Responses to Anonymous Referee #1 comments:

Thank you again for your thoughtful comments. We had to include some changes in the MS following comments from Referee #2, which affected our previous responses to your comments so we uploaded an updated version of our response. Our responses are indicated below in blue text.

Helman et al. identify a lack of ground-based ET/C-uptake measurements in water-limited environments as a motivation of their study. The authors test a biophysical approach based on a satellite derived estimates, with and without a seasonal drought stress factor, across paired forest and non-forest sites. Central to the method employed is the use of a drought stress factor.

1. In my eyes, it isn't really a drought stress modifier. Plants are responding not only to atmospheric demand, but also to supply limitations from the root-zone. The method used by the authors only considers atmospheric demand.

Au: The RS-Met model does not only consider atmospheric demand. Actually, it includes three main drivers: (1) radiation and temperature used to derive ET_o and PAR, (2) the f_{WA} and f_{WD} from rainfall data and calculated ET_o, which adjust for water supply conditions at the surface and the root zone, respectively, and (3) fVC and fAPAR (both based on satellite NDVI data) used as proxies for changes in vegetation cover and dynamics. Driver (1) expresses the atmospheric demand, while (2) expresses the water supply limitation and (3) the plant response to the changing conditions.

2. The one strength their approach does have is that it is cumulative over 2 months, so the effect is gradual. But why 2 months? Clearly the level of stress is sensitive to this assumption. I can't think of a physical justification for this decision. On a more fundamental level, this assumption overlooks the fact that vegetation in these water-limited regions would respond to water stress in a fundamentally different way to vegetation in more mesic regions. However, their modifier doesn't explicitly attempt to account for this in any fashion. This issue presents itself in various forms in the manuscript. The modifier appears to lack sensitivity at various key points (see comment below & fig 2).

Au: We have used a cumulative period of two months in the f_{WD} formulation because we think that this is the critical period for plants to respond to changes in water supply conditions at the root-zone in ecosystems in this region (Raz-Yaseef et al 2010; 2012). Actually, we did not get consistently better results when using a different single accumulation period for all sites, as we did with the 2-month period. We acknowledge that the model could be improved through calibrating the parameter values at the site level. However, the strength of the model in its present form is in its consistency (in terms of the used parameters) between the sites that allows the estimation of the ET and C- fluxes not only at the local scale but also at a wider spatial scale and in distinct ecosystems under different environmental and water supply conditions.

We explain this in the revised MS:

"The water availability factor (f_{WA}) is calculated as the simple ratio between the daily rainfall amount and the daily ET_o, both cumulated over a period of two months. Basically, the accumulation period could vary for different ecosystem types and environmental conditions. However, we have taken here a period of two months for the native shrublands and planted (and native) forests following previous observations that showed that this period is sufficient to maintain wet the topsoil layer for the whole rainy season in ecosystems in this region (Raz-Yaseef et al. 2010; 2012). Furthermore, changing the accumulation period did not gave us a consistently better results in all sites, as the two-month period gave us."

Yaseef, N. R., Yakir, D., Rotenberg, E., Schiller, G., & Cohen, S. (2010). Ecohydrology of a semi-arid forest: partitioning among water balance components and its implications for predicted precipitation changes. *Ecohydrology*, *3*(2), 143-154.

Raz-Yaseef, N., Yakir, D., Schiller, G., Cohen, S., 2012. Dynamics of evapotranspiration partitioning in a semi-arid forest as affected by temporal rainfall patterns. *Agric. For. Meteorol.* 157, 77–85. doi:10.1016/j.agrformet.2012.01.015

3. The authors highlight that they can't account for legacy effects and that they might need a "local" drought stress factor. That final point seems inconsistent with the method. How can the authors advocate for this method as an alternative to infill flux data, whilst simultaneously advocating for the need for a locally refined modifier?

Au: We agree with this comment and have deleted these lines from the MS because, as pointed out in the previous response, the good agreement between model and EC estimates despite of the use of a single cumulative period and similar Kc and RUE_{MAX} values for the different sites remarks the potential use of this model not only at the local scale but also at a wider spatial scale.

4. There is clearly some potential to the methods employed by the authors, but it is hard to grasp the extent of this. In part this is because a lot of their comparison is anchored in a comparison against the model without a stress factor. It is understandable why they take this approach, but there is an element of a straw man to it, meaning that we miss any interesting insight into how the new model performs. For example, in Fig. 3 if I look

carefully, the GPP model has a lot of day-to-day variability not captured at all by the model with drought stress. What drives this? Both the ET and GPP models over-estimate the fluxes, is this related the flat sensitivity to drought during J-A highlighted in Fig. 2?

Au: In general, we think that the model captures daily variability in the fluxes to a great extent. Of course, there are some discrepancies between model and EC estimates because of the many sources of error in model drivers (e.g. the meteorological data and NDVI from the satellite) and EC measurements. However, agreement between RS-Met and EC is in general quite good despite of these uncertainties (see also **Table 2**). The f_{WD} actually represents the water deficit conditions at the root-zone rather than drought stress conditions so it should adjust the model only when plant respond to a significant decrease in water supply from the root-zone, which is not captured by the NDVI. This is why it seems that f_{WD} has a flat sensitivity during the rainy season in Eshtaol (former **Fig. 2**), which maintain fairly wet conditions during this period. However, following this misunderstanding we have decided to change: (1) the name of this factor ('drought stress factor') to 'water deficit factor' along the MS (f_{WD}), and (2) former Fig. 2 and present the model and model drivers at the more xeric site of Yatir, which shows a higher sensitivity of the f_{WD} to changes in water supply (see revised **Fig. 2** below). Finally, we have added to the comparison the MOD16 ET and MOD17 GPP 1-km datasets (see comment below).

5. The other core weakness I see in this paper is a comparison to some other satellite derived product. How different are these results to using other satellite derived GPP/ET products? I feel like this is an avenue of exploration that would add considerable value to the paper. The comparison to a poor model without a drought modifier doesn't add much value in my eyes. But if the authors can demonstrate clear improvement over more widely used GPP/ET products using their drought modifier then I think that is publishable. It may offer insight and comment on the need to consider their drought modifier in other satellite products. Ultimately this is an area I'd suggest the authors consider in any revisions.

Au: We have included a comparison of our model with the MOD16 ET and MOD17 GPP 1-km datasets in the revised MS (please see revised **Figs. 4-9** added along and at the end of this document).

6. The authors explore WUE. I fail to see what insight this really brings? The products used rely upon the same input remote sensing data, so the measures aren't truly independent? I realise the authors aren't the first to take this approach, but it is worth reflecting if this really makes sense? Why not instead compare to a well-used satellite product? Or consider how they might adapt the drought modifier to reflect drought-adapted vegetation. The WUE bit seems tangential to the story.

Au: Former Fig. 7c (WUE) was removed from the main MS following this comment. However, we think this information is still relevant so we kept this figure in the Supplement of this MS (**Fig S5** in SM). We have decided to leave **Fig. 10** in the main MS due to the important topic of drylands afforestation and its particular interest in Israel.

7. Finally, I'm also struggling with PaVI-E as a validation metric? It is derived from satellite data as well, so is itself a product, one that isn't independent? Why not just stick to testing during the continuous flux period? There are also few details offered about what PaVI-E constitutes to the reader; this needs to be rectified.

Au: PaVI-E was included in the comparison because it has been validated in this region at the basin scale and has similar spatial resolution as RS-Met (250 m; Helman et al. 2015). It is true that it is a satellite-derived product but as already stated in the MS, it does not use meteorological information and thus provides an interesting comparison with ET from RS-Met. There is not enough flux information to derive annual EC estimates in the six sites (campaign-based sites) so using PaVI-E is intended to provide an additional assessment of the ET estimated from RS-Met at the annual scale. We compared our model with the MOD16 ET and MOD17 GPP 1-km datasets in the revised MS, following this and prior comments, which resulted in the following revised **Fig. 9**:

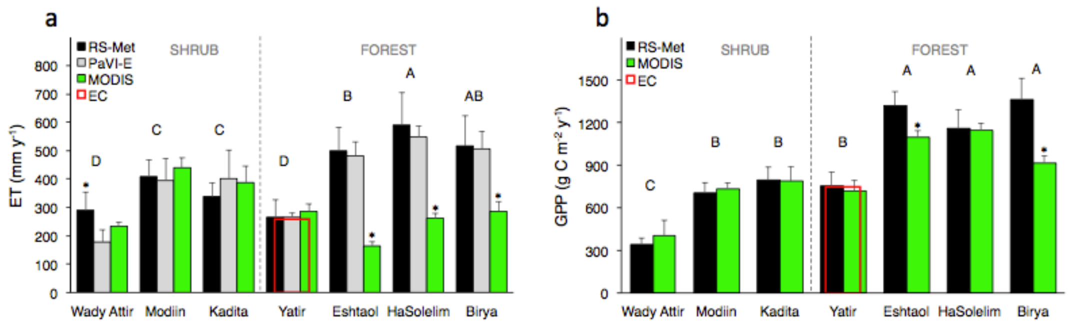

Fig. 9. Mean annual (2003-2013) estimates of ET **(a)** and GPP **(b)** from RS-Met (black), MODIS (green) and PaVI-E (grey; only for ET in **a**; Helman *et al.*, 2015a) at the seven sites. *Uppercase* letters indicate significant differences at $P<0.05$ between sites from Tukey HSD separation procedure following two-way ANOVA for the interaction of site × model (using PaVI-E and RS-Met in **a** and MODIS and RS-Met in **b**). Asterisks indicate significant different values from other models for the specific site, as indicated by Tukey HSD. The EC annual sums at Yatir are also shown (red).

We also included correlations of EC vs. MOD16 ET and MOD17 GPP using 8-day mean estimates of ET and GPP (i.e. the temporal resolution of MODIS products) in the following revised **Fig. 8**:

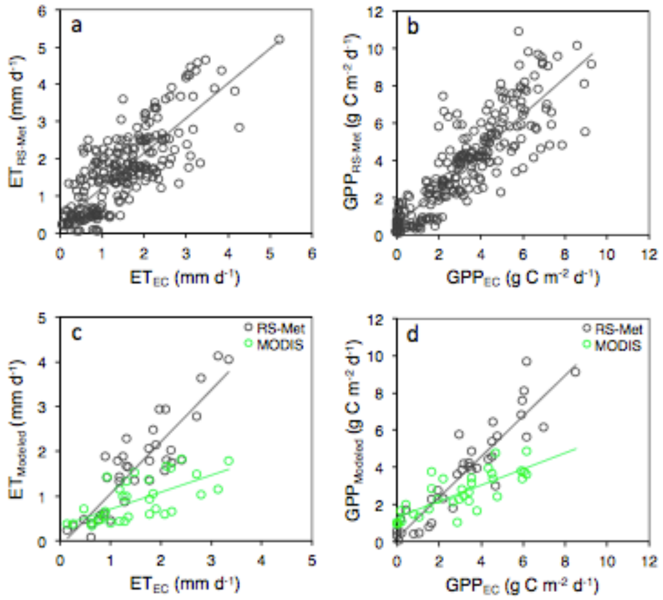

Fig. 8. Cross-site EC *vs.* model correlations of ET **(a, c)** and GPP **(b, d)**. In **(a)** and **(b)** are the EC *vs.* RS-Met (with f_{WD}) using all EC data from the six sites (each dot representing a single date), with linear fits of $ET_{RS\text{-}Met} = 0.936\ ET_{EC} + 0.281$ (R = 0.82; *P*<0.0001; N = 243) and $GPP_{RS\text{-}Met} = 0.990\ GPP_{EC} + 0.515$ (R = 0.86; *P*<0.0001; N = 252) for ET and GPP, respectively. In **(c)** and **(d)** are the same cross-site correlations but with data averaged over 8-day periods for comparisons with MODIS ET and GPP products (8-day averaged values). Linear fits for EC *vs.* RS-Met and MODIS in **(c)** are $ET_{RS\text{-}Met} = 1.16\ ET_{EC} - 0.11$ (R = 0.88; *P*<0.0001; N = 36) and $ET_{MODIS} = 0.38\ ET_{EC} + 0.33$ (R = 0.65; *P*<0.0001; N = 33), respectively. In **(d)**, linear fits are $GPP_{RS\text{-}Met} = 1.09\ GPP_{EC} + 0.21$ (R = 0.92; *P*<0.0001; N = 36) and $GPP_{MODIS} = 0.43\ GPP_{EC} + 1.31$ (R = 0.77; *P*<0.0001; N = 33) for EC *vs.* RS-Met and MODIS, respectively.

We added the necessary information on the PaVI-E model as suggested (L323-339):

"We also compared the RS-Met ET estimates to the annual ET derived from PaVI-E (Parameterization of Vegetation Index for the estimation of ET model; Helman et al., 2015a), at the six sites on an annual basis. The PaVI-E is an empirical model based on simple exponential relationships found between MODIS-derived EVI (and NDVI) and annual ET estimates from EC in 16 Fluxnet sites, comprising a wide range of plant functional types across Mediterranean-climate regions. This simple relationship (PaVI-E) was shown to produce accurate ET estimates at the annual timescale (mm y^{-1}) and at a

moderate spatial resolution of 250 m in this region (Helman et al., 2015a). It was validated against physical-based models (MOD16 and MSG LSA-SAF ETa) and ET calculated from water balances across the same study area. PaVI-E was used for ecohydrological studies in this region, providing insights into the role of climate in altering forest water and carbon cycles (Helman et al., 2017, 2016). The advantage of this model, is that it does not requires any additional meteorological information but is a proper function of the relationship between observed fluxes and satellite-derived vegetation indices. This makes it interesting to compare with the RS-Met model since the RS-Met is highly dependent on meteorological forcing."

Minor stuff ———– -

8. Line 83: The argument that remote sensing estimates of ET and GPP are too complex for other communities seems a strange one. Certainly, it ought to be substantiated. Given these data are provided as "products", should another community wish to use them I fail to see the complexity?

Au: We meant that the majority of the published models are not provided as "products" and that the basic algorithms of these models are too complex to be reproduced by the non-remote sensing community. We now mention MODIS ET and GPP products as accessible products in the revised text:

"Many RS models for the estimation of ET and GPP exist (see review in Kalma et al., 2008), but these algorithms are too complex and most of the models are not provided as accessible products for researchers outside the remote sensing community. Particular exceptions are the satellite-borne Moderate Resolution Imaging Spectroradiometer (MODIS)-based ET and GPP products (MOD16 and MOD17), which provide 8-day ET and GPP estimates at 1-km for 2000-2015, globally (Mu et al., 2007, 2011, Running et al., 2000, 2004)."

9. Fig 2. Perhaps try different colours? It is really hard to tell the lines apart. There is also essentially no sensitivity of the drought stress between J and A. Does that seem realistic? The authors argue the fluxes are reduced during drought to more realistic values, but I can't really conclude that from looking at the graph.

Au: We changed the colors in **Fig. 2** and revised this figure. It now shows model parameters and EC data at the xeric site of Yatir. The drought stress factor, which is actually a factor that reduces ET when water is less available at the root zone (hereafter changed to 'water deficit factor') should remain low during the summer and increase following the rainy season to adjust the high and low ET_o during summer and spring, respectively. The function of this factor is more pronouncedly observed in the case of Yatir:

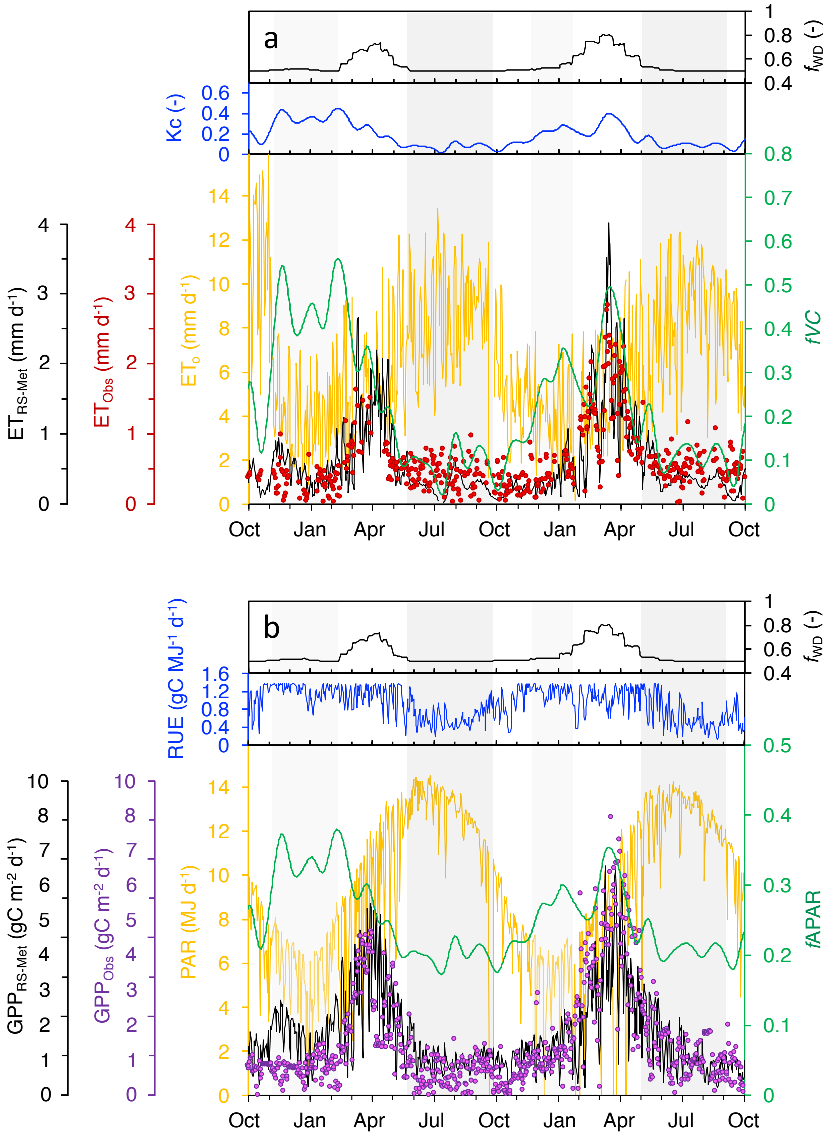

Fig. 2. Seasonal evolution of the water deficit factor (f_{WD}; black line in upper panel) and the main drivers of the modeled ET **(a)** and GPP **(b)** at the semiarid pine forest of Yatir ($ET_{RS\text{-}Met}$ and $GPP_{RS\text{-}Met}$, respectively; black line in lower panel) for the seasonal years 2008/9 and 2009/10. EC fluxes are also shown (ET_{Obs} and GPP_{Obs}, red and purple dots, respectively). The K_C and the radiation use efficiency (RUE) both without the addition of the f_{WD} (blue in middle panels) are shown together with the potential ET (ET_o; yellow in **a**), the fraction of vegetation cover (*fVC*; green in **a**), the photosynthetic active radiation (PAR; yellow in **b**), and the fraction of absorbed PAR (*f*APAR; green in **b**). Colored vertical bands indicate the critical periods when the addition of f_{WD} is particularly useful.

10. The caption and labels could be improved. It is also unclear what the "obs" are here? Are there any obs? How is the reader meant to come to a conclusion on "realism"? - All figure captions, labels could be improved for clarity.

Au: Please see our response to previous comment.

11. Line 583: The authors advocate that the water factor should account for past stores. No information is offered as to how they propose this would work in a practical sense. I don't see any obvious way this could be integrated with the scheme they've used here.

Au: Theoretically, this could be done by summing available water (calculated as P – ET) from previous year with the two-month summed P in the f_{WD}. Of course, this would be applied only after the first year of the ET estimation. We added this information in the text (L589-594):

"Thus, the 'transfer' of surplus rainwater from previous years should be also taken into consideration when adjusting the model with available water through the f_{WA} and f_{WD}, which are currently calculated only with the seasonal rainfall. Theoretically, this could be done by summing the available water from the previous year (calculated as P – ET) to the two-month summed P in the calculation of the f_{WA} and f_{WD}. Of course, this would be applied only after completing the ET estimation of the first year."

Revised Figs. 4-7:

Figure 4:

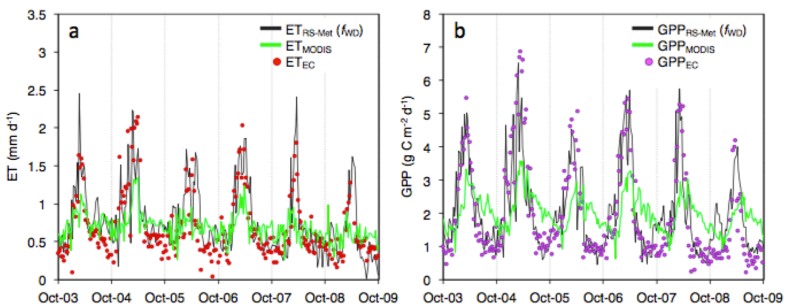

Fig. 4. Eight-day averaged values of ET **(a)** and GPP **(b)** from EC (dots), RS-Met (black) and MODIS (green) at Yatir. The R of the correlation for EC *vs.* RS-Met is 0.78 and 0.80 for ET and GPP, respectively (slope = 0.90 and 0.70; intercept = 0.19 and 0.66 for ET and GPP, respectively). The R of the correlation for EC *vs.* MODIS is 0.47 and 0.60 for ET and GPP, respectively (slope = 0.21 and 0.27; intercept = 0.56 and 1.47 for ET and GPP, respectively).

Figure 5:

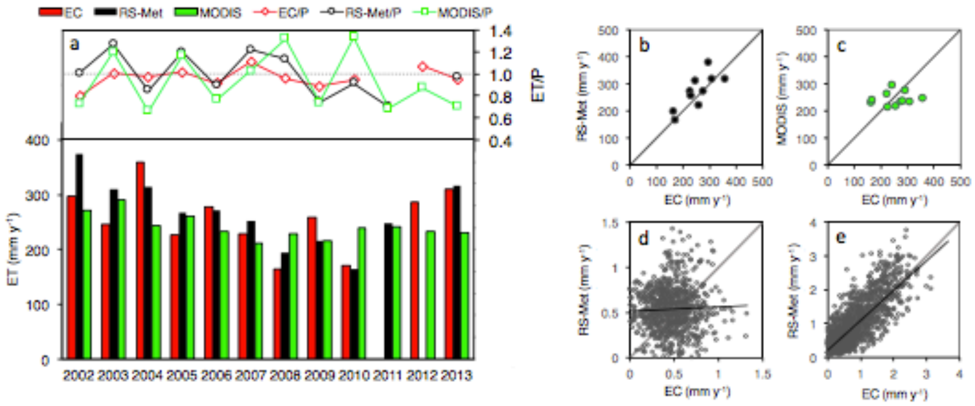

Fig. 5. Annual ET (mm y^{-1}) summed from daily RS-Met (with f_{WD}; black), MODIS (green) and EC (red) at Yatir forest site for 2003-2014 **(a)**. Linear regressions of the EC *vs.* RS-Met **(b)** and EC *vs.* MODIS **(c)**. Daily estimates from RS-Met in dry summer (June-August; **d)** and rainy (October-May; **e)** seasons. The R's of the linear fits for EC *vs.* RS-Met **(b)** and MODIS **(c)** are 0.78 (P<0.05; N=10) and 0.10 (P>0.1; N=11), respectively. The R's of the linear fits for the daily data in **(d)** and **(e)** are 0.05 (P>0.1; N=876) and 0.80 (P<0.0001; N=1570), respectively. The interannual trends in ET/P from EC, RS-Met and MODIS are presented in the upper panel of **(a)**. Note that the annual sums of ET from EC and RS-Met in 2012 and 2013, respectively, are not displayed due to the scarcity of available data during these years (>50% missing data).

Figure 6:

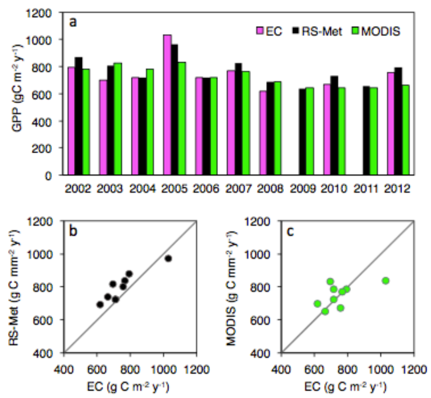

Fig. 6. Annual GPP sums (g C m^{-2} y^{-1}) from EC, RS-Met (with f_{WD}) and the 8-day MOD16 product (MODIS) at Yatir **(a);** and the linear regressions of EC vs. RS-Met **(b)** and MODIS **(c)**. The R of the linear fits is 0.91 (P<0.05; N=10) and 0.58 (P=0.08; N=10) for RS-Met and MODIS, respectively. Annual EC GPP for 2009 and 2011 were not calculated due to missing data.

Figure 7:

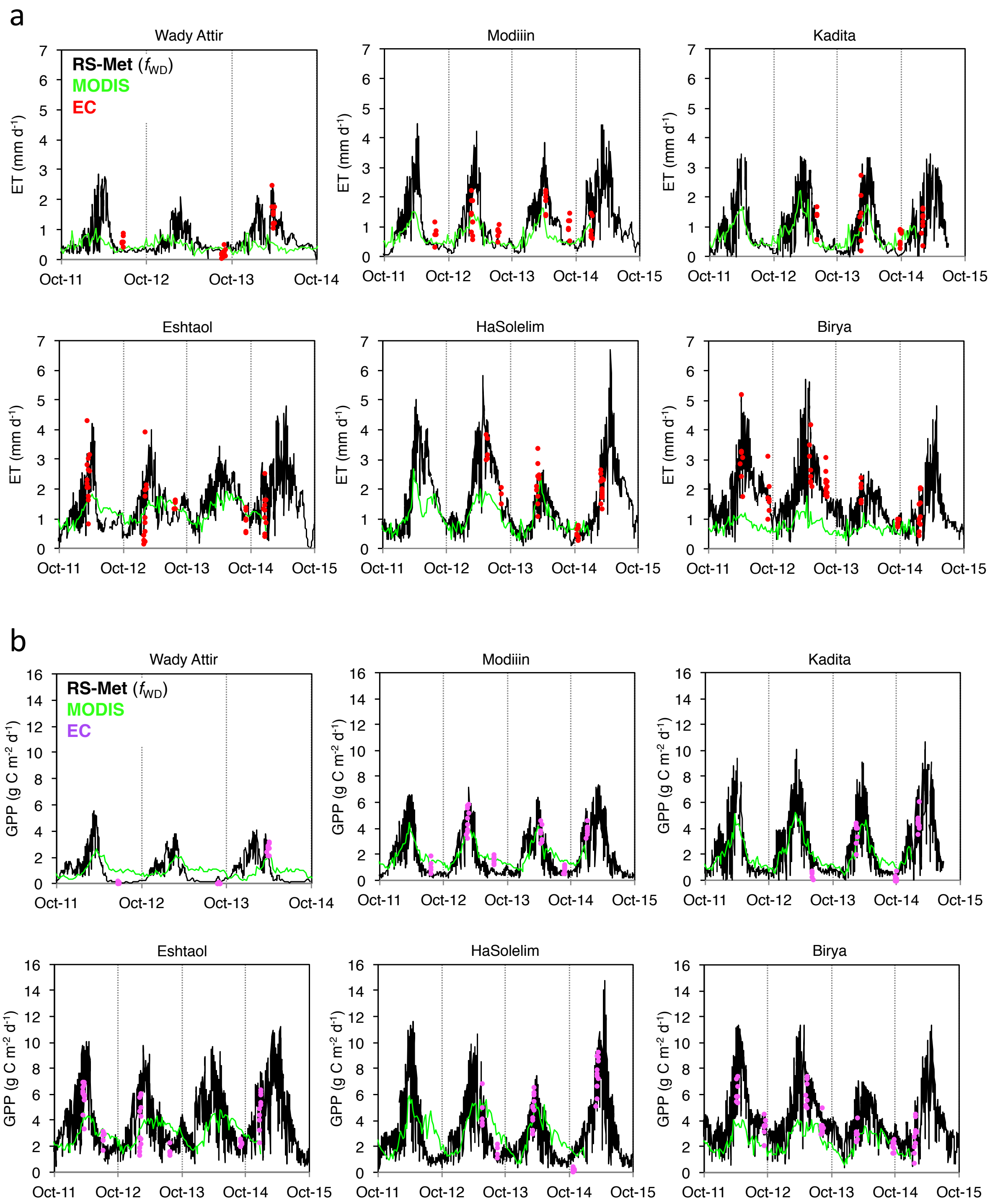

Fig. 7. ET **(a)** and GPP **(b)** from EC, RS-Met (with f_{WD}) and MODIS (MOD16A2 and MOD17A2 products, respectively) at the six forest and non-forested sites.