Hydrol. Earth Syst. Sci. Discuss., 10, 7325–7350, 2013
www.hydrol-earth-syst-sci-discuss.net/10/7325/2013/
doi:10.5194/hessd-10-7325-2013
© Author(s) 2013. CC Attribution 3.0 License.

Hydrology and Earth System Sciences Discussions
Open Access

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

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

C. Cammalleri, M. C. Anderson, and W. P. Kustas

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

Received: 6 May 2013 – Accepted: 19 May 2013 – Published: 7 June 2013

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Four upscaling methods for estimating daytime evapotranspiration (ET) from single time-of-day snapshots, as commonly retrieved using remote sensing, were compared. These methods are based on the assumption of 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 suggests that at-surface solar radiation is 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 did not perform as well as the solar radiation method, but similarly had errors with little seasonal dependency. 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.

Please consider using the term actual ET (ET_a) throughout

1 Introduction

Routine monitoring of 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., 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 example, the overpass time of the Landsat series, in sun-synchronous polar orbit, is around 10:00 local solar time, while MODIS sensors on board of Terra and Aqua platforms have an equator crossing time of 10:30 and 13:30, respectively. Remote ET estimates acquired with these instruments, as a single snapshot during the day, have to be upscaled to longer time scales (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 metric 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 in 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; Delogu et al., 2012; Lhomme and Elguero, 1999). Brutsaert and Sugita (1992) have 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). Additionally, 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.

Other commonly used upscaling methods use the incoming solar radiation, R_s (Jackson et al., 1983; Zhang and Lemeur, 1995), or even the top of atmosphere 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_o) as an upscaling variable, based on the assumption that ET_o incorporates most of the main meteorological factors that influence the evaporative process.

Previous analyses of upscaling methods were focused in many cases on 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 "integrated" daily variables – the nominal representation of "truth" – particularly 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), 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 "integrated" daily ET that can differ between studies, and can lead to different conclusions about optimal upscaling approach.

In this paper we suggest 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 four different upscaling methods is conducted using surface energy fluxes collected by 12 stations from the AmeriFlux network (http://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 "integrated" 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 modeled, 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 09:00 and 15:00 ST).

2 Materials

consider providing units when a flux energy variable is introduced

The daytime total 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 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).

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 EF method (Anderson et al., 1997); the effects of this assumption are discussed.

Since the literature has little information pertaining to systematic errors in RS, TOA and REF methods, especially in the case of daytime fluxes, β is assumed equal to 1 for all these cases. A notable exception is the analysis conducted by Van Niel et al. (2012), where a correction factor for the retrieval of 24 h ET was proposed as a function of day of the year, time-of-day and cloud conditions.

The dataset 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. Turbulent fluxes of sensible (H) and latent heat were obtained from the Level 2 standardized AmeriFlux dataset and observed G_0 values were corrected for heat storage. Data gaps were not filled, and only days with fully available half-hourly daytime data were used in the analysis.

missing

how?

Given the surface energy imbalance typical of EC data, three different "integrated" ET datasets were used in the following analyses: (i) the "Unclosed" dataset, where closure was not enforced; (ii) the "Residual" dataset, where λET is obtained as residual term of the surface energy budget ($\lambda ET = R_n - G_0 - H$); and (iii) the "Bowen" dataset, 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 "integrated" ET datasets. The choice of focusing on daytime fluxes instead of 24 h fluxes is 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 below. In the case of the REF methodology, half-hourly ET_o

was modeled using local meteorological data (Allen et al., 1998), and daytime values were obtained through temporal integration.

3 Analysis methods

3.1 The effects of energy budget closure technique

consider expanding equations

In order to evaluate the performance of the four upscaling methods, they were applied individually to each of the three "integrated" ET datasets. For each site, 7 different possible overpass times-of-day were analyzed (from 09:00 to 15:00 ST, at 1 h time steps), to test dependence of upscaling errors on time of clear-sky acquisition. The methods were applied over the whole year, but only on days that had clear skies 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). Two main statistical metrics were used as descriptors of method performance: (i) relative error, RE (%), computed as the ratio between mean absolute error (MAE; $E|ET_{d-X} - ET_d|$) 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.

move to results section

The plots in Fig. 1 summarize the performance of the different methods in upscaling half-hourly observations collected during the mid-daytime hours (09:00 to 15:00). Figure 1a–c shows RE, averaged over all 12 sites, for the "Bowen", "Residual" and "Unclosed" datasets, respectively, while Fig. 1d–f reports the corresponding average RB values. These results highlight the dependence of apparent method performance on closure technique applied to the EC flux data. This 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 "integrated" cases, and similarly the diurnal shape of both RE and RB lines.

3.2 Statistical analysis approach

needs re-writing

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 intercomparison method that explicitly accounts for the uncertainty in the "integrated" ET has been adopted in this study.

Combining the three "integrated" ET datastreams, the minimum (ET_{min}) and maximum (ET_{max}) daytime ET values are identified for each day. Typically ET_{min} is associated with the measured datastream (no closure correction), while ET_{max} is obtained from the residual closure method, although this is not always the case. The "true" state generally lies within these two boundaries. Moreover, two further thresholds defined as $ET_{min} - \Delta$ and $ET_{max} + \Delta$, with $\Delta = 0.5(ET_{max} - ET_{min})$, are used to discriminate upscaled estimates that are acceptable (ET_{min} to ET_{max}), those with moderate errors (ET_{min} to $ET_{min} - \Delta$ and ET_{max} to $ET_{max} + \Delta$) and major errors ($< ET_{min} - \Delta$ or $> ET_{max} + \Delta$). For each day, 21 daily ET estimates are potentially available for each method using Eq. (1), including 7 possible acquisition times × 3 "integrated" λET_t series. Estimates are combined over a given time interval (e.g., month, year, full two-year sample), then the frequency distribution of the estimates is reconstructed based on these thresholds. These frequency distributions can be used to quantify the accuracy of each method (percentage of estimates between ET_{min} and ET_{max}), as well as systematic positive or negative biases (values $> ET_{max}$ or $< ET_{min}$, respectively).

why "potentially"? It was stated before that only data having a complete daytime record were used (i.e. no gaps)

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 frequency class. The histograms show that all the models have similar accuracy, defined here as the percentage of upscaled values that matched the daytime "integrated" values within the uncertainties of the observations (ET_{min} to ET_{max}). Of the methods tested, RS and REF were most accurate, yielding a peak frequency 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 β = 1 to 44 % with β = 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 %.

what β would have been best? what β would be used for the other approaches? Why only test one extra value of β for only one of the approaches?

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 yields relatively uniform bias for various choices of acquisition time. This characteristic benefits ET retrieval approaches that can use TIR data from a combination of thermal satellite sensors with varying overpass times. The EF approach shows less bias for morning acquisition times (09:00 and 10:00) or for late afternoon (15:00), 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.

Another analysis was performed by splitting the data by month, assuming that a reliable monthly frequency distribution was obtained when more than 15 days of data were available. The plots in Fig. 4 report the all-site average results for the EF (panel a), RS (panel b), TOA (panel c) and REF (panel 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 variability in the 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, 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.

move some of this to methods. If someone wanted to know how many days defined a month, they would look in the methods section, not the results section.

5 Discussion

The statistical analysis of the accuracy of the different daytime upscaling methods discussed in Sect. 4, as quantified by 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 (~ 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.

presenting results in the discussion section

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 10:00 and after 15:00. 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 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 cases 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 significant negative linear correlation with the difference between the overestimation frequency for TOA and RS methods, with a determination coefficient (R^2) equal to 0.74 (Fig. 5).

use different word unless you did a statistical test.

This means that the two methods perform similarly when the sky is clear, while the overestimation in TOA increases under mixed cloud cover conditions. Following Eq. (1), it is clear that the relationship between $R_{s,t}$ and $R_{TOA,t}$ used in Eq. (1) is the same for partially cloudy and clear-sky days (clear-sky at the specific time-of-day is the only

constraint for remote sensing estimations), but during partially cloudy days $R_{TOA,d}$ is greater than $R_{s,d}$ by definition. Moreover, the monthly clear-sky fraction values obtained for all the sites are in general high, ranging only between 0.60 to 0.73; these values seem to 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 dataset, hence the TOA method performs reasonably for most of the days.

reasonably well?

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 small bias observed by Ryu et al. (2012) may related to use of daytime vs. 24 h total ET. The strong correlation observed between cloudiness and the overestimation in TOA in Fig. 5 suggests it might be possible to correct TOA-upscaled estimates when information on cloud-fraction is available. The observed negative relationship between cloudiness and TOA overestimation supports the results reported by Van Niel et al. (2012), although their analysis was limited to two Australian sites and "Unclosed" data only.

may be

already stated above

In terms of accuracy, the EF method performed similarly to RS, especially during the June–August timeframe. 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. To further investigate the root cause of this variability in performance, a more detailed analysis of the impact of the different components of available energy was conducted (Fig. 6).

presenting results in the

One test neglected daytime G_0 in the computation of EF, essentially assuming soil heat flux was 0 when integrated over the daytime hours (referred to as the RN method).

The 1st author has a paper published in Journal of Hydrology on this topic from 2012 that would seem suitable for citing in here somewhere.

This resulted in a further increase in the seasonality of method performance, increasing the temporal standard deviation over time in systematic underestimation from 13 to 18 % and in accuracy from 10 to 13 %. While the temporal variability of the underestimation increases when G_0 is neglected, 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 "integrated" flux was 24 h rather than daytime only. Another test used only the shortwave component of R_n (RSW method). This served to reduce monthly variability in accuracy to 4 %, close to the value observed for the RS method (3 %); however, signs of seasonality are still evident (Fig. 6b). This analysis suggests that the long-wave component of R_n is the main cause of the observed seasonality in the EF method. In general, the use of land-surface related variables appears to degrade the results compared to the simple R_s.

& presentation of results and introduction of new methods in the discussion section.

The results suggest 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), 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 hours 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.

In general, the use of daytime ET instead of 24 h ET as a "integrated" 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 ±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. On the other hand, nighttime eddy covariance measurements of ET are not very reliable, since observations often are made under low winds with minimal turbulence (Falge et al., 2001; Fisher et al., 2007); hence, the inclusion of nighttime EC fluxes in such an analysis can cause greater uncertainty and inconsistent results.

6 Conclusions

– needs re-writing – remove "discussion", Keep & enhance the conclusion text

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 approach that intrinsically accounts for the uncertainty in EC flux tower ET observations was adopted. The results discussed here are therefore independent of the closure method adopted, and better reflect the intrinsic accuracy of the different methodologies apart from measurement 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. Additionally, the relatively high accuracy of remotely-sensed R_s maps already available from geostationary satellites (Otkin et al., 2005; Journée and Bertrand, 2010; Cristóbal and Anderson, 2013) suggests that the RS method has utility for operational use in land surface models applied at large scales.

for example, this is not a conclusion of your paper, but it would be suitable for the discussion section.

The TOA method appears to be less accurate than the RS methods, and yields a systematic overestimation of daytime fluxes related to cloud coverage. Indeed, some authors have already suggested the use of an empirical correction coefficient based on cloud conditions (Van Niel et al., 2012). This solution may be appealing in some applications due to the minimal requirement of information for the assessment of R_{TOA}

maps. However, the need for cloud conditions makes this technique less appealing and straightforward for routine applications. The TOA model seems to perform better for afternoon clear-sky acquisitions, becoming more appealing for applications with sensor 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.

The REF technique returns consistent estimates in terms of accuracy, but with a stable negative bias. For early morning (10:00 LT) 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 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) and (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 long-wave radiation and soil heat flux estimates. Since the analysis was performed using locally observed daytime R_s and $(R_n - G_0)$ values, it likely that in practical applications the RS method would in general perform better than EF in a variety of conditions.

Acknowledgements. 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