Review of:
Boese et al., Carbon-water flux coupling under progressive drought

The authors analyse dry-down periods at 31 flux tower sites to evaluate semi-empirical water use efficiency models. The authors present interesting concepts for analysing the effects of water stress on water use efficiency and separating mechanisms controlling vegetation drought responses. Whilst I appreciate the authors' efforts to understand deficiencies in the WUE models, the manuscript currently offers little process-level understanding, rather focusing on metrics and locally calibrated empirical models. It would be great to see the authors expand on their findings. For example:

1) The authors conclude that the WUE models need to consider radiation and soil water limitation of transpiration to better capture WUE (ET) changes during water stress, in addition to GPP and VPD effects. Yet, the authors offer very little in the way of explanation for why these limitations are important. What are the specific mechanisms? The other points in the discussion are rather obvious (short vegetation is generally more drought-prone and seasonal climates are more resistant to drought stress) but the main conclusion is spared little attention.

2) It would also be useful to unpack the results for "tall" vs "short" vegetation further. Whilst I acknowledge the limited metadata available for flux tower sites, the authors have potentially missed an opportunity to identify which ecosystems responds to water stress more strongly. For example, are sites with specific vegetation types, climates (hot, cold, wet, dry…?), high/low leaf area more responsive? Similarly, are the additional Rg and SWL terms more important in specific environments? "Tall" and "short" vegetation seems a rather simplistic classification to understand how specific sites responds to water stress and the broad vegetation classes are not particularly informative (for example, is savanna really "tall" vegetation or mainly grasses?).

3) I would question the meaning of the S_{rem} variable. This measure doesn't account for antecedent conditions, which could play a large role in determining the rate of dry-down, especially during the short-term droughts analysed here. It is thus unclear what specifically can be learnt from the inclusion of this term? The authors might be able to test the sensitivity of their results to the assumption that antecedent conditions are negligible by e.g. calculating rainfall accumulation prior to the dry-down. I also note much of the manuscript discusses water use efficiency, yet all the equations and results are for ET?

Overall, the authors have provided a comprehensive and well-written, but rather superficial analysis of ET responses to water stress. It is not clear how the results can be used more widely to gain mechanistic understanding of ecosystem functioning under water stress, or improve the formulation of these processes in models, as they rely primarily on locally-calibrated statistical models. I would encourage the authors to dedicate less space on metrics and calibration schemes (some of this might be better suited to the supplementary?), and unpack their findings further.

Specific comments:

P1 L19: prime-sources should be primary sources?

P1 L22: interacting rather than interlocking?

P3 L3: Please correct spelling to La Thuile

P3 L8: Please specify what you mean by "the established methods"?

P3 L12: How did you define a precipitation event (> 0mm?)?

P3 L15: How did you handle observed vs. gap-filled data? If some of the dry-down periods were heavily gap-filled or missing, were these still analysed? If so, I would question what can be learnt from these sites as it seems unlikely the gap-filled data can accurately reflect fluxes during extreme conditions. Also how were the sites selected? On line L22 you mention 31 sites were used, but there are many more in the La Thuile release alone (of course not all with dry-downs). I'm surprised if there are only 47 dry-down events in the 200+ site records, but this is of course possible.

P3 L19: *Rn* not defined? Also, EF is normally defined as *Qle/(Qle+Qh)*, with the latter part equating to *Rn – G*

P3 L20: Because the definition of dry-down events is central to this manuscript, I would like to see more details presented here instead of the supplementary. Also suggest using "we" instead of "I" in the supplementary.

P3 L25: remove data-sets

P4 L2: Mass balance is also affected by output from runoff. Similarly in the following sentence, stored water can depend on subsurface runoff. It seems reasonable to assume these fluxes were small due to the lack of precipitation, but this should be mentioned

P4 L10: How many missing values did you allow for?

P4 L19: How was uWUE determined?

P5 Eq. 5: Should max(S_{remt}) be without the *t* subscript? Surely S_{remt} is a single value, at time *t*?

P5 L15: Compared instead of inverted?

P6 L2: Consider mentioning that 1.0 is the best possible MEF value for readers not familiar with the metric

P9 Eq. 14: What is ET_{rad}?

P9 L29: Remove "was that"

P15 L 6: Correct to "led"

All figures: Avoid analysing results in the figure captions (e.g. models underestimated…), the captions should merely explain what is shown in the figures.

Fig 6: It is hard to see the blue line especially