Biogeosciences Discuss., 5, 3221–3266, 2008
www.biogeosciences-discuss.net/5/3221/2008/


Biogeosciences Discussions

*Biogeosciences Discussions* is the access reviewed discussion forum of *Biogeosciences*

# Drivers of interannual variability in Net Ecosystem Exchange in a semi-arid savanna ecosystem, South Africa

**S. Archibald[1], A. Kirton[1], M. van der Merwe[1], R. J. Scholes[1], C. A. Williams[2], and N. Hanan[3]**

[1]CSIR Natural Resources and Environment, P.O. Box 395 Pretoria 0001 South Africa
[2]School of Geography, Clark University, USA
[3]Natural Resources Ecology Lab, Colorado State University, USA

Received: 8 July 2008 – Accepted: 16 July 2008 – Published: 22 August 2008

Correspondence to: S. Archibald (sarchibald@csir.co.za)

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





## Abstract

Inter-annual variability in primary production and ecosystem respiration was explored using eddy-covariance data at a semi-arid savanna site in the Kruger Park, South Africa. New methods of extrapolating night-time respiration to the entire day and filling gaps in eddy-covariance data in semi-arid systems were developed. Net ecosystem exchange (NEE) in these systems occurs as pulses associated with rainfall events, a pattern not well-represented in current standard gap-filling procedures developed primarily for temperate flux sites. They furthermore do not take into account the decrease in respiration at high soil temperatures. An artificial neural network (ANN) model incorporating these features predicted measured fluxes accurately (MAE 0.42 g C/m$^2$/day), and was able to represent the seasonal patterns of photosynthesis and respiration at the site. The amount of green leaf area (indexed using satellite-derived estimates of fractional interception of photosynthetically active radiation $f_{APAR}$), and the timing and magnitude of rainfall events, were the two most important predictors used in the ANN model. These drivers were also identified by multiple linear models (MLR), with strong interactive effects. The annual integral of the filled NEE data was found to range from −138 to +155 g C/m$^2$/y over the 5 year eddy covariance measurement period. When applied to a 25 year time series of meteorological data, the ANN model predicts an annual mean NEE of 75 (±105) g C/m$^2$/y. The main correlates of this inter-annual variability were found to be variation in the amount of absorbed photosynthetically active radiation (APAR), length of the growing season, and number of days in the year when moisture was available in the soil.

## 1 Introduction

Carbon dioxide flux measurements using the eddy covariance technique generate a raw dataset with a very high temporal resolution (generally 10–20 Hz). The first step in the analysis of these data is to screen them for spurious values, perform various



BGD

5, 3221–3266, 2008



corrections, and then integrate the fluxes over periods of about 30 min. The half-hour data provides important insights into many short-term physiological processes, but most ecological and management-relevant questions are framed over even longer timeframes – from days to years. A matter of particular interest to both ecologists and ecosystem managers is the inter-annual variability of primary production and carbon storage (Lauenroth et al., 2006). Semi-arid savannas are characterised by high inter-annual variability, in response to highly variable rainfall. This underlies many features of their ecology, including the likelihood and intensity of fires, the growth and migration of animal populations, and the stability of the tree-grass mixture (Higgins et al., 2000; Tyson, 1986; Reed et al., 1994; Ma et al., 2007; Serneels et al., 2007), and makes savanna systems particularly hard to manage.

Accumulating 30 min flux measurements to longer time periods is not a simple matter of adding them up, for two main reasons. The first is that even the best-run eddy covariance datasets have gaps, due to instrument failure or weather conditions that cause the eddy covariance flux assumptions to be violated. The second is that the eddy covariance measurement, net ecosystem exchange (NEE), is often not what is needed by ecologists who are often more interested in its components, gross primary production (GPP) and ecosystem respiration ($R_{eco}$):
NEE=GPP+$R_{eco}$ (observing the convention that fluxes from the atmosphere to the ground are given a negative sign)

A model is used to bridge the data gaps in what is intended to be an unbiased fashion. The same or different models can be used to deconvolve the NEE signal into its components. A wide range of standard procedures have been developed for this process, largely for application in temperate ecosystems (Falge et al., 2001; Papale et al., 2006; Moffat et al., 2007). These are not always appropriate for tropical wet-dry systems. They use phenomenological models, neural networks or process-based models to achieve their objectives. The readily-available ones have proven not to work for data from semi-arid sites in southern Africa. This is because they assume the major controls on flux processes to be solar radiation and temperature, whereas temperatures in the




semi-arid tropics are almost always warm enough to permit physiological activity, and insolation is sufficient, at least during non-cloudy days, for light saturation of part or all of the typically-sparse canopy. In arid and semi-arid systems, the main control on the rate and duration of many ecosystem processes is soil moisture.

As a further complication, in low-rain, high-evaporation ecosystems, where the soils dry out between successive rainfall events (so-called pulse-driven systems), the various terms in the carbon budget are highly dependent on the recent history of the system (Huxman et al., 2004). For example, following a rainfall event, respiration increases rapidly whereas it takes several days for the ecosystem to reach maximum photosynthesis (Williams et al., 2008[1]). Similarly, the magnitude of the system response depends not only on the size of the current rainfall event, but on the amount and timing of preceding events: after a long drought the response to a rain event is larger than to a similar-sized event during the middle of the rainy season, but the time taken to reach the peak response is longer (Veenendaal et al., 2004). Therefore, it is not possible to use instantaneous measures such as the soil moisture content as a sole proxy for the state of the system. Gap-filling therefore requires consideration of indices that have "memory": for instance, accumulators of water deficit.

Moreover, "phenomenological" models will only be appropriate when they truly represent the underlying responses (Falge et al., 2001). Most current respiration models define the relationship between respiration and temperature using an exponential- or logistic-shaped function; i.e. functions that either continually increase, or level off at a maximum value (Moffat et al., 2007). These models were developed in systems where temperature ranges are generally below 30°C (Fang and Moncreiff, 2001; Lloyd and Taylor, 1994). Physiologically, respiration is expected to decrease once temperature exceeds the optimum for microbial activity (Yamano and Takahashi, 1983). In tropical dry systems, the soil temperature in the top centimetres often exceeds 40°C.

[1]Williams, C. A., Hanan, N., and Scholes, R. J.: On the complexity appropriate for modelling observed variation of water and carbon dioxide flux responses to rainfall pulses in and African savanna, Oecologia, in review, 2008.

Thus more appropriate functional forms need to be developed before current gap-filling methodologies can be applied globally.

Improving functional relationships to include extreme conditions would also be valuable in the context of climate change. In coming decades, many ecosystems around the world are likely to be exposed to higher temperatures and reduced moisture availability. Information on ecosystem responses to high temperatures and intermittent droughts will be valuable in predicting responses to these changes.

We present a statistical approach to estimating annual NEE for a semi-arid savanna system in southern Africa. We tested the importance of six environmental drivers of daily photosynthesis (GPP) and respiration ($R_{eco}$) at the Skukuza flux tower in the Kruger Park (25.02° S, 31.50° E). Predictors commonly used in temperate systems were included, together with a range of environmental predictors chosen to reflect the effect of pulsed rainfall events. Predictive models were then used to interpolate annual fluxes over a 25 year time period, and to investigate the degree and possible causes of inter-annual variation in $CO_2$ exchange.

Our approach was motivated by the fact that there was a limited amount and duration of flux data (spanning 6 years with many gaps, which is too short for a reliable estimate of variance), but that a full time series of daily meteorological and phenological data was available for a 25 year period. Working at a daily time-step allowed us to bridge the gap between the half-hourly flux data and the crucial annual timescale, and to use the long-term meteorological data to estimate inter-annual variability. Process-based modelling would be ideal for these systems where previous conditions affect the response of the system to perturbation, but we chose to limit ourselves to a statistical analysis, given our imperfect understanding of the processes driving NEE in these systems. Results from this research will be used to develop more process-based models.

This paper aims to:

– Document new procedures for eddy covariance gap-filling that are appropriate for dry, hot ecosystems;





– Explore the factors associated with short-term (daily) variation in NPP, GPP and $R_{eco}$

– Calculate annual estimates of NEE and explore the main factors driving inter-annual variation in savanna carbon exchange at the Skukuza flux site in South Africa

## 2 Methods

### 2.1 Study site

A flux tower situated in a semi-arid savanna near Skukuza, in the Kruger National Park has been collecting data since February 2000. The site is 370 m above sea level with strongly seasonal rainfall occurring between November and April. Mean annual rainfall is 550±160 mm. The landscape is gently undulating, consisting of broad-leaved *Combretum apiculatum*-dominated savanna on the coarse sand crests and fine-leaved *Acacia nigrescens* savanna on sandy clay loam in the valleys (Scholes et al., 2001). The soils are about 0.6 m deep. The eddy covariance flux tower is situated at the ecotone between the two vegetation types.

The woody vegetation reaches 8–10 m in height and the flux sensors are at 17 m, giving the tower a footprint of about 500 m. The vertically projected tree canopy cover in this area is about 30% and woody basal area is 7 $m^2$ $ha^{-1}$. The grass layer is dominated by *Panicum maximum*, *Digitaria eriantha*, *Eragrostis rigidor*, and *Pogonarthria squarrosa*.

The tower is instrumented with a Gill sonic anemometer measuring wind velocity in three dimensions and a LICOR 6262 closed-path infrared gas analyzer measuring water vapour, $CO_2$ concentration, and pressure. The raw high frequency (10 Hz) data was processed following (Lee et al., 2004) to produce half-hourly measures of above-canopy turbulent fluxes of sensible heat, water vapour, and carbon dioxide. Heat and mass fluxes were calculated based on conventional equations and corrections (see





e.g. Moncrieff et al., 1997; Aubinet et al., 2000) and all fluxes are reported as positive upward from the land to the atmosphere. Canopy storage flux was estimated from the half-hourly time derivative of a 16 m column integral based on $CO_2$ concentrations measured at 0.75, 2.0, 3.5, 5.25, and 16 m, and added to the above-canopy turbulent flux for data analysis. Incoming and outgoing long- and shortwave radiation was measured with Kipp and Zonen shortwave and thermal radiometers mounted at 22 m.

Average half-hourly volumetric soil water content was estimated with 15 cm long Campbell Scientific frequency domain reflectometry probes installed horizontally at soil depths of 3, 7, 16, 30, and 50 cm in the clayey *Acacia*-dominated soils downhill of the tower, and 5, 13, 29, and 61 cm in the sandier *Combretum*-dominated soils uphill. Half-hourly averaged soil heat flux was obtained with HFT3 plates (Campbell Scientific) installed 5 cm below the ground both under and between tree canopies. Rainfall per half hour was measured with a tipping bucket rain gauge located on the tower top, along with other standard meteorological variables such as air temperature and humidity, wind speed and direction.

## 2.2 Data processing and gap filling

Flux data were available from February 2000 to December 2005 (the site continues to operate, but with an open-path IRGA). Of the half-hourly data, 41% was missing, which is slightly more than the average among flux sites of 35% (Falge et al., 2001). As rainfall occurs during summer months of November to April the flux data were summarised by rainfall years (July to June) which provided five full years of flux data – with data coverage ranging from 30 to 74% annually. Most of the data gaps were for a single half hour interval, but instrument failure due to lightning strikes resulted in six gaps of over two months duration, usually occurring during summer periods. These large, non-random gaps limit the types of gap filling approaches that can be used.

When a $u^*$ filter of 0.1 $ms^{-1}$ (Reichstein et al., 2005) was applied to eliminate periods of low turbulence during which eddy covariance measurements are unreliable, the missing flux data increased to 49%. Linear interpolation was used to fill gaps <2 h in




duration, which reduced the missing data to 44%. These half-hourly data were then summed to calculate daily NEE values for all days with unbroken 30-min time series. The result was 698 days of NEE data. These days were not randomly distributed through the year, with the rainy months (particularly December and January) represented by much less data than the dry months of June through September (Fig. 1). Dry, winter conditions are therefore over-represented in the sample. In addition, one of the periods of most continuous and cleanest observations spans an intense drought, 2002–2003 growing season, further biasing results.

Simple gap-filling techniques using mean daily averages are inadequate for filling gaps in the Skukuza data because the stochastic and variable NEE response over the course of a wetting event would not be well represented by a summary value, and because gaps in the data often span several weeks. Non-linear regression methods work well when there is just one main driver of carbon uptake or release (in temperate systems, temperature is normally used to drive respiration, and APAR to drive photosynthesis (Moffat et al., 2007). However, the presence of multiple drivers at the Skukuza site means that single-parameter non-linear methods are unlikely to be sufficient.

Similarly, Marginal Distribution Sampling (Reichstein et al., 2005) uses short periods (15 days) to define temperature response curves, which are then parameterised for specific conditions of soil moisture and leaf display at different times of year. This method presents similar problems to the mean daily average method for the Skukuza site, and does not allow for an explicit exploration of the relationships with any drivers other than temperature.

We used Artificial Neural networks (ANN) as our gap-filling approach, as this method accommodates non-linear relationships between variables but requires few a priori assumptions on the relative importance of different variables or their functional relationships. The usefulness of ANNs depends entirely on the appropriate selection of input variables – the standard ANN gap-filling approach used in the CarboEurope network does not include moisture as a controlling variable (Papale and Valentini, 2003). We also ran standard multiple linear regression models on the data to explore interactive

effects between the variables. This approach allowed us to investigate the important drivers of NEE, as well as develop models which could be used for prediction using long-term meteorological data.

### 2.3 NEE, photosynthesis, respiration

Half-hourly night-time fluxes were used to estimate the day-time respiration. A stricter $u^*$ threshold of 0.25 $ms^{-1}$ (Kutsch et al., 2008) was used for this analysis, as it was more important to have reliable data than large sample sizes. Respiration is controlled by temperature, which generally varies quite predictably over the course of a day, as well as variables such as soil water content and the amount of actively photosynthesising leaf material, which are relatively constant over a single day, but vary over longer time scales. We therefore took a two-scale approach to determining day-time ecosystem respiration: we derived a temperature response curve by fitting it to "optimum" respiration conditions – i.e. the maximum values measured at a range of temperatures (all valid half-hourly night-time fluxes were used for this). This curve was used to estimate the maximum potential respiration rate for each daylight interval, using the daytime temperature trend as input (see Appendix B for more details on this method). The actual respiration during any particular day was then estimated as the temperature-driven "potential" scaled by the ratio of observed night-time respiration to the potential night-time respiration for that day. This scaling factor was assumed to account for the effects of soil moisture and physiological activity. Unlike the MDS method of Reichstein et al. (2005) this method does not require a separate temperature response function to be derived for each day.

Conventional Arrhenius or Lloyd-Taylor temperature functions were not considered appropriate representations of the response functions, as day-time temperatures at the site often exceed that which is optimum for microbial activity (Yamano and Takahashi, 1983). An analysis of independently-collected respiration data from the site, collected using soil chambers, indicated that a generalised Poisson temperature relationship




produced the best fit to measurements of soil respiration (Kirton et al., 2008[2]).

We therefore used the following equation to describe the optimal temperature response:

$$\text{Respiration}=M\left(\frac{b-\text{Soil temperature}}{b-a}\right)^{c}\exp\left\{\left(\frac{c}{d}\right)\left[1-\left(\frac{b-\text{Soil temperature}}{b-a}\right)^{d}\right]\right\}$$

Parameters were estimated using a non-linear least squares by means of the Levenberg-Marquardt algorithm:

$\hat{M}$ = 1.0104 (0.0814);

$\hat{a}$ = 27.6815 (1.5876);

$\hat{b}$ = 11.4221 (10.2024);

$\hat{c}$ = 0.6676 (1.1693);

$\hat{d}$ = 4.1457 (1.7782)

where values in brackets represent the standard error of the estimate. Only days when there were more than three valid night time flux values with which to estimate the scaling parameter were used to interpolate day-time fluxes. See Appendix B for details on this method.

Negative night time fluxes were excluded from the model fitting. Interpolated respiration values that dropped below zero (which can occur at very high or low temperatures, using the parabolic curve) were given a value of zero. This method produces predicted respiration values with similar distributions to those recorded for all conditions of soil moisture and $f_{APAR}$ (Fig. 2).

Daily respiration ($R_{eco}$) values were obtained by calculating a half-hourly value (multiplying the per second value by 60×30) and summing this over the 48 half-hours. All

[2]Kirton, A., Archibald, S., Scholes, R. J., and Makhado, R.: Soil respiration at high temperatures: improving model fit, in preparation, 2008.





other daily values were calculated in the same way. Daily Gross Primary Production (GPP) was calculated by subtracting the interpolated day-time respiration values from the recorded daytime NEE values, and summing over the daylight hours. This resulted in a dataset with 372 valid records for $R_{eco}$ and 529 for GPP.

### 2.4 Drivers of NEE

In temperate systems incoming solar radiation (PAR) and temperature are the main drivers used to predict photosynthesis and respiration. In some models these are modified by measures of LAI and soil moisture (Moffat et al., 2007). We chose to test six input variables as predictors of GPP and $R_{eco}$ (see Table 1).

Only data that could be derived from standard daily South African Weather Services (SAWS) climate records or long-term low-resolution satellite vegetation indices were used as input predictors, in order that the models could be used in conjunction with the long-term records to estimate NEE over periods much longer than the eddy covariance data would permit. The daily time-course of temperature variables was estimated from daily maximum and minimum air temperature. Soil water content was modelled using a simple bucket model and Penman-Monteith evapo-transpiration functions (Archibald and Scholes, 2007). The half-hourly meteorological data available at the flux tower was used to validate these models (see Appendix A).

Three different measures were used to indicate the hydrological state and history of the ecosystem: Relative Plant Available Water (RPAW); water deficit (a function which accumulates the deficit for all days of water stress $\theta<\theta_{crit}$ until rewetting occurs); and time since wetting (the time since the last big wetting event – i.e. time since $\theta$ increased above $\theta_{crit}$). Equations for these indices can be found in Table 1. Mean air temperature – which correlates well with soil temperature (Appendix A) – was used as the predictor of $R_{eco}$, whereas mean daytime temperature was used at the predictor for GPP. The European Joint Research Centre 10-day $f_{APAR}$ product (Pinty et al., 2002) was linearly interpolated to create a daily $f_{APAR}$ parameter. A relationship between AVHRR-derived NDVI (the "GIMMS data", Tucker et al., 2004) and $f_{APAR}$ was used to define the daily

$f_{APAR}$ input for the period before the beginning of the Joint Research Centre (JRC) dataset (see Appendix A).

### 2.5 Modelling approach

Two different artificial neural network (ANN) methods were tested: Generalised Regression Neural Network (GRNN) and Multi-Layer Feed Forward Neural Network (MLF). The GRNN is based on a kernel smoothing approach and has the advantage of using non-parametric regression procedures (which makes no assumptions about the underlying data) and can be trained quickly as only the smoothing parameter needs to be estimated and optimised. As has been found in other studies (Cigizoglu, 2005; Currit, 2002; Kisi, 2006) this method is efficient for modelling non-linear systems and worked as well as the more traditional MLF, which required excessive fine-tuning to optimise the system architecture. Three separate models were developed for predicting $R_{eco}$, GPP, as well as daily NEE. Models were developed using 80% of the data for training and 20% for testing (proportions of 70–30% were also tried, without substantially changing the results).

Multiple linear regression equations with up to three-way interactions were examined for both photosynthesis and respiration. A combination of backward selection and stepwise selection was used to obtain significant predictors in the model. The ability of the MLR to explore the importance of different variables separately and in combination added value to the results of the ANN. However, there are strong theoretical reasons against using ordinary least squares (OLS) regression for data-filling (Richardson and Hollinger, 2005), which is why we restricted their use to exploring the relationships between variables. Many of the meteorological variables, at least over a certain range, are expected to have a near-linear relationship with respiration and photosynthesis. Temperature is an exception: therefore quadratic terms of temperature were also included during the model selection process.





## 3 Results and discussion

### 3.1 Carbon balance

The diurnal time-course of NEE is highly responsive to soil moisture and the presence of green leaves (Fig. 3). Interestingly, maximum $CO_2$ uptake occurs during periods of low soil moisture when green leaves are still present (Williams et al., 2008[1]), because under these circumstances the contribution of soil respiration is low, but a substantial amount of photosynthesis is still occurring using water stored in the plant, or accessed from deeper soil layers that do not contribute much to ecosystem respiration.

### 3.2 Gap-filling: modelling $R_{eco}$ and GPP

Despite the relative paucity of daily data both the ANN and multiple regression methods produced models which reasonably represented the input data (Table 2). Mean absolute error (MAE) ranged from 0.37 to 0.56 g C/m$^2$/day, which compares favourably to the 1–1.5 g C/m$^2$/day range of values reported by Moffat et al. (2007) for a range of gap-filling methods and vegetation types. Respiration was generally harder to predict than photosynthesis, and the linear models performed badly in predicting $R_{eco}$ ($r^2$ of 0.41, MAE of 0.68 g/m$^2$/day).

The ANN identified available green leaf material (indexed by $f_{APAR}$) to be the most important predictor of both $R_{eco}$ and GPP, but $f_{APAR}$ was relatively more important for predicting GPP than for predicting $R_{eco}$, as would be expected (Table 3). We interpret the role of $f_{APAR}$ in driving $R_{eco}$ as reflecting the availability of readily-respired substrate. For GPP the time since wetting event was the next most important predictor, which corroborates findings of Williams et al. (2008)[1] that there is a delay in the pulse of photosynthetic activity after a rainfall event. In terms of water relations, soil moisture content was the best predictor for $R_{eco}$, but water deficit and time since wetting were also identified as important. Interestingly, temperature did not prove to be useful in predicting either respiration or photosynthesis. This could reflect the daily time-step at



BGD
5, 3221–3266, 2008



which we did the analysis – in this sub-tropical system temperature variation between days and over the growth season is much less important than variation in leaf dynamics and soil moisture in driving NEE.

For respiration models using Multiple Linear Regression, $f_{APAR}$ and time since wetting were the most significant single predictors. Interactions between various soil moisture parameters and $f_{APAR}$ also significantly improved the fit of the respiration model. As can be seen in Fig. 3, the effect of a parameter like soil moisture greatly depends on the amount of photosynthesising green leaf material, so it is unsurprising that these interaction terms are important.

In the photosynthesis model soil moisture was very significant, and three-way interactions between $f_{APAR}$, soil moisture, PAR, and time since wetting were important in improving model fit. The importance of the interactive terms perhaps goes some way to representing the delayed photosynthetic response to wetting events identified by Williams et al. (2008)[1]. It usually takes 5–7 days in this system before photosynthesis reaches its maximum after a wetting event, and this response depends on how much leaf material is present. Temperature was included in both the GPP and $R_{eco}$ models as it produced significant interactions with other variables, but as a main effect it was not significant.

The ANN net ecosystem exchange model had the lowest error (Table 2), so this model was used to gap-fill the 6 year dataset . Compared with the gap-filling procedure used by CarboEurope and fluxnet (Papale et al., 2006), our ANN had a closer fit and represented the range of daily NEE values better (Fig. 4).

### 3.3 Inter-annual variability

Annually-integrated net ecosystem exchange varied from −138 to +155 g C/m$^2$/y over the 5 year period for which there was flux data (Table 5). In drought years limited carbon uptake occurs even during the height of summer, but in years with above average rainfall the site can be a sink of carbon for several months of the year (Fig. 5). Only two of the five years had negative NEE (in other words, were net carbon sinks at





the annual timescale). It is possible that our gap filling methods over-estimate the amount of respiration occurring at this site: there was very little data available during the summer months (Fig. 1), so the model was probably not well trained to identify days of maximum GPP in this system. To test this we will need to acquire a more extensive summer dataset for this site. In the meantime, it would be reasonable to assume that in years where the site had a very low positive predicted NEE (such as 2000/2001, with an NEE of 18) the site might actually have been close to carbon-neutral.

When the 25 year NEE sequence is predicted the pattern becomes more obvious (Fig. 6). The site was predicted to be a net sink for carbon in only 6 of the 25 years, but three other years (1989, 1996, and 2000) may have been near-sinks. The data give a long-term mean annual NEE of 75 (±105) g C/m$^2$/y. There is no intrinsic reason to suppose that this site is a long-term source of carbon – the composition and dynamics of the site have been relatively stable over the last 50 years. The population of *Acacia nilotica* trees at the site is senescing, which could add to the respiration term, but saplings of other species are growing up in their place.

Despite the possibility of over-estimation of NEE, the relative difference between years still provides information on the inter-annual variation for this site. Figure 7a indicates that there is a strong relationship between predicted annual NEE and available photosynthetically active radiation (APAR, which is PAR* $f_{APAR}$). This analysis suggests that once annually accumulated APAR exceeds about 675 MJ/m$^2$, the system becomes a sink for carbon (Fig. 7a).

The apparent lack of relationship with total annual rainfall is somewhat surprising. Even when photosynthesis and respiration are considered separately (Fig. 7b, c), by far the best relationship is found with APAR. These results suggest that the integrated response of green leaf material, while itself controlled largely by rainfall, is a better predictor of annual NEE. Perhaps this is because at an annual scale the distribution of rainfall through the year is more important in controlling NEE than the total amount. For example, in the 2003–2004 rainfall year the total annual rainfall was above average (618 mm) but 97 mm had fallen by 31 December, compared with a mean July–




December value of 300 mm. Thus it appears that integrated values of $f_{APAR}$ represent the growing conditions for a season better than total rainfall.

### 3.4 Other pathways of carbon loss from the system

A savanna carbon budget would be incomplete without a consideration of fire and herbivory. The fluxes of $CO_2$ to the atmosphere via these two pathways have not been directly measured at the Skukuza site, but can be inferred and constrained from other data. The abundant large mammalian herbivore (>5 kg body mass) community in this landscape consists of 14 species, mostly Bovidae. The combined herbivore biomass is 3155 kg km$^{-2}$ (Scholes et al., 2004). Taking into account the effect of body mass on metabolic requirements and digestability, this translates to a herbivore respiratory flux of 4.5 g C m$^{-2}$ y$^{-1}$ and a flux from the decomposition of dung of 5.0 g C m$^{-2}$ y$^{-1}$. The uncertainty range associated with these estimates is unknown, but thought to be around 20%, related mostly to errors in game census. The inter-annual variability is thought to be relatively low. The herbivore respiration and dung decomposition fluxes are subsumed in the ecosystem respiration measured by the eddy covariance system (Table 6).

The mean fire return time in this landscape in the KNP is 4.2 years (Van Wilgen et al., 2000). The most comprehensive set of fuel measurements for this landscape was taken in August 1992 at 10 locations within 30 km of the Skukuza site (Shea et al., 1996). The combusted material was predominantly dry grass (1442±975 kg ha$^{-1}$), tree litter (1452±636 kg ha$^{-1}$) and a contribution from dead wood (226±194 kg ha$^{-1}$) giving a total of 3120±1795 kg ha$^{-1}$. A multi-site, multi-year mean grass fuel load for the KNP is 3359 kg ha$^{-1}$, with a range of 1152–6728 (Trollope and Potgieter, 1985). The emission factor for $CO_2$, measured for the same fires as the above fuel loads (Ward et al., 1996) is 1699±33 g $CO_2$ kg DM$^{-1}$. Therefore, the long-term annualised emission of $CO_2$ through fire is around 136±58 g $CO_2$ m$^{-2}$ y$^{-1}$. An additional 6.4±3.9 g CO m$^{-2}$ y$^{-1}$ and 0.2±0.2 g $CH_4$ m$^{-2}$ y$^{-1}$ are also emitted from fires, so the total pyrogenic carbon

loses are around 40.0±17.5 g C $m^{-2}$ $y^{-1}$ (Table 6).

The flux site has burned five times since 2000[3], which suggests that the pyrogenic emissions during this period are probably about twice the long-term, landscape-scale averages calculated above. The pyrogenic fluxes are in principle part of ecosystem respiration, but in practice are not measured by the eddy covariance system because they occur briefly, and during that period exceed the measurement range of the infrared gas analyser. The inter-annual variability is high because a given site does not burn at all in most years, and the fuel load varies greatly in the years when it does burn, in response to the variability of rainfall in the preceding season.

## 4 Conclusions

Inter-annual variability in carbon exchange at the Skukuza flux site is on the same scale as an oak savanna in California (Ma et al., 2007). The variability seems to be largely controlled by variations in the length of time that green leaf is displayed by the trees and grasses, and by changes in seasonal patterns of water availability (Fig. 7) – both ultimately driven by variations in rainfall between years.

Estimates of annual $CO_2$ flux obtained through gap-filling using an ANN may be slight over-estimates (i.e. slightly biased toward the sink side), but the gap-filling procedure developed in this paper represents observed patterns of $CO_2$-exchange better than the standard Carbo-Europe methods (Fig. 4), largely because we explicitly include a soil moisture control, including indices of the wetting history. Results of the ANN gap-filling procedures and MLR models indicate a large degree of interaction between driver variables and lend support for the development of a process-driven model for this system. Such a model would need to include explicit measures of leaf mass, soil moisture and temperature.

The generalised Poisson function used here to fit an optimum temperature response

[3] August 2000, August 2001, April 2005, November 2006, May 2007





curve is an effective method for extrapolating day-time respiration in systems where temperatures often exceed 30°C – provided a scaling factor is used to control for the co-limiting factors of LAI and soil moisture. At a daily to seasonal level, however, temperature was shown to be less important than other factors in influencing NEE.

## Appendix A

### Comparison of meteorological data

Correlation between the flux tower variables and corresponding variables from other sources appears in Table A1. Strong linear relationships exist between the flux tower daily measurements for the mean soil temperature and the mean daytime temperature and the corresponding temperature variables derived from the minimum and maximum daily temperatures of the South African Weather Services (SAWS) data. There is also a strong linear relationship between the measured mean soil moisture and the modelled soil moisture using the SAWS data. There is also a fairly strong linear relationship between PAR derived from the shortwave radiation from the flux tower and the modelled PAR.

The correlation between the flux tower rainfall and the SAWS rainfall is significant, but not as strong as that of the previous comparisons to SAWS derived variables. The peaks of the environmental data are usually slightly higher than recorded from the flux tower, although there are few days when the flux tower recorded higher values. This could be due to localised rainfall events. Peaks in the data do not always correspond and this could be due to the measurements from the SAWS data being taken daily from a rain gauge, whereas the flux tower took instantaneous measurements of rainfall. Therefore daily rainfall events may not always correspond exactly. The pattern of rainfall during time appears to match for the two data sets. The annual sum of rainfall for the environmental data is always more than that for the flux tower data (Table A2). This is due to missing data from the flux tower.





There is a strong linear relationship between Gimms NDVI and $f_{APAR}$ (Table A1). Therefore a linear regression equation was derived to describe this relationship. The linear regression obtained a $r^2$-value of 0.71 and an MAE of 0.05. The estimated equation was: $f_{APAR} = -0.079 + 0.736 \times \text{Gimms}$.

The standard error for the intercept is 0.004 and the standard error for the slope is 0.009.

## Appendix B

### Interpolating day-time respiration

Fitting an optimal temperature function to the mass of night-time flux measurements involved making several assumptions about a) the shape of the temperature-respiration curve, and b) the values to use to fit the curve.

B1 Shape of the temperature-response curve

Field data indicate that a generalised Poisson function is the best descriptor of the effect of temperature on respiration, as it describes both the exponential increase of respiration with temperature and the sudden decrease once the temperature optimum has been reached (Kirton et al., 2008[2]). However, for this analysis we also tried a simple parabolic function.

B2 Values used to fit the curve

This interpolation method relies on deriving a curve that represents the temperature response under a certain set of environmental conditions. Any deviation from this line by an observed point is then assumed to be due to different environmental conditions. The curve can be pulled up and down to match this point, and thereby adjust for these varying environmental conditions, by the use of a scaling parameter. Missing respiration

values (day time points) can then be interpolated on this day (because the environmental conditions other than temperature are going to remain stable at a daily time step) by using the temperature at each point and the adjusted temp/resp equation.

With this in mind, extracting the points to be used could be done in a number of different ways. The easiest way to identify points where all factors other than temperature are constant would be to identify the maximum points for each temperature value (which would represent respiration under completely optimal conditions of soil moisture and LAI). We tried three different methods for extracting these values: manually picking the maximum respiration values, calculating the maximum respiration value for each degree temperature change, and calculating the 95th quantile for each degree temperature change (Fig. B1). We also tried manually picking values at the top of the thickest part of the cloud of respiration points. This approach would exclude any extreme outliers but could also be assumed to represent the same set of other environmental conditions. Because the curve is adjusted up and down based on the respiration values on the day in question, the position of the curve on the y axis is unimportant. It is the shape of the curve that will affect the interpolation.

Using the 95th quantile was not satisfactory as some temperature categories had orders of magnitude more respiration measurements than others. We therefore abandoned that method and tested six different respiration interpolation methods (Table B1): manually selected maximum points (fitting a parabolic and generalised Poisson), manually selected points at edge of data cloud (parabolic and GDP), and calculated maximum points (parabolic and GDP).

## B3 Results

Results indicate that the interpolated values are very resilient to the method used to fit the temperature response curve. The distribution of interpolated points was similar for all six methods (Fig. B2), and linear regression models show similar fits to the observed respiration data (Table B1). A visual assessment of the interpolated points (Fig. B3) indicates that the generalised Poisson interpolations fell more clearly within the main





data cloud. We therefore chose to use the calculated maximum value method fitted to the generalised Poisson distribution.

*Acknowledgements.* We would like to thank Walter Khubeka for his tireless data-collection. This research was funded by grants to Hanan from the US National Aeronautics and Space Administration (NASA) Terrestrial Ecology Program and the US National Science Foundation and CSIR Parliamentary Grant funding to the Natural Resources and Environment Operating Unit.

## References

Archibald, S. and Scholes, R. J.: Leaf green-up in a semi-arid African savanna – separating tree and grass responses to environmental cues, J. Veg. Sci., 18, 583–594, 2007.

Aubinet, M., Grelle, A., Ibrom, A., Rannik, J., Moncreiff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbingier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grunwald, T., Morgenstern, K., Pilegaard, K., rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology, Adv. Ecol. Res., 30, 113–175, 2000.

Cigizoglu, H. K.: Application of generalized regression neural networks to intermittent flow forecasting and estimation, J. Hydrol. Eng., 10, 336–341, 2005.

Currit, N.: Inductive regression: overcoming OLS limitations with the general regression neural network, Computers, Environment and Urban Systems, 26, 335–353, 2002.

Falge, E., Baldocci, 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., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B., Meyers, T., Moncreiff, J., Moors, E., Munger, 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 defensible annual sums of net ecosystem exchange, Agric. For. Meteorol., 107, 43–69, 2001.

Fang, C. and Moncreiff, J. B.: The dependence of soil $CO_2$ efflux on temperature, Soil Biol. Biochem., 33, 155–165, 2001.

Higgins, S. I., Bond, W. J., and Trollope, W. S. W.: Fire, resprouting and variability: a recipe for grass-tree coexistance in savanna, J. Ecol., 88, 213–229, 2000.





Huxman, T. E., Snyder, K. A., Tissue, D., Leffler, A. J., Ogle, K., Pockman, W. T., Sandquist, D. R., Potts, D. L., and Schwinning, S.: Precipitation pulses and carbon fluxes in semiarid and arid ecosystems, Oecologia, 141, 254–268, 2004.

Kisi, O.: Generalized regression neural networks for evapotranspiration modelling, Hydrol. Sci. J., 51, 1092–1104, 2006.

Kutsch, W. L., Hanan, N., Scholes, R. J., McHugh, I., Kubheka, W., Eckhardt, H., and Williams, C.: Response of carbon fluxes to water relations in a savanna ecosystem in South Africa, Biogeosciences Discuss., 5, 2197–2235, 2008, http://www.biogeosciences-discuss.net/5/2197/2008/.

Lauenroth, W. K., Wade, A. A., Williamson, M. A., Ross, B. E., Kumar, S., and Carieveau, D. P.: Uncertainty in calculations of Net Primary Production for grasslands, Ecosystems, 9, 843–851, 2006.

Lloyd, J. and Taylor, J. A.: On the temperature dependence of soil respiration, Functional Ecology, 8, 315–323, 1994.

Ma, S., Baldocchi, D. D., Xu, L., and Hehn, T.: Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, Agric. For. Meteorol., 147, 157–171, 2007.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D., 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., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agric. For. Meteorol., 147, 209–232, 2007.

Moncrieff, J. B., Massheder, J. M., deBruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide, J. Hydrol., 189, 589–611, 1997.

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

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

Pinty, B., Gobron, N., Melin, F., and Verstraete, M. M.: A time composite algorithm for FAPAR





products, Institute for Environment and Sustainability, Joint Research Centre, Ispra, Theoretical basis document EUR 20150 EN, 8, 2002.
Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O.: Measuring phenological variability from satellite imagery, J. Veg. Sci., 5, 703–714, 1994.
Reichstein, M., Falge, E., Baldocci, D., Papale, D., Aubinet, M., Berbingier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrahkova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourvical, J., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, C., and Valentini, R.: On the separtion of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biol., 11, 1424–1439, 2005.
Richardson, A. D. and Hollinger, D. Y.: Statistical modelling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models, Agric. For. Meteorol., 131, 191–208, 2005.
Scholes, M. C., Scholes, R. J., Otter, L. B., and Woghiren, A. J.: Biogeochemistry: the cycling of elements, in: The Kruger Experience: ecology and managment of savanna heterogeneity, edited by: Du Toit, J., Biggs, H. C., and Rogers, K. H., Island Press, Washington D.C., 130–148, 2004.
Scholes, R. J., Gureja, N., Giannecchinni, M., Dovie, D., Wilson, B., Davidson, N., Piggott, K., McLoughlin, C., van der Velde, K., Freeman, A., Bradley, S., Smart, R., and Ndala, S.: The environment and vegetation of the flux measurement site near Skukuza, Kruger National Park, Koedoe, 44, 73–83, 2001.
Serneels, S., Linderman, M., and Lambin, E. F.: A multilevel analysis of the impact of land use on interannual land-cover change in East Africa, Ecosystems, 10, 402–418, 2007.
Shea, R. W., Shea, B. W., Kaufman, J. B., Ward, D. E., Haskins, C. I., and Scholes, M. C.: Fuel mass and combustion factors associated with fires in savanna ecosystems of South Africa and Zambia, J. Geophys. Res., 101, 23 551–23 568, 1996.
Trollope, W. S. W. and Potgieter, A. L. F.: Fire behaviour in the Kruger National Park, Journal of the Grassland Society of Southern Africa, 2, 17–22, 1985.
Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D., Pak, E. W., Mahoney, R., Vermote, E., and El Saleous, N.: An Extended AVHRR 8-km NDVI Data Set Compatible with MODIS and SPOT Vegetation NDVI Data., Int. J. Rem. Sens., 26, 4485–5598, http://glcf.umiacs.umd.

edu/data/gimms/, 2005.
Tyson, P. D.: Climatic Change and Variability in Southern Africa, Oxford University Press, Cape Town, 284 pp., 1986.
Van Wilgen, B. W., Biggs, H. C., O'Regan, S. P., and Mare, N.: A fire history of the savanna ecosystems in the Kruger National Park, South Africa, between 1941 and 1996, South African J. Sci., 96, 167–178, 2000.
Veenendaal, E. M., Kolle, O., and Lloyd, J.: Seasonal variation in energy fluxes and carbon dioxide exchange for a broad-leaved semi-arid savanna (Mopane woodland) in Southern Africa, Global Change Biol., 10, 318–328, 2004.
Ward, D. E., Hao, W. M., Susson, R. A., Babbit, R. E., Shea, R. W., Kauffman, J. B., and Justice, C. O.: Effect of fuel composition on combustion efficiency and emission factors for African savanna ecosystems, J. Geophys. Res., 101, 23 569–23 576, 1996.
Yamano, H. and Takahashi, K.: Temperature effect on the activity of soil microbes measured from heat evolution during the degradation of several carbon sources, Acric. Biol. Chem., 47, 1493–1499, 1983.





**Table 1.** Defining the six input variables used in the models to predict GPP and $R_{eco}$. All input variables were derived from data available at a daily level from the SA Weather Services, so they could be used to produce long-term predictions.

| Parameter | | Derivation | GPP predictor | $R_{eco}$ predictor |
|---|---|---|---|---|
| Photosynthetically Active Radiation | PAR | Modelled (energy balance) | × | × |
| Mean temperature during the day | $T_{pn}$ | $T$min+0.75*($T$max–$T$min) | × | |
| Soil temperature | $T_{re}$ | ($T$max+$T$min)/2 | | × |
| Fraction of absorbed PAR | $f_{APAR}$ | Modelled from satellite-derived reflectances (JRC: http://fapar.jrc.it/Home.php) | × | × |
| Relative Available Water Content (RAWC) | $\theta$rel | ($\theta$-WP)/(FC-WP)×100 | × | × |
| Accumulated water deficit | water deficit | If ($\theta$<$\theta$crit) Σ ($\theta_i$ – $\theta$crit) If($\theta$>$\theta$crit) 0 | × | × |
| Period of wet soils | time since wetting | (While $w$def=0) Σ days since $w$def=0 | × | × |



BGD
5, 3221–3266, 2008





**Table 2.** Comparison of model performance. Artificial Neural Networks (ANN) generally performed better than multiple linear regressions (MLR), but MLR's still managed to explain a large proportion of the variance in photosynthesis.

| | ANN | | | MLR | |
|---|---|---|---|---|---|
| | $R_{eco}$ | GPP | NEE | $R_{eco}$ | GPP |
| MAE (g C/m$^2$/day) | 0.56 | 0.37 | 0.42 | 0.85 | 0.62 |
| $r^2$ | – | – | – | 0.41 | 0.68 |
| $n$ | 372 | 529 | 698 | 372 | 529 |





**Table 3.** Relative importance (percentage) of the different variables used to predict ecosystem respiration, gross primary productivity, and net ecosystem exchange using an ANN.

| $R_{eco}$ | | GPP | | NEE | |
|---|---|---|---|---|---|
| $f_{APAR}$ | 36% | $f_{APAR}$ | 46% | $f_{APAR}$ | 27% |
| RAWC | 19% | time since wetting | 19% | RAWC | 26% |
| PAR | 18% | PAR | 14% | time since wetting | 14% |
| time since wetting | 14% | RAWC | 12% | water deficit | 14% |
| water deficit | 13% | water deficit | 5% | $T_{pn}$ | 10% |
| $T_{re}$ | 0% | $T_{pn}$ | 4% | $T_{re}$ | 6% |
| | | | | PAR | 3% |

**Table 4.** Results of a multiple linear regression to predict ecosystem respiration **(a)**, and GPP **(b)**. The best respiration model included $f_{APAR}$, time since wetting, soil temperature, and relative available water content, and two-way interactions between these variables. This corroborates the findings of the ANN model, but does not produce a good prediction ($r^2$=0.41, MAE=0.85 (g C/m$^2$/day)). The best GPP model included $f_{APAR}$, time since wetting, relative available water content, mean daytime temperature, and three-way interaction between several variables. This also corroborates ANN results, and produces a reasonable prediction ($r^2$=0.68, MAE=0.62 (g C/m$^2$/day)).

| **(a)** | Estimate | Std. Error | t-value | P | |
|---|---|---|---|---|---|
| $f_{APAR}$: time since wetting | 1.21 | 0.33 | 3.70 | 0.000 | *** |
| $f_{APAR}$ | 45.91 | 14.71 | 3.12 | 0.002 | ** |
| RAWC: $T_{re}$ | 0.02 | 0.01 | 2.92 | 0.004 | ** |
| time since wetting | −0.27 | 0.10 | −2.80 | 0.005 | ** |
| $f_{APAR}$: PAR: time since wetting | −0.13 | 0.05 | −2.56 | 0.011 | * |
| $f_{APAR}$: time since wetting:RAWC | −0.03 | 0.01 | −2.54 | 0.012 | * |
| $f_{APAR}$: $T_{re}$ | −1.48 | 0.62 | −2.38 | 0.018 | * |
| time since wetting:RAWC | 0.01 | 0.00 | 2.33 | 0.020 | * |
| $f_{APAR}$: RAWC | −0.36 | 0.18 | −1.97 | 0.049 | * |
| (Intercept) | −4.24 | 2.42 | −1.75 | 0.081 | . |
| RAWC | −0.19 | 0.13 | −1.52 | 0.131 | |
| $T_{re}$ | 0.16 | 0.11 | 1.49 | 0.139 | |
| PAR: time since wetting | 0.02 | 0.02 | 1.45 | 0.149 | |
| PAR | 0.22 | 0.28 | 0.78 | 0.437 | |
| $f_{APAR}$: PAR | 1.01 | 1.39 | 0.73 | 0.469 | |





**Table 4.** Continued.

| (b) | Estimate | Std. Error | t-value | P | |
|---|---|---|---|---|---|
| RAWC | 0.99 | 0.15 | 6.72 | 0.000 | *** |
| $f_{APAR}$: PAR: RAWC | 0.40 | 0.08 | 5.35 | 0.000 | *** |
| $f_{APAR}$: RAWC | −1.89 | 0.42 | −4.50 | 0.000 | *** |
| RAWC: $T_{pn}$ | −0.02 | 0.01 | −4.33 | 0.000 | *** |
| PAR: time since wetting: RAWC | 0.00 | 0.00 | −4.27 | 0.000 | *** |
| $f_{APAR}$: time since wetting: RAWC | 0.03 | 0.01 | 4.25 | 0.000 | *** |
| PAR | 2.00 | 0.52 | 3.88 | 0.000 | *** |
| $f_{APAR}$: water deficit | 0.93 | 0.25 | 3.69 | 0.000 | *** |
| $f_{APAR}$: PAR | −6.33 | 1.75 | −3.63 | 0.000 | *** |
| water deficit | −0.12 | 0.03 | −3.50 | 0.001 | *** |
| PAR:time since wetting: $T_{pn}$ | 0.00 | 0.00 | 3.39 | 0.001 | *** |
| PAR: RAWC | −0.08 | 0.02 | −3.29 | 0.001 | ** |
| $f_{APAR}$: PAR: time since wetting | 0.17 | 0.06 | 3.08 | 0.002 | ** |
| PAR: $T_{pn}$ | −0.05 | 0.02 | −2.77 | 0.006 | ** |
| $f_{APAR}$: time since wetting: $T_{pn}$ | −0.07 | 0.03 | −2.73 | 0.007 | ** |
| PAR:time since wetting | −0.09 | 0.03 | −2.68 | 0.008 | ** |
| time since wetting:RAWC | −0.02 | 0.01 | −2.48 | 0.013 | * |
| time since wetting: RAWC: $T_{pn}$ | 0.00 | 0.00 | 2.41 | 0.016 | * |
| $f_{APAR}$ | −31.56 | 13.95 | −2.26 | 0.024 | * |
| $f_{APAR}$: $T_{pn}$ | 1.06 | 0.61 | 1.74 | 0.083 | . |
| (Intercept) | −5.34 | 3.27 | −1.63 | 0.103 | |
| $T_{pn}$ | 0.21 | 0.14 | 1.53 | 0.126 | |
| $f_{APAR}$: time since wetting | 0.77 | 0.63 | 1.22 | 0.223 | |
| time since wetting | 0.20 | 0.19 | 1.06 | 0.291 | |
| time since wetting: $T_{pn}$ | 0.00 | 0.01 | −0.26 | 0.792 | |





**Table 5.** Summary of NEE over the 5 year period for which there was flux data. Negative values represent an overall sink of carbon. Data gaps were filled using an ANN and predictors $f_{APAR}$, water deficit, relative soil moisture content, mean day time temperature, time since wetting, and mean soil temperature, in that order of importance. [2mm] Also reported are annual summaries of rainfall, available photosynthetically active radiation, length of the growing season, and number of growth days (days when soil moisture content is greater than $\theta_{crit}$ (7% by volume)).

| Rainfall year (July to June) | Annual NEE | Annual Rainfall | Annual APAR (sum) | Length of growing season | Number of growth days |
|---|---|---|---|---|---|
| 00_01 | 42 | 659 | 662 | 244 | 245 |
| 01_02 | 155 | 572 | 523 | 191 | 169 |
| 02_03 | 150 | 303 | 406 | 156 | 166 |
| 03_04 | −138 | 618 | 555 | 188 | 81 |
| 04_05 | −83 | 760 | 665 | 197 | 186 |





**Table 6.** Annualised summary of the different contributions to the carbon balance at the Skukuza flux site.

| | Mean annual flux |
|---|---|
| Herbivory | 9.5 g C $m^{-2}$ $y^{-1}$ (unknown error ?20%) |
| Fire | 33.6±14.7 g C $m^{-2}$ $y^{-1}$ |
| Flux measurement (incl. herbivory) | 75±105 g C $m^{-2}$ $y^{-1}$ |
| Total | 108.6±119.7 g C $m^{-2}$ $y^{-1}$ |





**Table A1.** Summary of comparisons between flux tower derived variables and corresponding variables derived from other sources.

| Variables compared | Pearson correlation | 95% confidence interval |
|---|---|---|
| Mean flux tower soil temperature and derived soil temperature from SAWS data (Tre). | 0.92 | (0.92; 0.93) |
| Mean flux tower daytime temperature and derived daytime temperature from SAWS data (Tpm). | 0.96 | (0.95; 0.96) |
| Scaled flux tower soil moisture and derived scaled soil moisture from SAWS data ($\theta_{rel}$) | 0.78 | (0.75; 0.80) |
| Daily flux tower rainfall and SAWS rainfall data. | 0.61 | (0.58; 0.64) |
| $f_{APAR}$ and GIMMS NDVI. | 0.84 | (0.83; 0.85) |
| PAR calculated from the flux tower data and the modelled PAR data | 0.62 | (0.58; 0.66) |





**Table A2.** Annual rainfall over time.

| Annual Rainfall Sum from SAWS Environmental Data | | | | | | |
|---|---|---|---|---|---|---|
| 99/00 | 00/01 | 01/02 | 02/03 | 03/04 | 04/05 | 05/06 |
| 363 | 659 | 572 | 302 | 618 | 760 | 249 |
| **Annual Rainfall Sum from Flux Tower Data** | | | | | | |
| 99/00 | 00/01 | 01/02 | 02/03 | 03/04 | 04/05 | 05/06 |
| 415 | 671 | 427 | 310 | 276 | 582 | 209 |





**Table B1.** The six different methods used to fit a temperature response curve to the measured night-time (respiration) fluxes. Two different fitting functions were used, and three different methods for identifying points to fit the curve to. The distributions of the data interpolated with each method were very similar to each other (Fig. 2), and fell well within the bounds of the observed respiration data (Fig. 3).

| | Parabolic | | | Generalised Poisson | | |
|---|---|---|---|---|---|---|
| Name | Observed max parObsMax | Calculated max parCalcMax | Observed fit to datacloud parObsMain | Observed max poisObsMax | Calculated max poisCalcMax | Observed fit to datacloud poisObsMain |
| $r^2$ | 0.57 | 0.58 | 0.56 | 0.56 | 0.58 | 0.56 |
| slope of linear model | 0.61 | 0.61 | 0.6 | 0.56 | 0.58 | 0.56 |
| Median predicted value mg $CO_2/m^2/s$ | 0.070 | 0.070 | 0.069 | 0.071 | 0.070 | 0.069 |
| Minimum predicted value mg $CO_2/m^2/s$ | 0.002 | 0.002 | 0.000 | 0.002 | 0.002 | 0.000 |
| Maximum predicted value mg $CO_2/m^2/s$ | 0.98 | 0.68 | 0.81 | 0.81 | 0.71 | 0.81 |





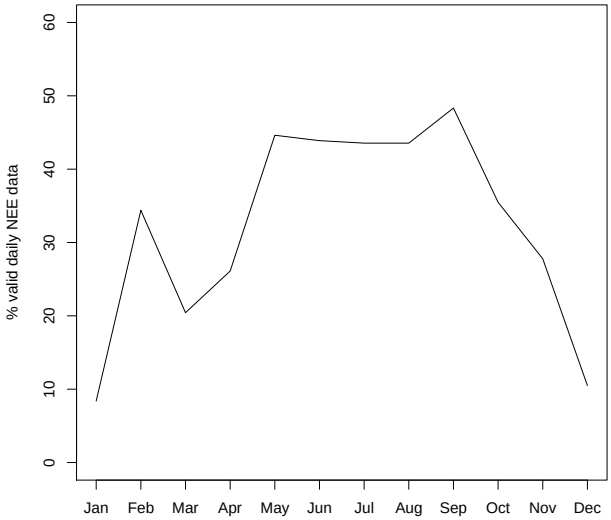


**Fig. 1.** Seasonal distribution of valid NEE data points from a six-year long dataset at the Skukuza flux tower.



BGD

5, 3221–3266, 2008



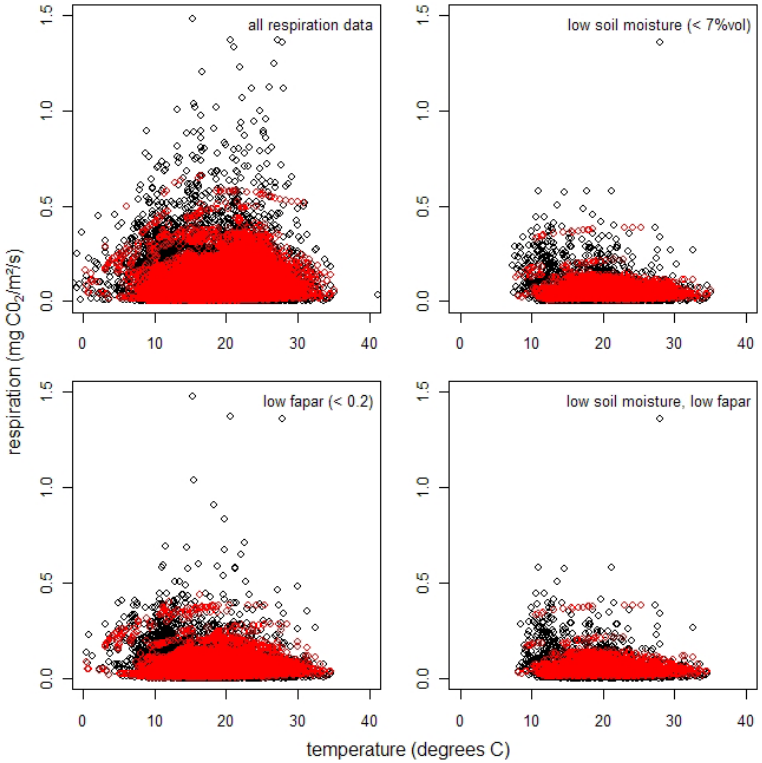


**Fig. 2.** Distribution of observed (black) and interpolated (red) half-hourly respiration values over temperature. Data are presented for all conditions, for periods of low soil moisture, for periods with little leaf material (low $f_{APAR}$), and for conditions of low soil moisture and $f_{APAR}$. Interpolated values lie well within the distribution of observed values for all conditions. It is also clear that respiration drops off at high temperatures, and that temperature-response functions need to include this reduction at high temperatures if they are to be appropriate for this site.





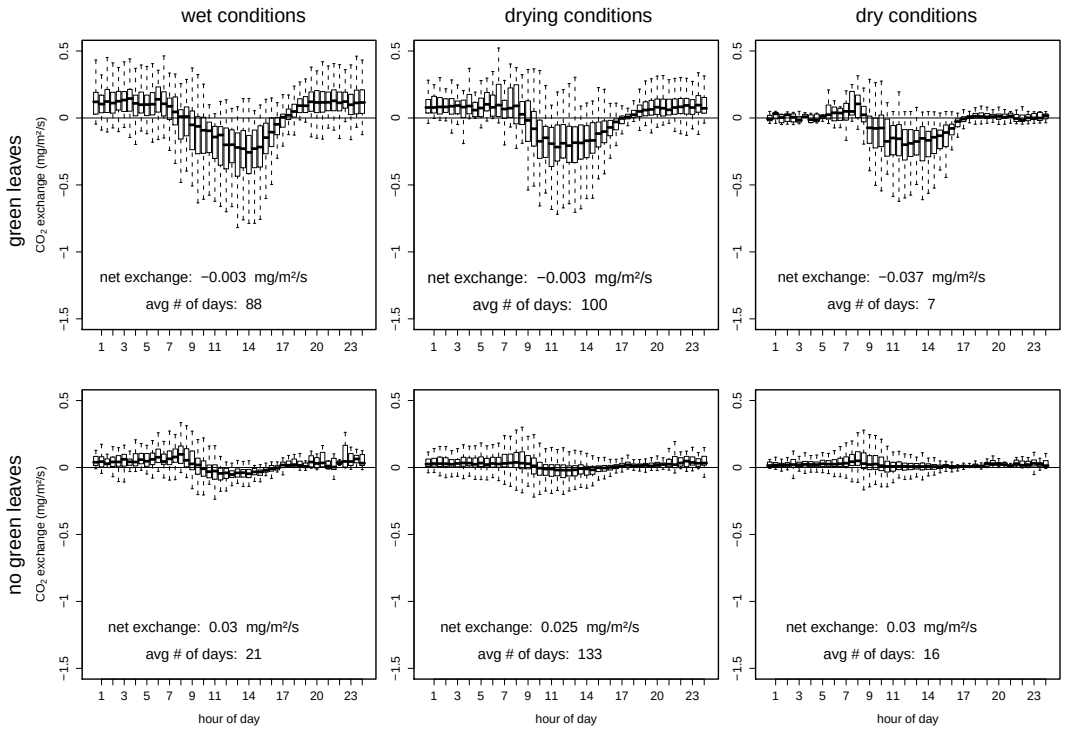


**Fig. 3.** Daily time-course of NEE averaged over 5 years of measurements and for six combinations of environmental conditions at the Skukuza flux site. Maximum $CO_2$ sequestration occurs when soil moisture is low but green leaves are still present. Wet conditions were defined as periods when the soil moisture was greater than 9% volumetric water content, dry conditions, less than 6%. Periods with green leaves were defined as periods when the $f_{APAR}$ value was greater than 0.2. The average number of days each year for each combination of physiological and soil moisture conditions are shown.



BGD
5, 3221–3266, 2008



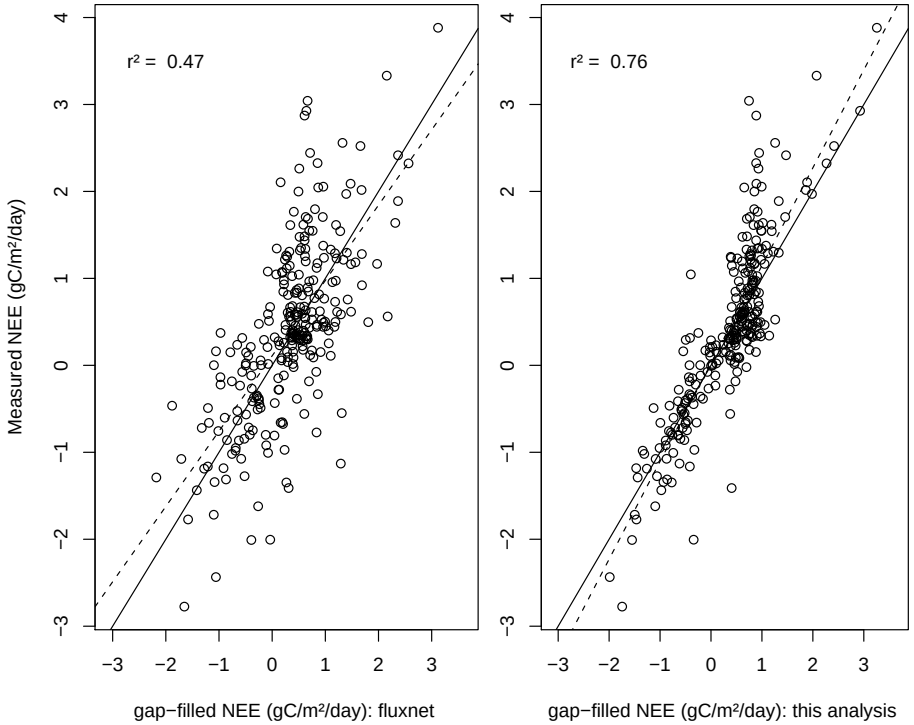


**Fig. 4.** Comparison of measured NEE and that modelled using the standard Fluxnet neural network approach (no soil moisture inputs; Papale et al., 2006), and the neural network developed in this manuscript (three different hydrological inputs). Units are the same for the x- and y-axes (g C/m$^2$/day) Dashed line represents the linear fit of the data, solid line a 1:1 relationship. The predictions using our model have a closer fit, and represent the range of measured values better. Fluxnet gap-filled data were only available from January 2000 to December 2003.

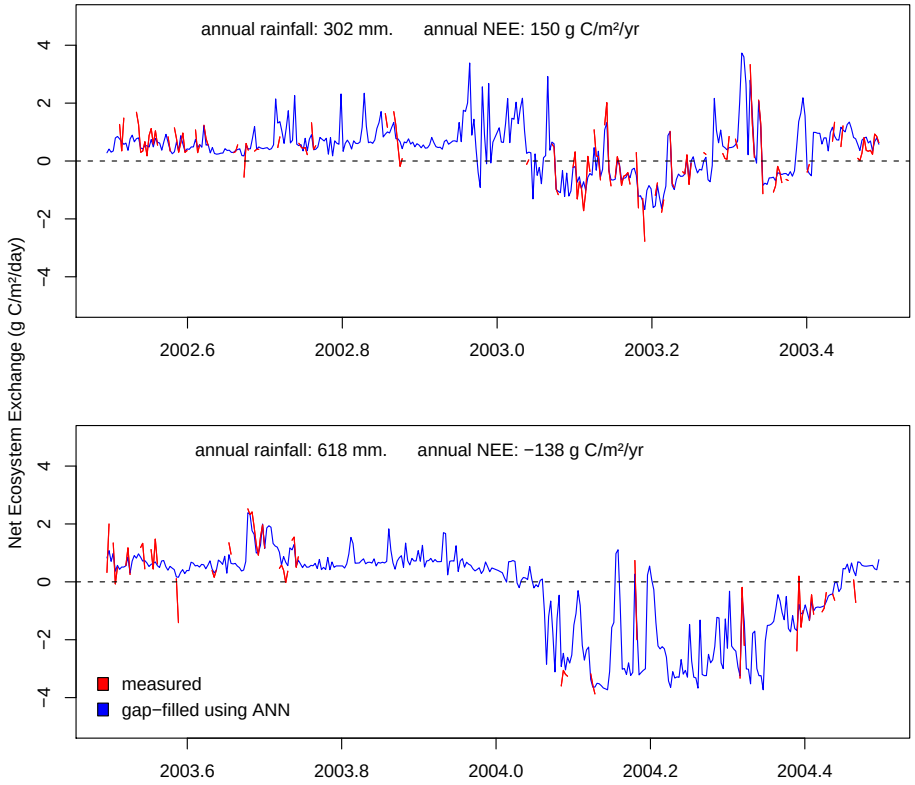


**Fig. 5.** Annual time course of NEE for two consecutive rainfall years (a dry year and a near average year) at the Skukuza flux tower. Red line represents measured daily NEE, blue is modelled using an artificial neural network and inputs of $f_{APAR}$, soil moisture, temperature, time since wetting, and water deficit.

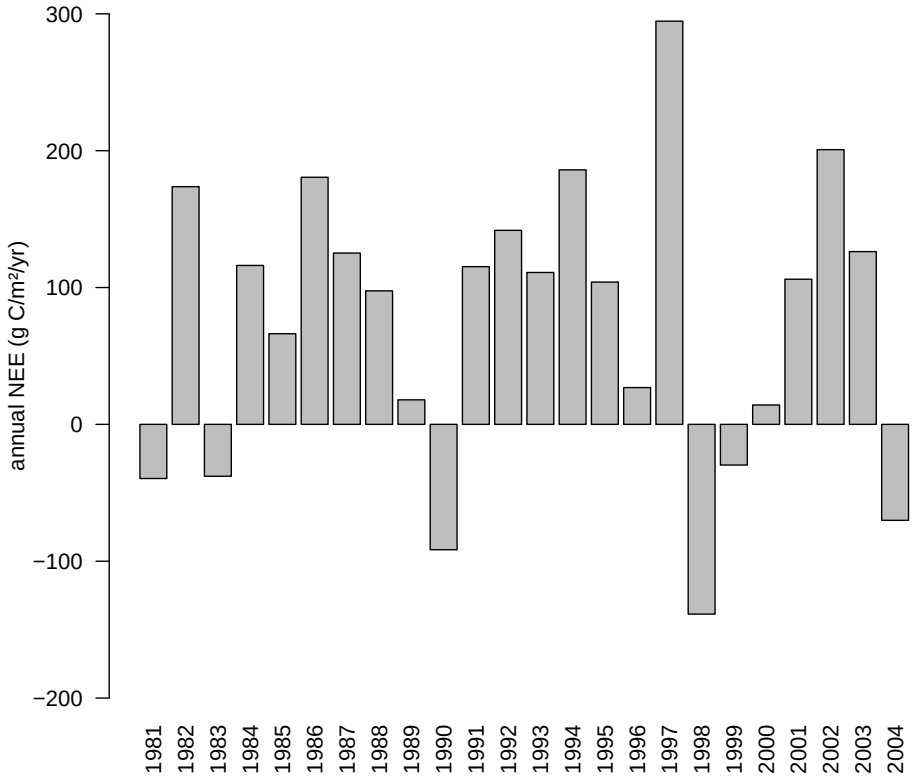


**Fig. 6.** Annual NEE estimated over a 25 year time-series at the Skukuza Flux site.





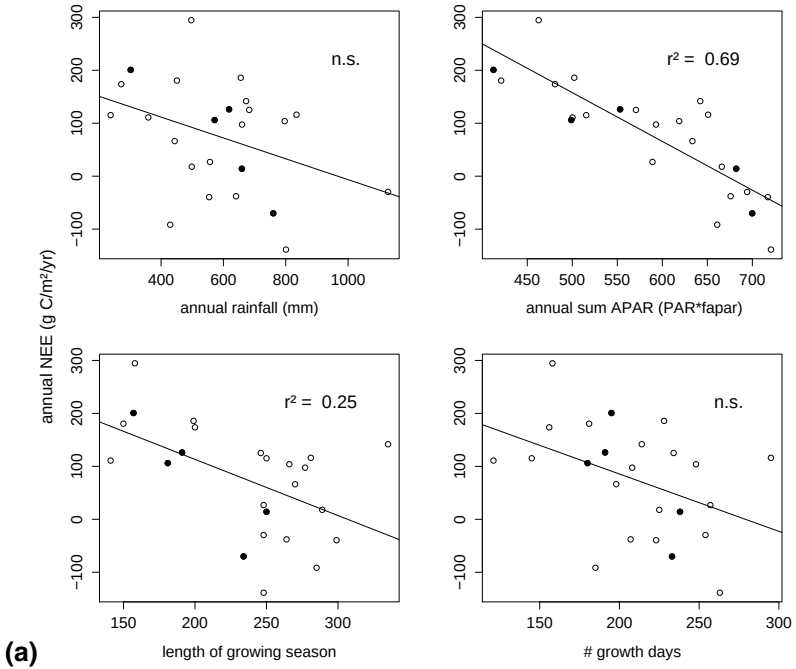


**Fig. 7.** Relationship between annual NEE **(a)**; $R_{eco}$ **(b)**, and GPP **(c)** and four potential drivers of inter-annual variability in carbon uptake: annual rainfall, available photosynthetically active radiation, length of the growing season, and number of growth days. Annual rainfall seems to be the least significant, compared with parameters that include seasonal variation in leaf display (APAR and length of growing season), and the seasonal distribution of rainfall. Solid circles represent years 2000–2005 for which flux data were available to constrain the model.

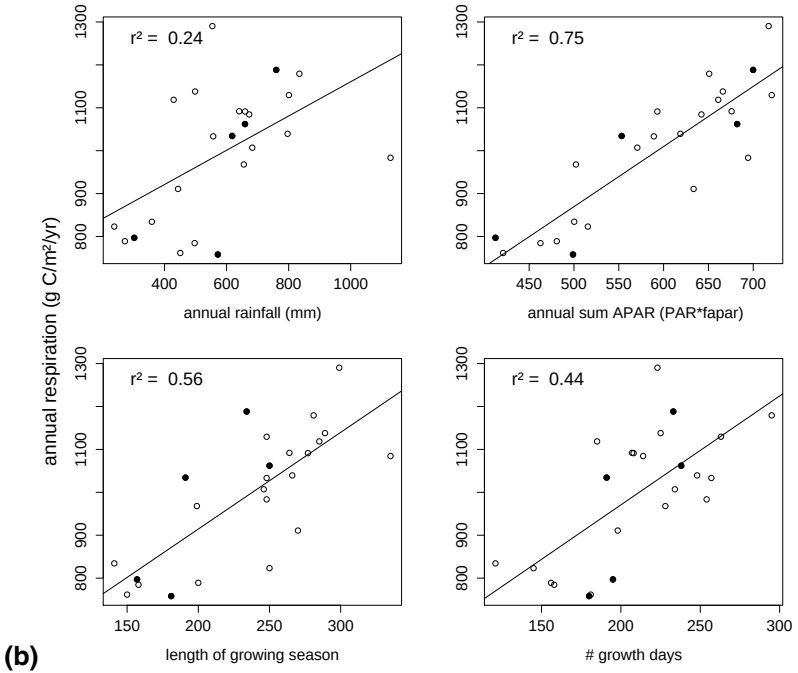


**Fig. 7.** Continued.

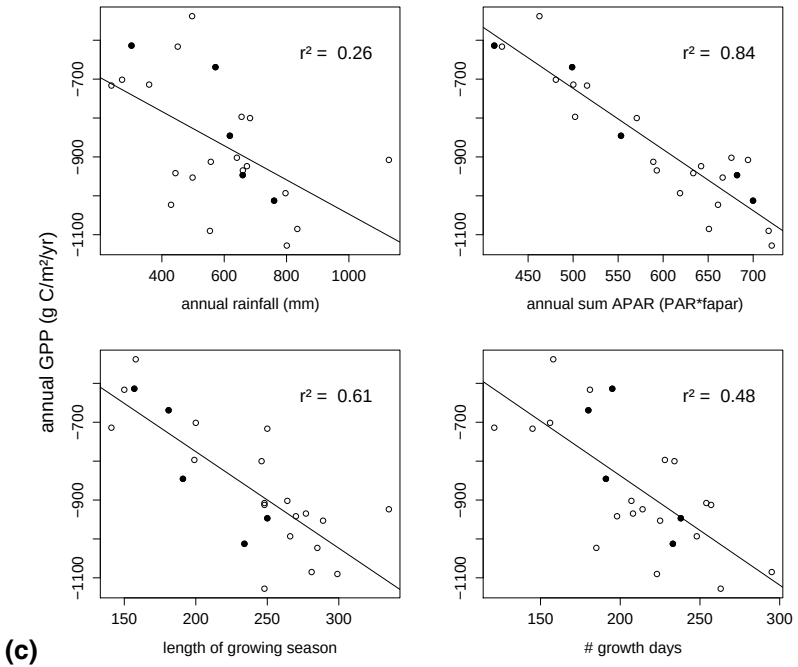


**Fig. 7.** Continued.



BGD
5, 3221–3266, 2008





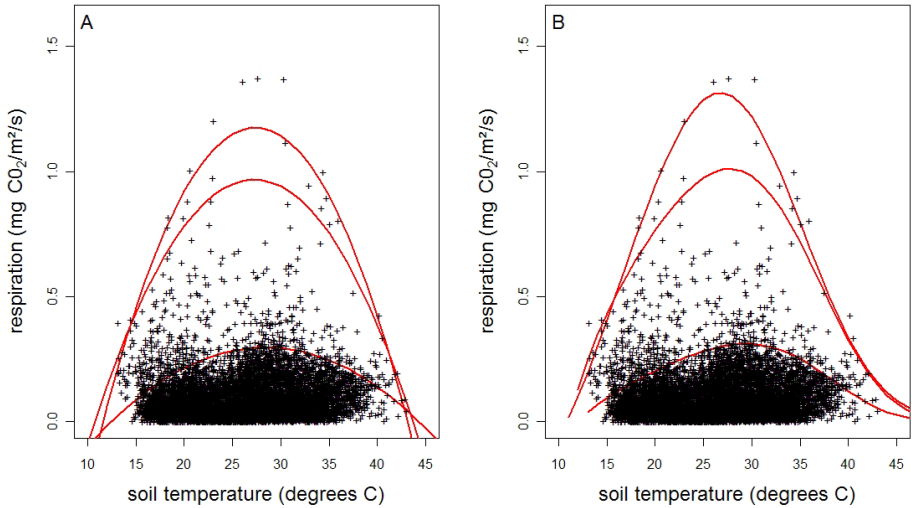


**Fig. B1.** Showing the six temperature response functions fitted to the half-hourly night time fluxes (respiration). Pabel **(A)** shows the parabolic functions fitted over the manually selected maximum points (top function), the automatically selected maximum points (middle function) and the manually selected top of the data mass (bottom function). Panel **(B)** shows the Generalised Poisson function fitted over the same three selection of points.

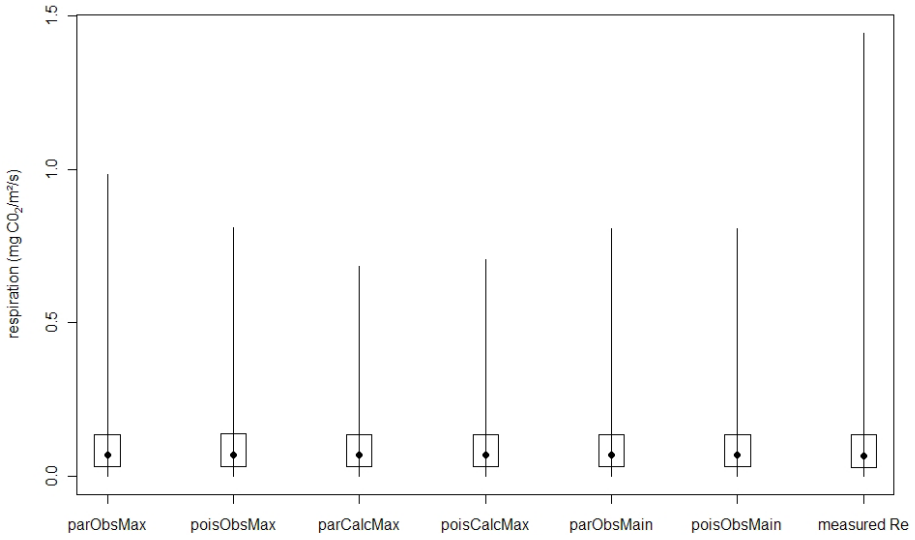


**Fig. B2.** Showing the distribution of the respiration data interpolated using six different methods (solid points: median values, box: ±25% quantiles, bar: data range). The median and ±25% quantiles are very similar for each method, but the method that calculates the fitted values had slightly lower maxima than the other two methods. All data are well within the range of measured Re values ($u^*$-corrected half-hourly night-time fluxes).



BGD
5, 3221–3266, 2008





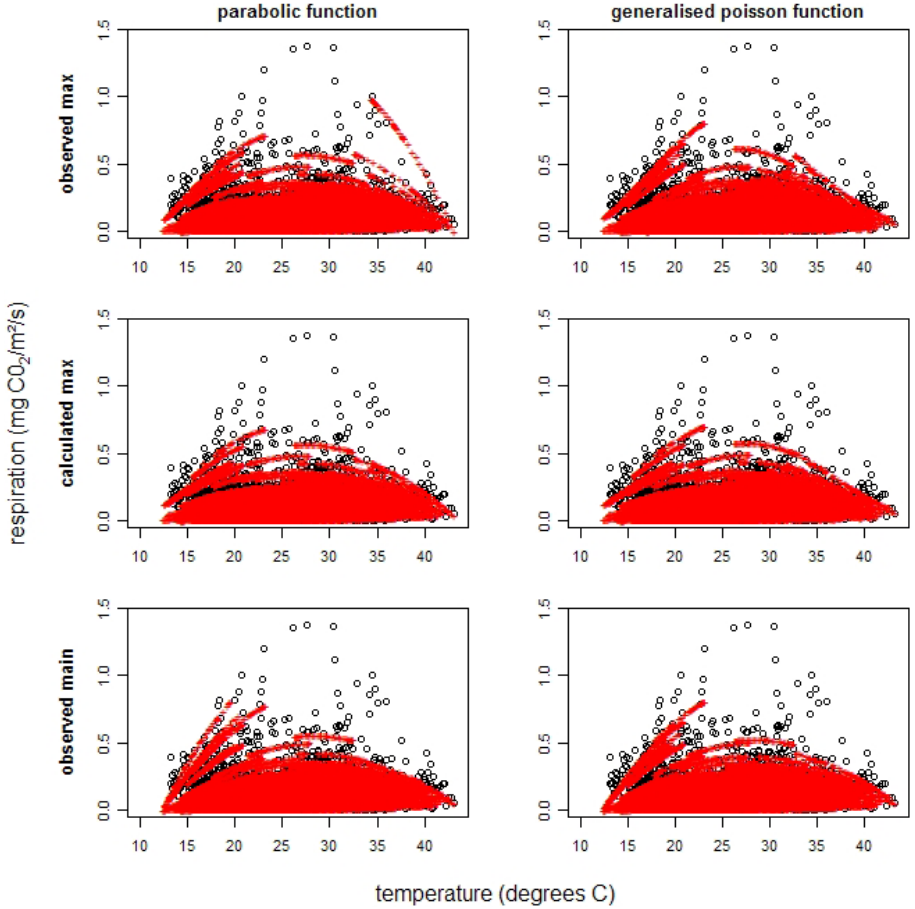


**Fig. B3.** The distribution of measured half-hourly night-time fluxes (black circles) and interpolated half-hourly respiration (red crosses) along a temperature axis. Interpolated fluxes represent all half-hour values which had soil temperature data and at least three night-time fluxes to estimate the scaling parameter.