Hydrology and Earth System Sciences

Modeling moisture fluxes using artificial neural networks: can information extraction overcome data loss?

A. L. Neal[1], H. V. Gupta[1], S. A. Kurc[2], and P. D. Brooks[1]

[1]Department of Hydrology and Water Resources, University of Arizona, 1133 James E. Rogers Way, Tucson, AZ 85721, USA
[2]School of Natural Resources and the Environment, University of Arizona, 325 Biosciences East, P.O. Box 21004, Tucson, AZ 85721, USA

Abstract. Eddy covariance sites can experience data losses as high as 30 to 45% on an annual basis. Artificial neural networks (ANNs) have been identified as powerful tools for gap filling, but their performance depends on the representativeness of data used to train the model. In this paper, we develop a normalization method, which has similar performance compared to conventional training approaches, but exhibits differences in the timing of fluxes, indicating different and previously unused information in the data record. Specifically, the differences between half-hourly model fluxes, especially during summer months, indicate that the structure of the information content in the data changes seasonally, diurnally and with the rate of data loss. Extracting more information from data may not improve model performance and indicates the need for improved data and models to address flux behavior at critical times. We advise several approaches to address these concerns, including use of separate models for day and nighttime processes and the use of alternate data streams at dawn, when eddy covariance may be particularly ineffective due to the timing of the onset of turbulent mixing.

1 Introduction

Automated field data collection often produces discontinuous data sets as a result of instrument malfunction, power failure, or various other technical and non-technical problems. These discontinuities prove especially problematic for micrometeorological measurements. The expanded use of micrometeorological systems for ecological studies (Baldocchi et al., 2001) has resulted in an increased interest in methods to interpolate values for missing data. Eddy covariance measures landscape-scale energy and mass fluxes in a wide variety of ecosystems at high temporal resolution (generally 30 min accumulations). Towers are employed in a wide array of geographic regions including agricultural lands, temperate forests, tropical rainforests, and a range of arid and semi-arid landscapes (Baldocchi et al., 2001; Kurc and Small, 2007; Scott et al., 2006; Wohlfahrt et al., 2008), leading to important insights into the nature of the soil-vegetation-atmosphere system.

Data acquired using eddy covariance typically has significant gaps caused by insufficient turbulent mixing (Blanken et al., 1998; Goulden et al., 1996), the sensitivity or failure of equipment, as well as poorly identified source areas (Brown-Mitic et al., 2007). Such factors can lead to violations of the assumptions of the eddy covariance technique, resulting in data being discarded during processing. These gaps are often serially correlated to particular events or periods important for observation, such as extreme weather events or nighttime carbon exchange (Falge et al., 2001) and transpiration (Dawson et al., 2007; Fisher et al., 2007).

To develop daily, seasonal, and annual estimates of fluxes, a method to fill gaps by approximating values for missing data is crucial. Gap-filled flux data is also applied in land-surface modeling studies, which often require continuous data streams for parameter identification. The method used for gap filling should result in flux time series that provide realistic assessments of moisture flux, energy and carbon exchange. For example, carbon flux gap filling has been shown to alter estimates of annual carbon exchange, including changes in the source/sink behavior of the carbon flux (Moffat et al., 2007), which presumably better reflects the true behavior of the ecosystem.

Gap filling requires awareness of the nature of missing data; i.e. the user should know what periods have missing data, why those data are missing, and what relationships will produce an appropriate estimate for that data. Gap-filling
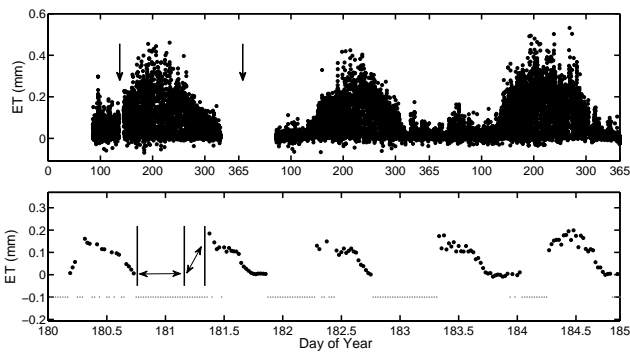
Fig. 1. The time series of latent heat flux used in this study shown as the full record (upper plot) and highlighting in more detail the concerns associated with gap filling, including properly identifying the length of near-zero fluxes at night, and the timing and rate of flux increase during the day (lower plot).

efforts should concentrate on times where flux records are critical, e.g. periods when flux data are important for modeling physical processes, such as the breakup of the stable nocturnal boundary layer, or when values may alter annual flux budgets, such as nocturnal fluxes. In all gap-filling scenarios, the presence of some valid data during similar periods is necessary to identify and model the appropriate relationships to replicate expected values when data are lost. Of course, at very high data loss rates, even a well-trained model may not be sufficient to reproduce the true behavior of a system.

Gap filling requires the use of a model to estimate missing data from existing reference data – numerous models for gap filling have been explored (Falge et al., 2001; Gove and Hollinger, 2006; Hui et al., 2004; Knorr and Kattge, 2005; Papale and Valentini, 2003). Analysis of several models by Moffat et al. (2007) suggests that artificial neural network (ANN) methods can provide substantial benefits for gap-filling studies; they perform similarly to other methods without requiring prior assumptions regarding model structure and are computationally less expensive. Moffat et al. (2007) also note in their conclusion the need to expand testing of gap filling models to a variety of different ecosystems, including arid sites. While arid sites may behave differently under physically based models due to differences in energy and water partitioning, ANNs should demonstrate similar model performance levels in humid and arid regions.

Neural network methods identify input-output relationships in a manner dependent on the information contained in the input and output data sets (MacKay, 2003). For ANN training, the use of data that maximize information about the input-output relationship is key. Because most nighttime eddy covariance data is filtered out, maximizing information requires sampling the input-output pattern for any valid nighttime data. Validity is often established by a criterion such as friction velocity, which describes the degree of turbulent mixing at the sensor (Blanken et al., 1998; Goulden et

al., 1996). Eddy covariance data filtering may include other factors such as advection and stationarity of flow. By extracting the maximum available information from data in conditions near the filter threshold, we can improve the results of gap filling. This study considers an approach to treating data that modifies the distribution of data to extract information from data near the filter threshold.

In brief, gaps in flux records pose problems for the development of seasonal and annual estimates of evapotranspiration at the landscape scale. These gaps also make it difficult to conduct model-based investigations of forcing-response relationships at the land surface. Here, we investigate the use of an ANN framework to fill gaps in energy fluxes, with a particular focus on the probability distributions of flux data and the associated information applied during ANN model training. We apply a standardization technique that converts the probability density function of the flux data to an approximate normal distribution. This transformation moves extreme events toward the central tendency while expanding the region of near-zero fluxes where much of the sensitivity in gap-filled flux records is found (Falge et al., 2001). By altering the shape of the distribution, we can improve the ability of the ANN training algorithm to detect extreme events.

2 Methods

2.1 Data and model structure

Data used for this study were collected at a mesquite woodland site in southeastern Arizona, near Tombstone, AZ (Scott et al., 2004). The data record spans three years (2001–2003) – the tower was not operational during the winter of 2001–2002. Data were filtered according to Scott et al. (Scott et al., 2004). Time series of the latent heat flux (Fig. 1) contain data gaps typical of eddy covariance records, with both long (multiple-day to week) and short gaps (several hours to individual 30-min intervals). Similar patterns of missing data have been found in flux tower records from mesic and humid systems including a range of forest types (tropical and temperate, broadleaf and needleleaf – Falge et al., 2001; Moffat et al., 2007).

Fluxes were filtered prior to modeling for data spikes as well as using a friction velocity (u*) filter (Scott et al., 2006). Fluxes were also corrected for density changes (Webb et al., 1980). Data were tested for subsidence by analyzing the probability density function of vertical wind speed at night. While nearly 50% of the nighttime data had wind speeds above $0.3\,\mathrm{ms}^{-1}$, the u* filter rate was above 70% as well. Applying the remaining data had minimal effect on model results. Nighttime data also show minimal bias for wind direction (not shown).

The input-output structure was similar to simple equations of evapotranspiration (e.g. Penman equation, Alavi et al., 2006) as well as land-surface models (Pitman, 2003). Energy
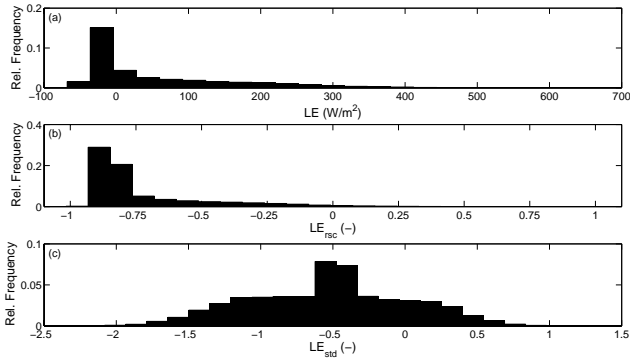
Fig. 2. Histograms of relative frequency of latent heat flux at the flux tower site. Frequency of the raw data is shown in **(a)**. Plot **(b)** is the frequency of the LE$_{rsc}$ (the *mapminmax* function of the Matlab Neural Network Toolbox was used to perform scaling). Plot **(c)** is a histogram of standardized LE as described in Data and Methods. Note that the *mapminmax* retains the original distribution while the standardization transforms the distribution to something near normal for ET.

balance, determined as the slope of a linear fit on turbulent (latent and sensible heat) fluxes and net energy (net radiation minus ground heat flux) was 59%. Energy balance at midday was generally between 50% and 80%, while at night the average was 45%.

The ANN models used in this study compute estimates of latent and sensible heat fluxes based on inputs of precipitation, relative humidity, wind speed, air pressure, net radiation and temperature. Input meteorological data come from a site less than 1 km from the tower and were gap-filled independently using a mean diurnal value method. After running the model, the latent heat fluxes were then converted into evapotranspiration (ET, mm/30-min) for ease of interpretation.

2.2 ANN training and information content

2.2.1 Approaches to ANN training data

When using ANNs, pre-processing methods are used to transform input and output data onto the range ±1 (e.g. Matlab Neural Network Toolbox, MathWorks, Inc., Natick, MA). This ensures that the model will predict outputs based on the scaled variation in and among the data, rather than one dominant data stream of large magnitude. As a way to assess this scaling process, we examine the probability distribution function of the output training data (the measured values of ET) before applying the ANN for gap filling. Two different scaling techniques are applied, the first a simple scaling, reducing values to the range $(-1\ 1)$ and the second making a standardized distribution, which is described below.

In order to standardize the flux data, the raw data is assumed to belong to a gamma distribution. The absolute value of the minimum is added to each point in the raw data so that the entire record is positive. After this shift, a gamma

distribution was fit using a maximum likelihood method in the Matlab statistical functions (*gampdf* and *gamcdf*). These statistical functions were used to transform the data to a normal distribution using a method similar to the standardized precipitation index (SPI, after Mishra and Desai, 2006). The normalization is achieved using an approximation for the normal distribution; (see Mishra and Desai, 2006 for more details):

$$Z = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \text{ for } 0 < H(x) \le 0.5$$
$$Z = \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \text{ for } 0.5 < H(x) < 1 \tag{1}$$

where, Z is the standardized index. The term, t, is calculated as:

$$t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)} \text{ for } 0 < H(x) \le 0.5$$
$$t = \sqrt{\ln\left(\frac{1}{(1 - H(x))^2}\right)} \text{ for } 0.5 < H(x) < 1 \tag{2}$$

where $H(x)$ is the cumulative gamma function, and the coefficients are:

$c_0 = 2.515517$ $c_1 = 0.802853$ $c_2 = 0.010308$
$d_1 = 1.432788$ $d_2 = 0.189269$ $d_3 = 0.001308$

The transformation alters the distribution of the latent heat flux data applied to the model as shown in Fig. 2. The results of this transformation are used directly as the training data for the model. The subsequent output is then converted back to values of latent heat flux using a third-order polynomial fit ($r^2 = 0.98$). By applying this transformation, we are able to alter the sampling patterns used by the ANN during training, and thus enhance the sampling from extreme values of latent heat flux.

2.2.2 Information theory and ANN training

To quantify the effects of different methods of data scaling, we consider the training data in terms of its information content, that is, the behavior "seen" by the model after it is sampled for training. Maximizing the information extracted from data should lead to the best possible model performance. We use the Shannon index to quantify the information associated with the frequency of certain values of latent heat flux. The Shannon index (Shannon, 1948), quantifies the amount of information, $h(x)$, characteristic of a value (or range), x, as a function of its probability $p(x)$, according to the expression:

$$h(x) = -p(x)\ln(p(x)) \tag{3}$$

and the total information content, H, for all values of x in the set, S:

$$H = \sum^{S} h(x) = -\sum^{S} p(x)\ln(p(x)) \tag{4}$$

As noted above, most implementations of ANNs, including the default settings of most software packages, simply

Table 1. Shannon information index (H) values of latent heat flux used for training ANN models. Bins are identical to those used in the histograms in Fig. 2.

ET (mm)	H_{rsc}	ET (mm)	H_{std}
< 10	0.2777	< −3.2	0.0009
10–38	0.1397	(−3.2)–(−1.2)	0.0019
38–66	0.092	(−1.2)–(−0.1)	0.0022
66–95	0.0792	(−0.1)–0.31	0.0758
95–123	0.0694	0.31–0.39	0.1042
123–151	0.0644	0.39–0.51	0.0476
151–179	0.0594	0.51–0.63	0.0287
179–207	0.0531	0.63–0.75	0.0123
207–235	0.0511	0.75–1.4	0.1153
235–264	0.0445	1.4–4.2	0.1172
264–292	0.0396	4.2–11	0.1191
292–320	0.0293	11–22	0.1193
320–348	0.021	22–40	0.1998
348–376	0.017	40–64	0.1921
376–404	0.0138	64–95	0.1198
404–432	0.0096	95–132	0.1094
432–461	0.007	132–176	0.1065
461–489	0.0033	176–227	0.104
489–517	0.0035	227–285	0.0879
517–545	0.0029	285–351	0.0554
545–573	0.0016	351–424	0.0317
573–601	0.0007	424–504	0.0133
601–630	0.0007	504–593	0.005
629–658	0.0002	593–689	0.0015
> 658	0.0002	> 689	0.0002
Total	1.08	Total	1.77

rescale the distribution of the raw data to an interval (−1 1) (e.g. the *mapminmax* function in Matlab). Under this typical rescaling approach, the total Shannon index value of the distribution shown in Fig. 2b is 1.08. To increase the information extracted from extreme low and high values of latent heat flux, the data were transformed from their original distribution into a (near) normal distribution, as described below. The resulting transformation yielded a total Shannon index of 1.77. Shannon values for individual bins in Fig. 2 are summarized in Table 1. At very low values of latent heat flux (less than 95 W/m^2, or 0.04 mm/d), the information content in the original distribution is 0.59 compared to 1.26 in the standardized distribution. The demonstrable increase in information is due to expanding the distribution of low flux values across a greater number of bins.

We estimate the Shannon index, using Eqs. (3) and (4), in the training record. These index values describe the information content passed to the ANN during training. Equations (3) and (4) are not explicitly used to inform ANN training, but instead provide offline information about the characteristics of the data applied to the model.

The two ANN models developed here were trained using (1) the typical approach of rescaling the latent heat flux data, referred to here as the "rescaled" model (similar to the "standard ANN" of Moffat et al., 2007) and (2) the new approach which used a standardized approach, referred to as the "standardized" model. Both models were trained using MAE, as noted above. Model training used data from the year 2003, which represents the longest near-continuous subset of the data that incorporates a full range of seasonal behavior. A one-year data record is the shortest span which can reproduce seasonality without risk of overtraining at longer times (Neal, 2008). From the year 2003 data, 80% of the data were used for training and 20% were used for validation of the model selected randomly. The validation data identify when training is terminated, i.e. when model performance ceases to improve.

Model training used the mean absolute error (MAE) to account for the error structure of individual flux observations (Richardson et al., 2006). Training was performed on the data from 2003, comprising the longest near-continuous record in the data, and was validated on the 2001–2002 time series. Gap filling performance was evaluated using existing gaps in the data to assess the two models in a "real" gap filling application. Where past studies have used data sets with artificial gaps (Moffat et al., 2007), generated by randomly removing existing data, our aim here was to train models with as much data as possible, rather than testing the sensitivity of models to the structure of data loss. Here, model performance on existing data is used to evaluate the predictive capability of the model.

2.3 Marginal Distribution Sampling (MDS)

To provide a further basis of comparison for the two ANN models employed here, we applied the marginal distribution sampling method (Moffat et al., 2007; Reichstein et al., 2005). In short, this method samples flux values within a variable window surrounding a missing data point. The missing data point is filled by the average value under similar meteorological conditions within a ±14-day window. Similar meteorological conditions are defined as net radiation within 50 Wm^{-2}, vapor pressure deficit within 0.5 kPa, and temperature within 2.5 °C. Thresholds were selected to ensure that sufficient data were available to develop average values under consistent meteorological conditions. Any missing data without valid data points within the 14-day window were left missing.

2.4 Model performance

After running the models, the output latent heat flux was converted to evapotranspiration (ET) for ease of comparison between models and to conventional ET measurements from other studies. Results are reported for the raw, unfilled data set (ET$_{raw}$), the rescaled model (ET$_{rsc}$) and the standardized model (ET$_{std}$).

Table 2. Summary of model performance for both ANNs and MDS (ET_{MDS}). Model training used mean absolute error (MAE) to identify parameters.

| Metric	2003			2001–2002		
	ET_{rsc}	ET_{std}	ET_{MDS}	ET_{rsc}	ET_{std}	ET_{MDS}
MAE (mm)	0.0252	0.0329	0.024	0.0272	0.0348	0.0289
RMSE 30min (mm)	0.0403	0.0517	0.041	0.0413	0.0522	0.2273
RMSE daily (mm)	0.7395	0.9301	1.0331	1.0544	1.0876	1.3616
Rel RMSE (–)	0.0369	0.0475	0.0176	0.392	0.4954	0.4828
Correlation	0.8951	0.8326	0.9094	0.8638	0.7916	0.8079

Model performance was determined using several metrics: root mean squared error (RMSE), relative RMSE, MAE and Pearson's correlation. These metrics were calculated for the training period (year 2003) as well as the remaining data record (2001–2002). Model residuals were also used to quantify model performance as a function of data loss. These residuals were calculated for each half hour interval and averaged for the entire record as well as seasonally (see below for a description of season delineation). The fraction of missing data for each half hour during these time periods was used to characterize data loss.

For comparison, uncertainty bounds were determined for the raw data during the training period. Because the tower was only operating during the growing season of 2001 and 2002, uncertainty estimates were only calculated based on the 2003 data. Uncertainty was determined for each 30-min interval using the difference between flux values on days of similar environmental conditions to identify the measurement uncertainty (Richardson et al., 2006).

As noted above, patterns in model performance were analyzed on daily and half-hourly time steps. Performance was also examined based on seasonal behavior. Seasons were defined as winter (December-January-February); monsoon, describing the North American Monsoon (Gochis et al., 2006) and identified by a climatologically defined rainy season (Kurc and Small, 2007) between late June and early September; and pre- and post-monsoon, the remainder of the year not contained in the other two seasons. These seasons are effectively winter, summer and spring/fall, but we use the climatologically defined monsoon to better identify moisture availability during the wet season.

3 Results

Model training resulted in similar performance for both models (Table 2). While the rescaled neural network model (ET_{rsc}) has a slightly lower RMSE and a slightly higher correlation, the difference in both is within the measurement uncertainty for the data. At a daily level, the difference in RMSE values for the two ANN models is lower than the half-hourly RMSE, suggesting that the ANN model training may
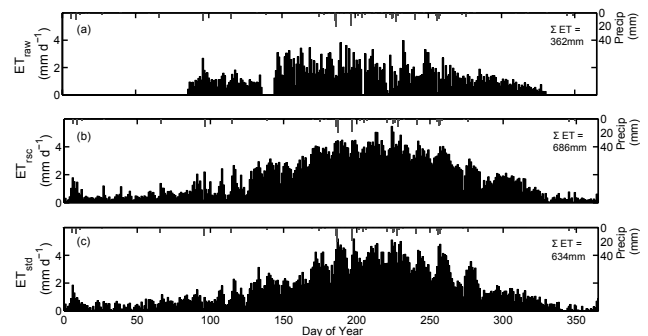

Fig. 3. Precipitation (P) and daily ET for **(a)** ETraw, **(b)** ETrsc, and **(c)** ETstd for the year 2001. Sum of ET over the year is shown in the box associated with each plot (sum of precipitation is 263 mm).

lead to compensation between under- and overestimation at different times during the day.

The MDS method performed similarly to the ANN model during 2003, in which data were available throughout the year, with MAE values within 0.01 mm (Table 2). The MAE and RMSE values for all methods approach the measurement uncertainty of 0.023 mm. However, in 2001 and 2002, when the flux tower was operating only during the growing season, MDS performance is markedly worse than the ANN performance, especially when measured by RMSE (Table 2). In general, based on the objective metrics used here, the performance of the two ANN models and MDS can be judged to be approximately equivalent.

Time series of daily ET for the raw and model data during the year 2001 follow an expected seasonal trend (Fig. 3). The difference between the annual ET derived from the two neural network methods is approximately 50mm. Both of the model methods (ET_{rsc}, and ET_{std}) generally reproduce the seasonal pattern of ET at the daily level, though they differ in the magnitude of ET response to precipitation events, with ET_{std} producing more ET following storms but regressing to lower ET between storms.

Note that ET is greater than annual precipitation (253 mm in 2001). This difference is likely due to the influence of groundwater near the riparian corridor. Plant access to and
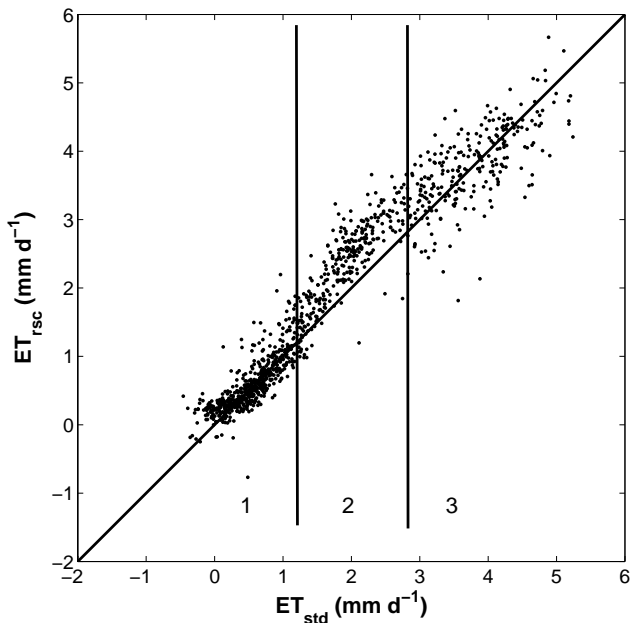
Fig. 4. Scatterplot of ETrsc and ETstd with labels indicating good model agreement (1, winter), relative overestimation in ETrsc (2, pre- and post-monsoon), and high intra-model variability (3, monsoon).

use of groundwater at this site is discussed in several other studies, and has been linked both to inputs from the aquifer as well as bank storage following high flow events in the San Pedro River (Scott et al., 2006; Scott et al., 2004; Scott et al., 2008). The strong atmospheric demand on moisture from the semi-arid climate suggests that the availability of groundwater will have little impact on the ability to model ET based on meteorological variables. In terms of this study, omission of a groundwater term is not likely to affect either model in comparison to the other.

Comparing the two ANN models in terms of daily ET (Fig. 4), the two models deviate from each other at daily ET values around 2 mm/day. These periods (region 2 in Fig. 4) correspond to the pre- and post-monsoon season. At high values of daily ET, i.e. during the monsoon season, comparing the two models shows a high degree of scatter around the 1:1 line. At low values of ET, which generally occur during winter, the two models correspond well at the daily level.

Along with daily ET, the diurnal patterns of ET flux should be reproduced by a gap-filling model. In the overall data set as well as in each of the three seasons, data loss is high during the nighttime and low, though still substantial around midday (Fig. 5). Rates of data loss generally follow the pattern of sunlight hours for each season. These rates of data loss, from the entire data set and each season independently, are used to estimate model performance as a function of data loss.

Applying the rate data loss as a predictor of model performance, using several forms of the model residual, the two

ANN models appear relatively consistent (Fig. 6). Residuals calculated include the absolute (ET_{model}–ET_{raw}) and relative (as a fraction of ET_{raw}) values as well as the residual normalized by the standard deviation of the data and the Nash-Sutcliffe efficiency. The models both show slight improvement in the absolute value of residuals and worsening performance in terms of relative residuals. Most notably, however, the residuals from ET_{std}, which are much larger in magnitude than ET_{rsc} at low rates of data loss, actually indicate improved performance than ET_{rsc} through intermediate levels of data loss (between 40 and 70% loss, Fig. 6).

The diurnal pattern of model performance (Fig. 7) indicates poor model results during dawn in all three seasons and in the overall record. Errors during dawn and near-dawn periods are larger even than dusk and near-dusk times. Because the near-dawn period is associated with the breakup of stable nighttime air and the return of turbulent flux at the boundary layer, poor model performance is related to this change in the nature of the measured flux data.

4 Discussion and conclusions

At a qualitative level, the two ANN models both reproduce the seasonal and event-based patterns in the data record. Timing of peak ET events occurs at similar times in the daily record when viewed on an annual level (Fig. 3). The overall performance of both models is similar, though the standardized model generally had greater error. When examined more closely, the differences in model behavior become more noticeable and indicate the shortcomings of both data and models.

4.1 Model performance and information extraction

The standardized ANN model, with a higher Shannon index value, should yield a trained model that better represent the observed pattern of ET, especially at low flux magnitudes. However, in terms of error metrics, the two ANN models perform similarly (Table 2). The rescaled data appears to be biased toward lower values of ET, while the standardized model favors larger values of ET. In both cases, the performance of each model does not fully replicate the distribution of the raw data. However, the representativeness of any of the three distributions is complicated by the frequent loss of data, especially at low ET.

As shown here, ANNs as gap-filling tools are insensitive to data treatment (rescaled or standardized). The differences in performance between the two models (Table 2) are much smaller than the magnitude of the errors. Slight differences between the two models suggest that problems with model implementation, e.g. input data identification and/or data loss, are more substantial than problems with information extraction.
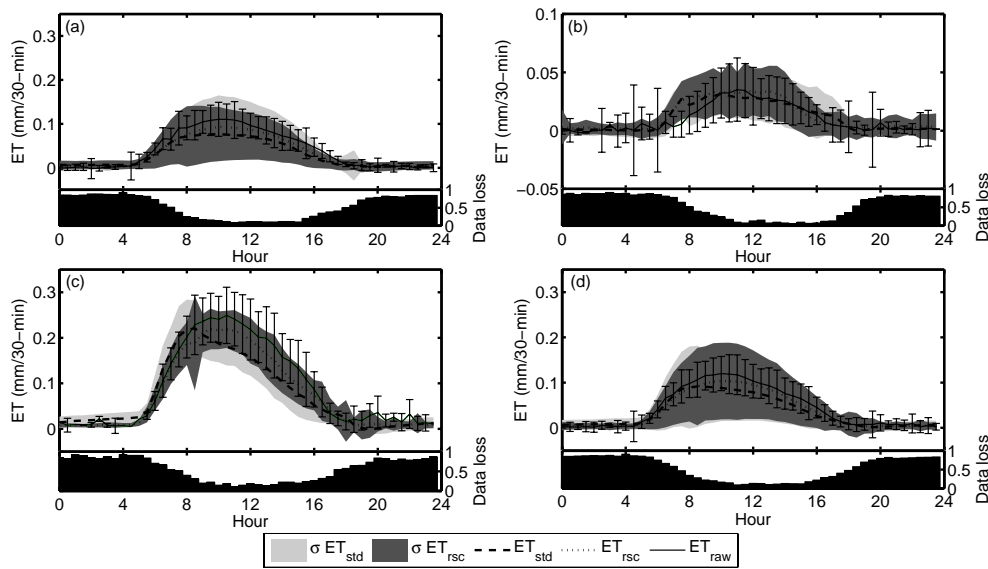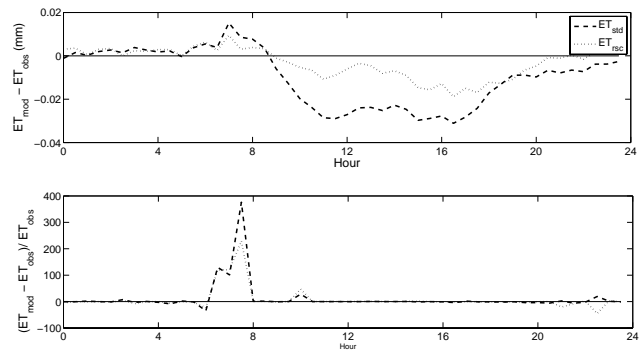
Fig. 5. Diurnal plots from **(a)** pre- and post-monsoon seasons, **(b)** winter, **(c)** monsoon, and **(d)** the full time series. For each plot the upper figure shows the two model results and the raw data. Lower plots show the rate of data loss during that period. Error bar on ET_{raw} representing uncertainty after Richardson et al. (2006).
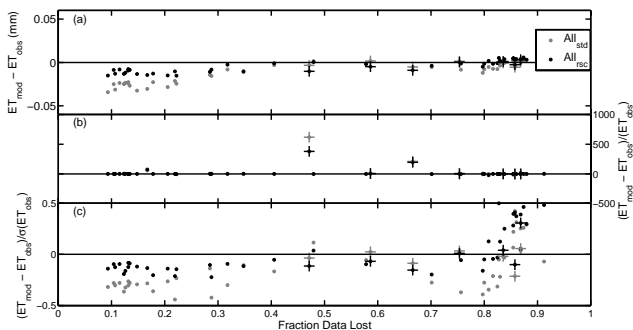
Fig. 6. (a) Model residuals, **(b)** normalized model residuals and **(c)** standardized model residuals plotted as a function of data loss for the full data record. Crosses indicate data between 05:30 and 08:00.

Fig. 7. Diurnal plots of model residuals from the full time series.

The problem of data loss is most noticeable when comparing ANN and MDS results. Long gaps in the flux record during 2001 and 2002 lead to periods when MDS is unable to produce quality gap filled data because the 14-day window contains little or no data. Long gaps may also bias MDS-filled data at the edge of the gap, where the data sampled to develop the replacement data value are consistently at one end of the meteorological "window". This bias at the start and close of the growing season may explain the increased RMSE in MDS during 2001 and 2002, while the relative RMSE and MAE are at values similar to the ANN methods. Overall, the inability of MDS to fill longer gaps makes it a less-attractive option compared to ANN models.

Model residuals as a function of data loss further indicate that the performance of ANNs is insensitive to the treatment applied to the training data. While the standardized method showed slight improvement as a function of data loss, especially at intermediate levels of loss (Fig. 6), the range of error values under both treatments were similarly large and the standard error estimates from both methods fully enveloped those of the other model. Model performance cannot be discriminated due to the large standard deviations of error based on a diurnal pattern. Since both models yield similar error values for the full time series and similarly over/under-estimate the ET, model improvement from data treatment may be limited.

Our findings are contrary to our initial hypothesis, that altering the data distribution would improve model performance by making more information available during model training. Changing the probability distribution of the data should have improved the sampling rate of extreme events

Table 3. Comparison of annual and seasonal sums of ET and the nighttime (ET_n) to daytime (ET_d) ratios for the two ANN models and MDS using the seasonal divisions described in Sect. 3. Data from winter 2001 and 2002 are not valid for ET_{MDS} and ET_{raw} due to limitations in the MDS method (see Sect. 4.1) and tower inoperation, respectively.

| | | ET_{std} | | ET_{rsc} | | ET_{MDS} | | |
		ET	ET_n/ET_d	ET	ET_n/ET_d	ET	ET_n/ET_d	ET_{raw}
2001	Pre/Post	273	0.11	325	0.05	381	0.06	205
	Winter	47	0.13	57	0.07	n/a	n/a	n/a
	Monsoon	260	0.06	267	0.04	290	0.07	132
	All	574	0.10	645	0.06	788	0.03	n/a
2002	Pre/Post	303	0.07	355	0.04	356	0.08	233
	Winter	57	0.04	68	0.04	n/a	n/a	n/a
	Monsoon	166	0.07	174	0.06	188	0.12	101
	All	524	0.07	596	0.05	n/a	n/a	n/a
2003	Pre/Post	263	0.09	306	0.05	395	0.04	291
	Winter	47	0.05	50	0.11	42	0.06	34
	Monsoon	247	0.08	264	0.05	318	0.10	221
	All	554	0.09	618	0.05	750	0.07	544

by narrowing the range of values in the standardized index (Fig. 2). Based on the objective functions (Table 2) and the wide variability of errors as a function of data loss (Fig. 6), it appears that the hypothesis may not be true. This result suggests either that the model structure may be flawed or that the missing data presents a much greater obstacle than the transformation can overcome. In such a case, alternative sources of data must be applied not only for direct analysis but to computational models as well.

4.2 Nocturnal evapotranspiration

One important result when comparing the two ANN models is the difference in nocturnal ET (Table 3). Both models have reduced errors at night, and are similarly prone to over and underestimation, especially when rates of data loss are high (Fig. 6). In a diurnal sense, the standardized model has a tendency to predict higher nighttime ET, especially around dawn, when the model shifts into a mode of increasing ET earlier than the rescaled model. The MDS gap-filled record yields greater total ET in all seasons than the two ANN models, but the ratio of nighttime to daytime ET is similar (Table 3).

Several other studies have reported on ET at night from deciduous forest sites in more humid regions (e.g. Novick et al., 2009). Few have reported on nighttime activity in arid riparian systems. At this site, nighttime ET is estimated at 0.06 and 0.09 fraction of daily ET based on the rescaled and standardized models, respectively. These fractions are similar to those reported for humid forests, indicating that semi-arid riparian species with persistent access to groundwater evapotranspire at similar rates as humid forests during the night. Evaporative demand is dramatically reduced at night when

radiative forcing is absent and relative humidity rises, even in arid locations.

Comparing nocturnal ET to estimated values of daytime and nighttime evaporation (E) and transpiration (T), we find that nighttime ET is a similar fraction of daytime ET as nighttime T to daytime T ($\sim$0.5–0.1, – Fisher et al., 2007). This suggests that plant transpiration scales directly with total ET from day to night. Even under the different model treatments proposed here, the ratio of nighttime to daytime ET are consistent. As noted in the error analysis, the results of gap filling analysis using two different training data sets does not dramatically alter predictions of the nighttime to daytime ET ratio.

4.3 Near-dawn evapotranspiration

The poor model performance at and near dawn provides strong evidence of the flaws associated with eddy covariance data at those times. Conventional methods of eddy covariance data filtering apply a u*-filter or other methods based on turbulence theory. However, the greater error at dawn compared to other nighttime intervals, despite similar friction velocities, indicates that the onset of turbulence is problematic for both measurement and modeling. Because the theoretical basis for eddy covariance is built on strong turbulent mixing, measurements during and immediately following periods of high stability pose substantial difficulties that may not be overcome in data-dependent regression models. Further investigation into near-dawn energy and moisture fluxes will provide insight into how ecosystems use water during this critical period.

In the scope of model-data fusion approaches, the question of eddy covariance data quality remains a problem for

the research community. Model development and calibration is dependent on continuous data records. However, the poor quality and frequent loss of data, especially at night, may require that model results are taken as the standard for comparison against data. The difficulty is in properly identifying the source of the best information for nighttime fluxes. In this study, we show that altering the sampling strategy may not significantly improve model results. We postulate that, in future studies, applying ancillary data streams (e.g. chamber flux measurements or soil moisture data) or using alternate model structures may have a greater effect on gap filling results.

4.4 Implications of this study

Overall, this study points toward several areas for future study in modeling land-atmosphere interactions and gap filling. We reinforce the conclusion of Moffat et al. (2007) that most gap filling methods yield similar performance, although ANNs may have slightly better results. When filling data with lengthy gaps, the disadvantages of statistical methods like MDS become apparent.

Principles of information theory can indicate what flux values are the most informative for modeling applications. Gap filling as a modeling exercise attempts to restore information where it is lost. As shown here, increasing the information content of low flux values may not dramatically improve overall model performance, but performance is not dramatically reduced either. Other methods to extract greater information from the data may prove more effective.

An information theory approach may not satisfy deficiencies inherent to the original data. The poor performance of both ANN models during the near-dawn periods can be linked to the timing of the onset of turbulent mixing as the nocturnal boundary layer breaks up. In this study, both data treatments lead to overestimation of fluxes between 05:30 and 08:00, when the stable nighttime air is becoming unstable due to surface warming. The meteorological variables used as model inputs indicate increasing fluxes earlier than the actual data. Because this time period represents the threshold of feasible data collection by eddy covariance, other methods (e.g. leaf-level measurements of transpiration) should be used to corroborate the data or develop independent models.

Because of the differences in turbulent mixing between nighttime and daytime periods, parallel gap filling models may be an appropriate solution to the information extraction problem identified here. Where most approaches use a single model for nighttime and daytime, assuming that the flux mechanisms and controlling variables are consistent throughout, we propose using different models for night and day, focusing on identifying appropriate model structures for each. Applying two models will also provide another avenue to explore filter criteria under less-than-ideal measurement conditions (van Gorsel et al., 2008).

This study also points out one of the potential problems with model training, that improved model performance under certain conditions (here low flux values) may result in poor performance under other conditions. The "black box" nature of ANN model development means that these trade-offs in model performance come without the ability to fully trace model outcomes to changes in the model structure or training. Comparing the diurnal plots for the two models, it appears that the standardized model under-estimated midday ET, which may have been ignored during training in favor of improving ET performance at other times of the day. The risk of this compensatory effect in model training suggests that different models should be used for filling gaps at night and during the day, especially in light of the different micrometeorological conditions at work in those times. Using separate models would allow one model to track patterns under stable nighttime conditions, while another would follow the flux behavior under turbulent mixing.

Edited by: S. Thompson

References

Alavi, N., Warland, J. S., and Berg, A. A.: Filling gaps in evapotranspiration measurements for water budget studies: Evaluation of a Kalman filtering approach, Agricultural and Forest Meteorology, 141(1), 57–66, 2006.

Baldocchi, D., Falge, E., Gu, L. H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X. H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, U., K. T. P., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, Bulletin of the American Meteorological Society, 82(11), 2415–2434, 2001.

Blanken, P. D., Black, T. A., Neumann, H. H., Den Hartog, G., Yang, P. C., Nesic, Z., Staebler, R., Chen, W., and Novak, M. D.: Turbulent flux measurements above and below the overstory of a boreal aspen forest, Bound.-Lay. Meteorol., 89(1), 109–140, 1998.

Brown-Mitic, C., Shuttleworth, W. J., Harlow, R. C., Petti, J., Burke, E., and Bales, R.: Seasonal water dynamics of a sky island subalpine forest in semi-arid southwestern United States, J. Arid Environ., 69(2), 237–258, 2007.

Dawson, T. E., Burgess, S. S. O., Tu, K. P., Oliveira, R. S., Santiago, L. S., Fisher, J. B., Simonin, K. A., and Ambrose, A. R.: Nighttime transpiration in woody plants from contrasting ecosystems, Tree Physiolog, 27(4), 561–575, 2007.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bern-hofer, C., Burba, G., Ceulemans, G., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for long term energy flux data sets, Agricultural and Forest Meteorology, 107(1), 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 Physiology, 27(4), 597–610, 2007.

Gochis, D. J., Brito-Castillo, L., and Shuttleworth, W. J.: Hydrocli-matology of the North American Monsoon region in northwest Mexico, J. Hydrol., 316(1–4), 53–70, 2006.

Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C., and Wofsy, S. C.: Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accu-racy, Glob. Change Biol., 2(3), 169–182, 1996.

Gove, J. H. and Hollinger, D. Y.: Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange, J. Geophys. Res.-Atmos., 111(D8), 21, doi:10.1029/2005JD006021, 2006.

Hui, D. F., Wan, S. Q., Su, B., Katul, G., Monson, R., and Luo, Y. Q.: Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations, Agr. For-est Meteorol., 121(1–2), 93–111, 2004.

Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, Glob. Change Biol., 11(8), 1333–1351, 2005.

Kurc, S. A. and Small, E. E.: Soil moisture variations and ecosystem-scale fluxes of water and carbon in semiarid grass-land and shrubland, Water Resour. Res., 43(6), W06416, doi:10.1029/2006WR005011, 2007.

MacKay, D.: Information Theory, Inference, and Learning Algo-rithms, edited, p. 640, Cambridge University Press, 2003.

Mishra, A. K. and Desai, V. R.: Drought forecasting using feed-forward recursive neural network, Ecol. Model., 198(1–2), 127–138, 2006.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D. F., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agr. Forest Meteorol., 147(3–4), 209–232, 2007.

Neal, A. L.: Toward a Model-Based Method for Gap Filling Flux Data for a Semi-Arid Site, University of Arizona, Tucson, AZ, 2008.

Novick, K. A., Oren, R., Stoy, P. C., Siqueira, M. B. S., and Katul, G. G.: Nocturnal evapotranspiration in eddy-covariance records from three co-located ecosystems in the Southeastern US: Impli-cations for annual fluxes, Agr. Forest Meteorol., 149(9), 1491–1504, 2009.

Papale, D. and Valentini, A.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Glob. Change Biol., 9(4), 525–535, 2003.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, Int. J. Climatol., 23(5), 479–510, 2003.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Mat-teucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob. Chfange Biol., 11(9), 1424–1439, 2005.

Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flanagan, L. B., Katul, G. G., Munger, J. W., Ricciuto, D. M., Stoy, P. C., Suyker, A. E., Verma, S. B., and Wofsy, S. C.: A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, Agr. Forest Meteorol., 136(1–2), 1–18, 2006.

Scott, R. L., Huxman, T. E., Williams, D. G., and Goodrich, D. C.: Ecohydrological impacts of woody-plant encroachment: sea-sonal patterns of water and carbon dioxide exchange within a semiarid riparian environment, Glob. Change Biol., 12(2), 311–324, 2006.

Scott, R. L., Edwards, E. A., Shuttleworth, W. J., Huxman, T. E., Watts, C., and Goodrich, D. C.: Interannual and seasonal varia-tion in fluxes of water and carbon dioxide from a riparian wood-land ecosystem, Agr. Forest Meteorol., 122(1–2), 65–84, 2004.

Scott, R. L., Cable, W. L., Huxman, T. E., Nagler, P. L., Hernandez, M., and Goodrich, D. C.: Multiyear riparian evapotranspiration and groundwater use for a semiarid watershed, J. Arid Environ., 72(7), 1232–1246, 2008.

Shannon, C. E.: A Mathematical Theory of Communication, The Bell System Technical Journal, 27, 379–423, 623–656, 1948.

van Gorsel, E., Leuning, R., Cleugh, H. A., Keith, H., Kirschbaum, M. U. F., and Suni, T.: Application of an alternative method to derive reliable estimates of nighttime respiration from eddy co-variance measurements in moderately complex topography, Agr. Forest Meteorol., 148(6–7), 1174–1180, 2008.

Webb, E. K., Pearman, G. I., and Leuning, R.: Q. J. Roy. Meteor. Soc., 106(447), 85–100, 1980.

Wohlfahrt, G., Fenstermaker, L. F., and Arnone, J. A.: Large annual net ecosystem CO_2 uptake of a Mojave Desert ecosystem, Glob. Change Biol., 14(7), 1475–1487, 2008.

Hydrol. Earth Syst. Sci., 15, 359–368, 2011

www.hydrol-earth-syst-sci.net/15/359/2011/