We thank **Anonymous Reviewer (#1)** for his insightful comments. Following the Editors' comment (at the initial stage), we also addressed here the suggestion posted at the initial evaluation of the manuscript by **reviewer #1** (comment **1**).

Reviewer #1 comments:

1) An interesting study showing that empirical modeling for ET is as good as more physically based modeling. This is fine to say the empirical way is 'good enough'. However, it would be better to say something about the generality of the PAVIE approach and commentary on the way forward for better ET estimation accuracy. The impression the paper leaves me is that we have reached the practical limits of accuracy for estimating ET from satellites, so it would be helpful for authors to state their opinion on this issue explicitly.

We thank you for this suggestion. Regarding the limits of accuracy for estimating ET from satellites we cannot say much because it was out of the scope and goals of this paper. However, what we can say is that VIs could not be used for estimating seasonal ET at complex Mediterranean vegetation systems comprising both annual (ephemeral) and perennial (mostly evergreen) vegetation. This is primarily because those two vegetation components have distinct phenology in Mediterranean systems (Helman et al., 2015). Annual and perennial vegetation usually constitute two different vertical layers in those systems and satellites only detect their combined signal. This is illustrated in Figure 1 and was explained in the revised MS:

Lines 3-12(P06):

Perennial and annual vegetation in Mediterranean regions have distinct phenology contributing differently to the VIs signal (Helman et al., 2015; Karnieli et al., 2003; Lu et al., 2003). Here we examined VIs - ET relationships in vegetation systems comprising both annual and perennial vegetation (i.e., forests, woodlands, savannah and shrublands, hereafter PA) separately from those comprising only annual vegetation (i.e., croplands and grasslands, hereafter AN).

We found that annual vegetation in the understory of PA systems might contribute significantly to VIs while having very small contribution to the total ecosystem ET. In some cases this results in an apparent phase shift between ET and VIs (Fig. 1) leading to negative or a lack of correlations."

And **Lines 27-31(P13)**:

"Intra-annual relationships were poor probably due to the mixed VI signal contributed by annual and perennial vegetation that constitute different vertical layers (Helman et al., 2015). While the annual vegetation (in the understory) was the main contributor to the intra-annual VI change, it was only a minor contributor to the total ecosystem ET in complex Mediterranean systems."

2) The scientific value of this work is not significant due to its statistical, locally calibrated approach. A more productive research avenue would be to develop greater, not less, modeling sophistication, and to find better ways to combine all lines of evidence into the ET estimation process. The currently available data sets, even confining oneself to remote sensing data alone, are rich; so it is hard to justify approaches that avoid using such data.

We acknowledge that developing a more general model for estimating ET at a global scale would be of a greater value. However, such task is out of the scope of this paper, though it is a possible future direction. Nevertheless, the importance of developing regional scale models in the EM should not be underestimated. Chen et al. 2014 is a good example for the importance of developing models at a regional scale presenting a model for China (published in ACP and cited in this MS). Moreover, ET models were hardly proposed or tested and validated at the EM. Developing remote sensing ET models without using ancillary data (except for the initial calibration using the EC data) is important for this region that lacks the required density of weather stations to assess ET at a high to moderate spatial resolution (250-1000m). Providing an ET product at such spatial resolution (250 m) is essential for the scientific community and forest and water resource managers of this region. We added this to the title:

"Annual evapotranspiration retrieved from satellites' vegetation indices for the Eastern Mediterranean at 250 m spatial resolution"

And the following lines to the revised MS:

Lines 28(P01) – 2(P02):

"In the lack of high-resolution (<1 km) ET models for the EM the proposed model is expected to contribute to the hydrological study of this region assisting in water resource management, which is one of the most valuable resources of this region."

And **Lines 6-7(P14)**:

"PaVI-E is the first ET model with such high-resolution (250 m) for the Eastern Mediterranean region."

While the need for simplicity can be important for common adoption of the method by others, the practitioners of ET science are a sophisticated group and can handle more complexity. Considering the ground work done with simplified approaches using vegetation indices and temperature more than 30(!) years ago by people such as Carlson, Price, Seguin, Gurney…plus many others I don't mention… I am wondering how proposing such a simple approach could be considered tenable. Has the world not progressed with ET research much since then? My opinion is that this work goes in the wrong direction.

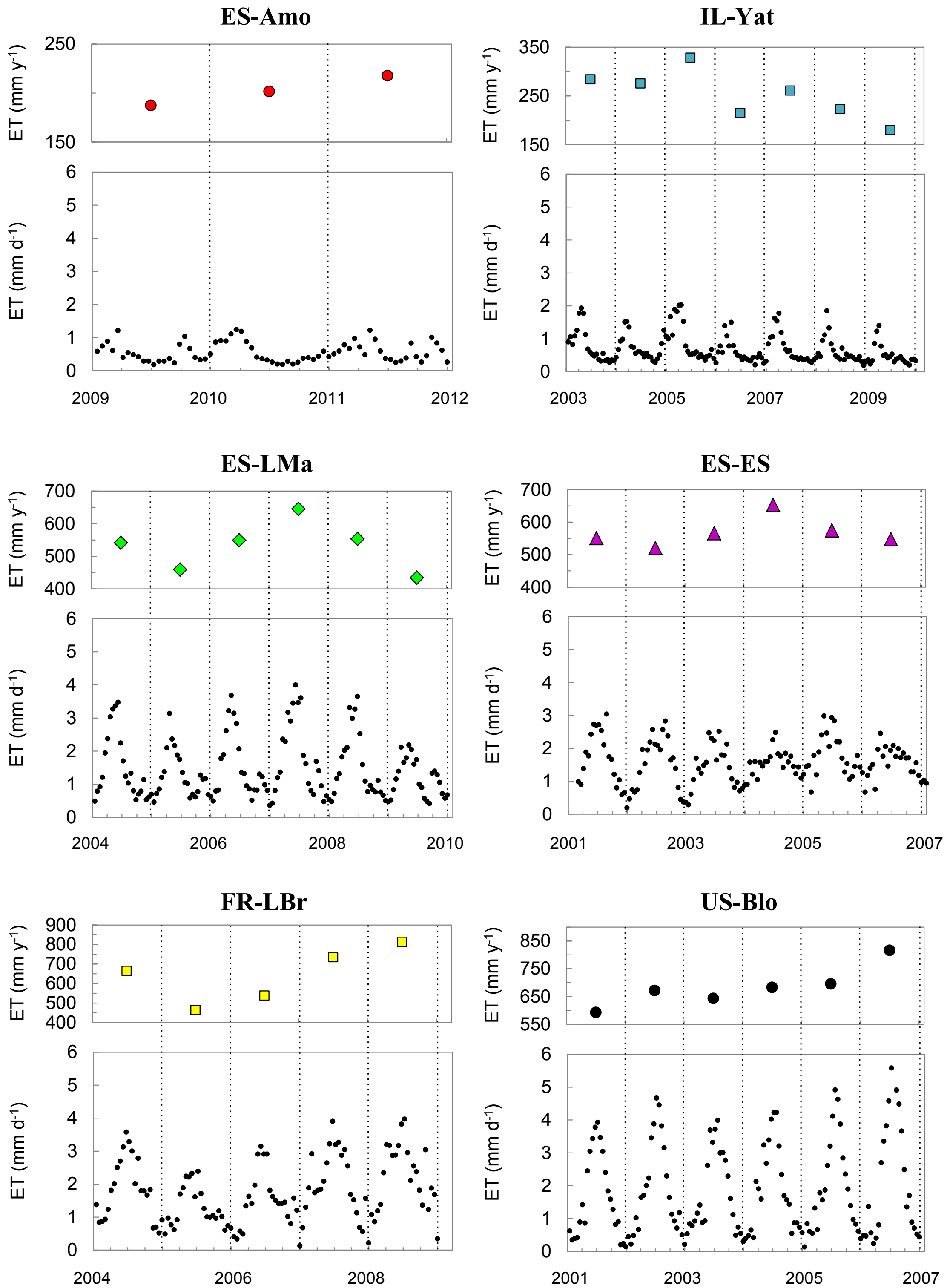
The practitioners of the remote sensing ET science are certainly a sophisticated group that can handle complex models and multiple data sources. However, from our experience, water resource and forest managers do not implement much of the models proposed by the remote sensing community because of their complexity. Moreover, some models report a limited accuracy at the sites they were tested and do not guarantee a better accuracy at the site to be implemented. Those models also have often too coarse spatial resolution for local to regional scale studies. In this paper, we propose a simpler model at relatively high spatial resolution (250m) that could be easily implemented by the scientific community and stakeholders at the EM with an acceptable accuracy, similar to (or better than) the proposed by more complex models. We see no reason to introduce additional data that would complicate the model while its current accuracy is within that of the EC data used for its calibration/validation. Glenn et al. (2010) in their review of existing remote sensing-based ET models already concluded that: *“The methods…range from the very simple (e.g. Groeneveld and Baugh 2007) to the more complex (e.g., Fisher et al. 2008), without a concomitant increase in accuracy.”* They also stated that: *“One reason that simple models work as well as multi- factor models is that plant productivity (and transpiration) tend to be limited by a single factor at a given time and place (Liebig’s law of the minimum) (Paris 1992).”*

The excellent work presented by Carlson, Price, Gillies, Smith, Kustas and more… founded the principles for using VIs and LST in ET (and soil moisture) assessment from remote sensing. However, they all used air temperature data from weather stations (or models) and a SVAT model to derive ET. The “triangle method” that does not use meteorological data is based on the simple relationship between VIs, LST and ET (or soil moisture), which are not straightforward at complex Mediterranean vegetation systems as illustrated by **Figure 1** (see also results from the 16-day regressions in **Tables 4-5**). The aim of our empirical approach was to test directly VIs/LST-ET relationships and examine the feasibility of using those simple relationships to estimate actual ET. Results imply that satellite-derived VIs could not be used for seasonal estimation of ET in complex Mediterranean systems due to the mixed contribution of annual and perennial vegetation to the VIs signal. However, ET could be estimated on an annual basis in those systems. This is an important contribution of this paper. As far as we know this is the first study that derives annual ET at such spatial resolution (250m) and accuracy (70-90% and MAE of 12%) from satellite data in the EM. Such annual ET product is essential for studying inter-annual changes in ET at the EM.

3) More details are needed to show readers statistical aspects of the 15-year data set: seasonal and annual variability of ET at each of the flux sites would help.

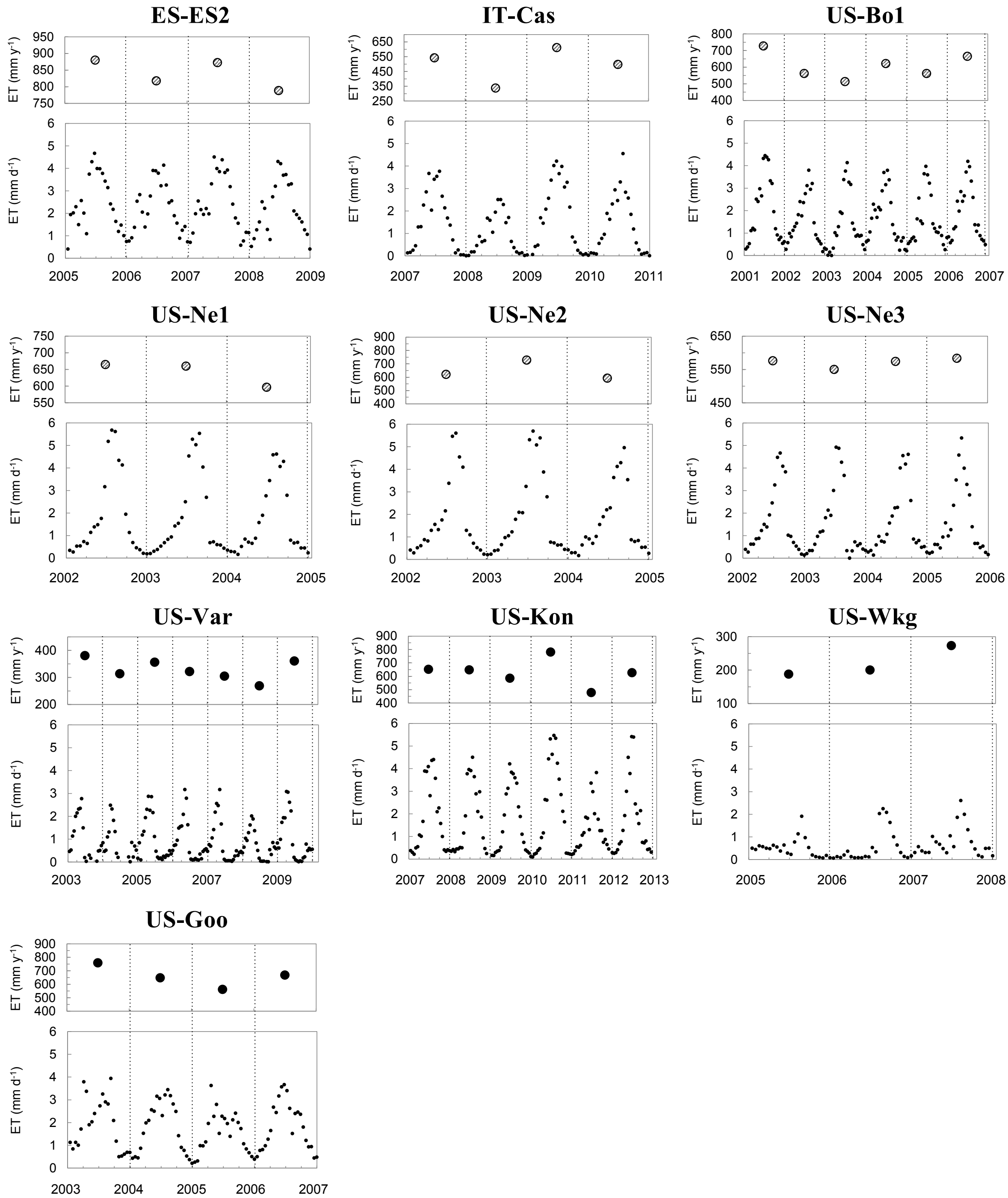
We added seasonal and interannual time series of the eddy covariance ET for each one of the 16 FLUXNET sites. However, to keep the flow of this paper we added this in the Supplementary Information (new **Figures S3** and **S4** in the revised SI).

Figure S3:



ES-Amo
ET (mm y-1)
250
150
ET (mm d-1)
6
5
4
3
2
1
0
2009
2010
2011
2012
IL-Yat
ET (mm y-1)
350
250
150
ET (mm d-1)
2003
2005
2007
2009
ES-LMa
ET (mm y-1)
700
600
500
400
ET (mm d-1)
2004
2006
2008
2010
ES-ES
ET (mm y-1)
700
600
500
400
ET (mm d-1)
2001
2003
2005
2007
FR-LBr
ET (mm y-1)
900
800
700
600
500
400
ET (mm d-1)
2004
2006
2008
US-Blo
ET (mm y-1)
850
750
650
550
ET (mm d-1)
2001
2003
2005
2007

and **Figure S4**:



4) What can you say about the accuracy of ET from the flux sites?

Following Glenn et al. (2010) and Kalma et al. (2008) that reviewed previous works that used the same data we used here, we can say that in general the accuracy of the ET flux data is ~70-90% (**Lines 18-20 P02** and **28-30 P09**).

5) How can you verify your confidence in your regression equations?

We dedicated two sections to validate and compare the results from our regression function models. First, a comparison with ET products from MODIS (based on Penman-Monteith) and MSG (based on H-TESSEL SVAT) for one year was conducted. Second, we validated our results using our mean annual estimates (2000-2013) with those calculated from water balances at the EM. Results showed good agreement in both cases. This is the best we can do with the scarcity of ET data for this region.

6) How representative is 2011 for comparing your empirical ET against MODIS and MSG ET estimates.

The ET product from MSG is only available for 2011-2013. We used one of those years for the comparison. PaVI-E was used for the same year, so it should be representative at least for that specific year.

7) Your statistical analyses, when considered as cover classes, do not seem strong with only ~7 samples per regression.

Although the number of per-site samples is relatively small (one sample per year), most relationships were statistically significant at $p<0.05$. The aim of the per-site comparison was to show possible relationships between VIs/LST and ET at annual time scales. Following this comment we indicated which correlations were statistically significant in **Table 3**. Cross-sites correlations were all statistically significant at $p<0.001$.

8) A lesser concern is your choice of R vs. R^2 for statistics; of course using R^2 values will decrease the apparent strength of your regressions.

The coefficient of determination (R^2) gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable, while R measures the strength and also the direction of the relationship. We used R to indicate strength and type of relationships (positive or negative) but referred to the percent of the variance in ET explained by VIs (R^2) in the text (**Lines 27-28 P09**).

9) No improvement of results at annual scales using LST data is not especially significant: VI data represent the long term vegetation patterns while LST data excel in identifying shorter term water stress events.

The 16-day LST reflect actual soil moisture conditions while mean annual LST is more representative of the water availability for ET in a specific site at an annual timescale. Good examples are the IL-Yat and ES-LMa sites, which had strong and significant (p<0.05) negative ET correlations with annual LST. In those sites, LST did improve correlations at the annual timescale for the multiple regression models (with EVI).