal behavior of the EF method can be compared to that of the simplified EF methods (RN and RSW) (Fig. 6). Seasonal variability in accuracy from the RN method (Fig. 6a) is stronger than that obtained with EF, increasing the temporal standard deviation over time in systematic underestimations from 13 to 18% and in accuracy from 10 to 13%. On the other hand, the RSW method serves to reduce monthly variability in the accuracy to 4%, close to the value observed for the RS method (3%). However, signs of seasonality are still evident (Fig. 6b).

5 Discussion

The statistical analysis of the accuracy of the different day-time upscaling methods discussed in Sect. 4, as quantified by

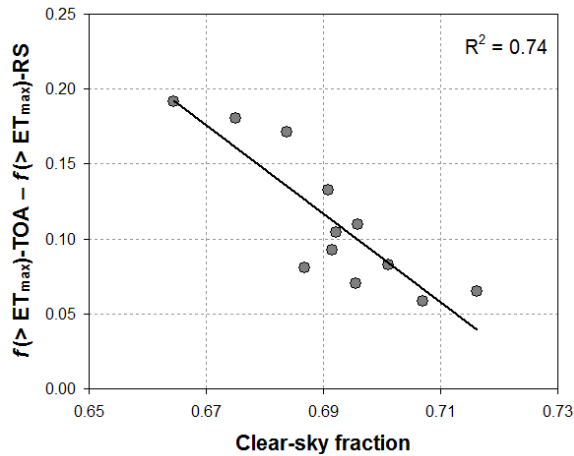


Figure 5. Correlation between all-site average monthly clear-sky fraction ($R_{s,d}/R_{TOA,d}$) and the difference between the overestimation frequency ($> ET_{max}$) for TOA and RS methods.

the frequency of retrievals falling between the minimum and maximum daytime ET values calculated from the observed flux datastreams, suggests that each method could be used with comparable results under certain conditions. While the methods yield similar levels of accuracy ($\sim 45\%$ of upscaled values falling between ET_{min} and ET_{max} in each case), the RS method demonstrates more robust overall performance both in terms of accuracy (46%) and site-to-site variability (5%). Furthermore, the analysis of systematic errors identified the RS approach as yielding the lowest bias at the monthly to annual timescale, with bias characteristics relatively uniform through the seasons. In contrast, both the EF and REF methods systematically underestimate the observed daytime fluxes, while TOA tends to systematically overestimate. These behaviors can be explained by looking in more detail at the error characteristics segregated by specific time-of-day and at the monthly scale.

The variability in the bias from the EF method with time-of-day shows a concave-down pattern (Fig. 3), with minimum bias for acquisition times early in the morning and late in the afternoon. In agreement with prior studies (e.g., Lhomme and Elguero, 1999; Gentine et al., 2007), this behavior suggests that self-preservation of EF is not achieved in general, and the systematic underestimation of the method is partially compensated by the higher EF values observed before 1000 and after 1500. To operationally use this approach, the time-dependent β correction factor suggested by van Niel et al. (2011) and Hoedjes et al. (2008) may be effective. Of the upscaling methods here tested, only the RS method is minimally affected by diurnal overpass time variability in both accuracy and bias, further confirming the robustness of this approach in its application to a variety of satellite sensors. RS also shows stable results at the monthly scale over the annual cycle, with an average temporal variability in accuracy represented by a standard deviation of 3%. The rel-

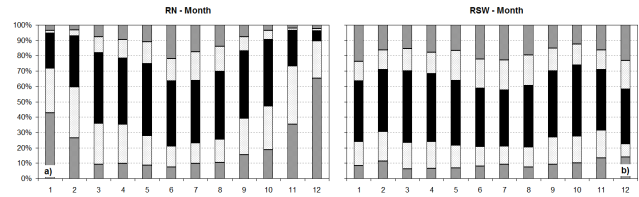


Figure 6. Monthly variability of the accuracy of EF upscaling methods using different components of available energy. RN method (a) neglects daytime G_0 , while RSW method (b) uses shortwave net radiation only. See caption of Fig. 5 for the description of the color bars.

atively high accuracy of remotely sensed R_s maps already available from geostationary satellites (Otkin et al., 2005; Cristóbal and Anderson, 2013), as well as the fact that no auxiliary information is required for the application of the RS method, provide further motivation for adoption of this technique in large-scale applications.

The seasonality in the overestimation for the TOA can be explained by cloud climatology, as evidenced by the strong correlation with clear-sky fraction in Fig. 5. The relatively high values of the monthly clear-sky fraction values obtained for all the sites, ranging only between 0.60 to 0.73, suggest that partly cloudy days (clear-sky at the specific time-of-day but cloudy on average) are just a minor fraction of the entire data set, and this is the reason why the TOA method performs reasonably well for most of the days in this and other studies (e.g., Ryu et al., 2012). However, from Eq. (1) it is clear that TOA will produce (in error) the same upscaled ET estimates regardless of whether a day is completely clear or partly cloudy, whereas RS is better able to discriminate impacts on evaporative fluxes of variable radiation load due to clouds. Some authors have suggested the use of an empirical correction coefficient for TOA estimates based on cloud conditions (Van Niel et al., 2012), and the observed negative relationship between cloudiness and TOA overestimation obtained in this study supports the reliability of this approach.

The good performance of the RS method and the small differences with TOA are consistent with the findings of Van Niel et al. (2012), who observed for two sites in Australia that RS returned the lowest error at the monthly scale compared to the EF and TOA methods. Despite this, the authors observed a systematic underestimation of measured daily ET values by RS, which may be associated to their use of 24 h integrated ET instead of daytime only as a time-integrated reference. Another source of disparity may be the use of Unclosed ET data only by Van Niel et al. (2012). The results obtained here for the TOA method do not differ significantly from those reported by Ryu et al. (2012) using 8-day average ET. The smaller bias observed by Ryu et al. (2012) may be related to the use of daytime vs. 24 h total ET.

In terms of accuracy, the EF method performed similarly to RS, especially during the June–August time frame.

However, the strong seasonality (temporal standard deviation up to 13 %) observed in EF monthly errors impacts the reliability of the model during the September–March period. Figure 6a shows that this seasonality is further increased if G_0 is neglected, while the magnitude of the bias is generally reduced from March to September due to the increased value of X_d in Eq. (1). Since accurate estimations of daytime G_0 are difficult to achieve from remote sensing data due to the effects of variation in soil thermal properties and soil moisture, this result highlights a further limiting factor in the applicability of EF method, particularly over sparsely vegetated areas where the contribution of G_0 is particularly relevant. However, it should be pointed out that the impact of G_0 may be less important if the “daily” flux was 24 h rather than daytime only (Cammalleri et al., 2012). Figure 6b demonstrates that the longwave component of R_n is the main cause of the observed seasonality in the EF method; upscaling using net shortwave radiation only (RSW) yields more uniform performance across seasons. In general, the inclusion of land-surface-related variables (i.e., G_0 , surface temperature) appears to degrade the results compared to the simple R_s .

The results suggest an imperfect conservation of EF, confirming previous observations by Gentine et al. (2007) using modeled values. The introduction of a constant correction factor $\beta = 1.1$ for EF partially reduced the systematic underestimation observed in similar recent studies by Ryu et al. (2012) and Van Niel et al. (2012), improving the performance of the method in terms of both accuracy and bias for daytime ET estimates. A value of $\beta = 1$, however, may be more reliable for 24 h ET fluxes, especially when negative nighttime fluxes are observed. As discussed by Van Niel et al. (2011), a time-dependent calibration may further improve EF performance. The results for EF obtained in this study indicate better performance than that reported by Ryu et al. (2012). This may be associated with the use of all-sky conditions by Ryu et al. (2012), including days when skies were cloudy at the specific overpass time. This assumption might cause the presence of outliers in their analysis due to non-representative “instantaneous” EF values under cloudy conditions (see Fig. 1a and b in Ryu et al., 2012).

The accuracy of the REF method (44 %) and associated systematic underestimation suggest that ET_o is not an improvement in comparison with using all-sky insolation as a scaling flux. While REF does not show seasonality in its error statistics, and its performance is more stable in time than EF or TOA, overall RS provides more robust results. A possible limitation of the REF approach, as implemented here, may be related to the differences in aerodynamic properties between the reference surface and the actual landscape around the flux measurement site. While this method has demonstrated good performance over agricultural irrigated areas (Allen et al., 2007; Trezza, 2002), application over natural semi-arid and forested sites may be less optimal. For example, Colaizzi et al. (2006) obtained very good results with the REF method for alfalfa and irrigated cotton fields in Bushland (TX) using

24 h ET, but poor results over bare soil where ET decreases rapidly for a drying soil, deviating significantly from reference ET. This may suggest limitations of the methodology in the presence of rapidly changing soil-water stress and strong surface heterogeneity. Additionally, since conditions at flux sites may be in many cases very different from reference conditions, particularly for semi-arid areas or forested sites, the accuracy of ET_o estimates computed from the local weather data will not be a true “reference ET”, potentially compromising the reliability of ET_o as upscaling quantity. The systematic positive biases in REF-upscaled ET fluxes evidenced in Figs. 3 and 4 suggest that this method could also benefit from a calibration of the β coefficient analogous to the EF method.

Overall, the use of daytime ET instead of 24 h ET as a “daily” upscaled quantity appears to reduce the systematic underestimation observed in previous studies using the RS method. Solar radiation is a good relative descriptor of daytime fluxes, but it cannot account for variability in nighttime fluxes. Implicitly assuming a constant contribution from nighttime ET may not be reasonable. Ryu et al. (2012) identified several flux sites with either high positive or negative nighttime ET fluxes depending on local climate and moisture conditions, constituting about $\pm 10\%$ of the annual sum of ET. As a consequence, the reliability in the estimation of 24 h fluxes is obviously related to the sign of nighttime fluxes, which are commonly positive in dry and advective environment (Kustas et al., 1994; Tolk et al., 2006) and negative (dew formation) in temperate climates.

6 Summary and conclusions

Four methodologies for upscaling daytime (sunrise to sunset) ET fluxes from a single time-of-day ET observation based on the self-preservation hypothesis were evaluated. The analysis was performed using flux observations collected at 12 AmeriFlux EC towers located across the US. A preliminary analysis highlighted the significant effect of surface energy imbalance and treatment thereof on upscaling method performance. Consequently, an alternative ensemble approach that intrinsically accounts for the uncertainty in EC flux tower ET observations was adopted. The results discussed here therefore better reflect the intrinsic accuracy of the different methodologies, apart from measurement and closure issues.

The results suggest that the RS method is a robust approach for daytime upscaling of ET, yielding the highest accuracy of the methods tested and an absence of systematic bias, as well as a negligible seasonality and diurnal variability. Accurate hourly insolation products, derived from geostationary satellite data, are becoming increasingly available for much of the globe and will have utility for operational applications of the RS over large scales. Continued efforts to improve satellite-based insolation data sets through calibration with local measurements (e.g., Journée and Bertrand,

2010) will further support future uses of the RS method. Analyses of the accuracy of remote-sensing-derived insolation products may be useful to evaluate the effective applicability of RS methods and the transferability of the results obtained using local data.

The TOA method appears to be less accurate than RS, and yields a systematic overestimation of daytime fluxes related to cloud coverage. While the correction of TOA estimations using coefficients based on cloud cover fraction may improve the results, the need for empirical calibration makes this technique less appealing and straightforward for routine applications. The TOA model seems to perform better for afternoon clear-sky acquisitions, and may be better suited for applications with sensors such as MODIS-Aqua. For operational purposes, it may be appropriate to use the TOA method along with the RS approach to fill spatial and/or temporal gaps where accurate solar radiation data are not available. This solution may be appealing due to the minimal requirement of information for the assessment of R_{TOA} maps globally.

The REF technique returns consistent estimates in terms of accuracy, but with a stable negative bias. For early morning (1000 local time) acquisitions, the model results are practically unbiased, suggesting that reliable estimates can be obtained using MODIS-Terra or Landsat data. However, given that ET_o estimates require insolation data, as well as other meteorological variables, it may be difficult to justify the use of this variable instead of R_s as reference for upscaling in generalized and routine applications. The results of this study suggest that systematic biases in the REF method may be further reduced by introduction of a correction factor of value $\beta > 1$. The data seem to suggest that a value of β close to 1.2 is an optimal solution; however, the use of this value should be supported by a more extensive analysis of the reliability of ET_o data over non-standard surfaces.

The accuracy of the EF method similar to that of the other methods (43 %), but the systematic underestimation and the seasonality in the errors can significantly limit its applicability, especially during winter months (November to January). The good performance obtained during June–August supports use of EF for agricultural application during the common growing season. The observed diurnal variability in the biases confirms the possibility of improving the model performance by means of daytime-variable correction factor, as suggested by van Niel et al. (2011, 2012). However, the current accuracy of remote-sensing-based estimations of daytime available energy is a limiting factor for the use of EF method operationally, and further studies are required to improve daytime net longwave radiation and soil heat flux estimates. Since the analysis was performed using locally observed daytime R_s and $(R_n - G_0)$ values, it is likely that in practical applications the RS method would in general perform better than EF in a variety of conditions.

The results reported in this study were obtained for daytime ET only; the inclusion of nighttime fluxes into the analysis could marginally alter these conclusions. However, day-

time fluxes commonly account for the majority of ET (about 90 %), suggesting that the adopted reference ET data set represents a robust target at daytime temporal scale and a good proxy for 24 h fluxes. Future studies on this topic should be focused on alternative methods (i.e., lysimeter) to provide reliable nighttime observations to be used as reference “daily” data sets.

Acknowledgement. The authors want to thank all the scientists and the supporting staffs at the AmeriFlux sites used in this study for data collection and access. The US Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual’s income is derived from any public assistance program (not all prohibited bases apply to all programs). Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotope, etc.) should contact USDA’s INTEGRATED Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

Edited by: N. Romano

References

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration, Guideline for computing crop water requirements, FAO irrigation and drainage paper n. 56, 326 pp., Rome, Italy, 1998.
- Allen, R. G., Tasumi, M., Morse, A., and Trezza, R.: A Landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning, *Irrig. Drain. Syst.*, 19, 251–268, 2005.
- Allen, R. G., Tasumi, M., and Trezza, R.: Satellite-based Energy balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)-model, *J. Irrig. Drain. Eng.*, 133, 380–394, 2007.
- Anderson, M. C., Norman, J. M., Diak, G. R., Kustas, W. P., and Mecikalski, J. R.: A two source time-integrated model for estimating surface fluxes using thermal infrared remote sensing, *Remote Sens. Environ.*, 60, 195–216, 1997.
- Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., González-Dugo, M. P., Cammalleri, C., d’Urso, G., Pimstein, A., and Gao, F.: Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery, *Hydrol. Earth Syst. Sci.*, 15, 223–239, doi:10.5194/hess-15-223-2011, 2011.
- Anderson, M. C., Allen, R. G., Morse, A., and Kustas, W. P.: Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources, *Remote Sens. Environ.*, 122, 50–65, 2012.

- Brutsaert, W. and Chen, D.: Diurnal Variation of Surface Fluxes During Thorough Drying (or Severe Drought) of Natural Prairie, *Water Resour. Res.*, 32, 2013–2019, 1996.
- Brutsaert, W. and Sugita, M.: Application of self-preservation in the diurnal evolution of the surface energy budget to determine daily evaporation, *J. Geophys. Res.*, 97, 18377–18382, 1992.
- Cammalleri, C., Ciraolo, G., La Loggia, G., and Maltese, A.: Daily evapotranspiration assessment by means of residual surface energy balance modeling: A critical analysis under a wide range of water availability, *J. Hydrol.*, 452–453, 119–129, 2012.
- Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolck, J. A.: Comparison of five models to scale daily evapotranspiration from one-time-of-day measurements, *Trans. ASABE*, 49, 1409–1417, 2006.
- Crago, R.: Conservation and variability of the evaporative fraction during the daytime, *J. Hydrol.*, 180, 173–194, 1996.
- Crago, R. and Brutsaert, W.: Daytime evaporation and the self-preservation of the evaporative fraction and the Bowen ratio, *J. Hydrol.*, 178, 241–255, 1996.
- Cristóbal, J. and Anderson, M. C.: Validation of a Meteosat Second Generation solar radiation dataset over the northeastern Iberian Peninsula, *Hydrol. Earth Syst. Sci.*, 17, 163–175, doi:10.5194/hess-17-163-2013, 2013.
- Delogu, E., Boulet, G., Olioso, A., Coudert, B., Chirouze, J., Ceschia, E., Le Dantec, V., Marloie, O., Chehbouni, G., and Lagouarde, J.-P.: Reconstruction of temporal variations of evapotranspiration using instantaneous estimates at the time of satellite overpass, *Hydrol. Earth Syst. Sci.*, 16, 2995–3010, doi:10.5194/hess-16-2995-2012, 2012.
- Falge, E., Baldocchi, D. D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Ta Lai, C., Law, B. E., Meyers, T., Moncrieff, J., Moors, J., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A. E., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for long term energy flux data sets, *Agric. Forest Meteorol.*, 107, 71–77, 2001.
- Fisher, J. B., Baldocchi, D. D., Misson, L., Dawson, T. E., and Goldstein, A. H.: What the towers don't see at night: nocturnal sap flow in trees and shrubs at two AmeriFlux sites in California, *Tree Physiol.*, 27, 597–610, 2007.
- Fuchs, M. and Tanner, C. B.: Calibration and field test of soil heat flux plates, *Soil Sci. Soc. Am. J.*, 32, 326–328, 1968.
- Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G., and Duchemin, B.: Analysis of evaporative fraction diurnal behaviour, *Agric. Forest Meteorol.*, 143, 13–29, 2007.
- Hoedjes, J. C. B., Chehbouni, A., Jacob, F., Ezzahar, J., and Boulet, G.: Deriving daily evapotranspiration from remotely sensed instantaneous evaporative fraction over olive orchard in semi-arid morocco, *J. Hydrol.*, 354, 53–64, 2008.
- Jackson, R. D., Hatfield, J. L., Reginato, R. J., Idso, S. B., and Pinter Jr., P. L.: Estimation of daily evapotranspiration from one time-of-day measurements, *Agric. Water Manage.*, 7, 351–362, 1983.
- Journée, M. and Bertrand, C.: Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements, *Remote Sens. Environ.*, 114, 2692–2704, 2010.
- Kalma, J. D., McVicar, T. R., and McCabe, M. F.: Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data, *Surv. Geophys.*, 29, 421–469, 2008.
- Kustas, W. P., Perry, E. M., Doraiswamy, P. C., and Moran, M. S.: Using satellite remote sensing to extrapolate evapotranspiration estimates in time and space over a semiarid Rangeland basin, *Remote Sens. Environ.*, 49, 275–286, 1994.
- Leuning, R., van Gorsel, E., Massman, W. J., and Isaac, P. R.: Reflections on the surface energy imbalance problem, *Agric. Forest Meteorol.*, 156, 65–74, 2012.
- Lhomme, J.-P. and Elguero, E.: Examination of evaporative fraction diurnal behaviour using a soil-vegetation model coupled with a mixed-layer model, *Hydrol. Earth Syst. Sci.*, 3, 259–270, doi:10.5194/hess-3-259-1999, 1999.
- Mu, Q., Zhao, M., Kimball, J. S., McDowell, N. G., and Running, S. W.: A remotely sensed global terrestrial drought severity index, *Bull. Am. Meteorol. Soc.*, 94, 83–98, doi:10.1175/bams-d-11-00213.1, 2013.
- Otkin, J. A., Anderson, M. C., Mecikalski, J. R., and Diak, G. R.: Validation of GOES-Based Insolation Estimates Using Data from the U.S. Climate Reference Network, *J. Hydrometeorol.*, 6, 460–475, 2005.
- Ryu, Y., Baldocchi, D. D., Black, T. A., Detto, M., Law, B. E., Leuning, R., Miyata, A., Reichstein, M., Vargas, R., Ammann, C., Beringer, J., Flanagan, L. B., Gu, L., Hutley, L. B., Kim, J., McCaughey, H., Moors, E. J., Rambal, S., and Vesala, T.: On the temporal upscaling of evapotranspiration from instantaneous remote sensing measurements to 8-day mean daily-sums, *Agric. Forest Meteorol.*, 152, 212–222, 2012.
- Tolk, J. A., Howell, T. A., and Evett, S. R.: Nighttime evapotranspiration from alfalfa and cotton in semiarid climate, *Agron. J.*, 98, 730–736, 2006.
- Trezza, R.: Evapotranspiration using a satellite-based surface energy balance with standardized ground control. PhD dissertation, Logan, UT: Utah State University, Bio. Irr. Eng. Dept., 2002.
- Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R., Meyer, T. P., Prueger, J. H., Starks, P. J., and Wesely, M. L.: Correcting eddy covariance flux underestimates over a grassland, *Agric. Forest Meteorol.*, 103, 279–300, 2000.
- Van Niel, T. G., McVicar, T. R., Roderick, M. L., van Dijk, A. I. J. M., Renczullo, L. J., and van Gorsel, E.: Correcting for systematic error in satellite-derived latent heat flux due to assumptions in temporal scaling: assessment from flux tower observations, *J. Hydrol.*, 409, 140–148, 2011.
- Van Niel, T. G., McVicar, T. R., Roderick, M. L., van Dijk, A. I. J. M., Beringer, J., Hutley, L. B., and van Gorsel, E.: Upscaling latent heat flux for thermal remote sensing studies: Comparison of alternative approaches and correction of bias, *J. Hydrol.*, 468–469, 35–46, 2012.
- Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D. D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., and Verma, S.: Energy balance closure at FLUXNET sites, *Agric. Forest Meteorol.*, 113, 223–243, 2002.
- Zhang, L. and Lemeur, R.: Evaluation of daily evapotranspiration estimates from instantaneous measurements, *Agric. Forest Meteorol.*, 74, 139–154, 1995.