Response to interactive comment by M. G. De Kauwe (Referee) on “Water stress induced breakdown of carbon-water relations: indicators from diurnal FLUXNET patterns”

Nelson et al. present a novel exploration of FLUXNET data to derive two new data-driven drought metrics. I think their approach is very interesting and this could form part of a more nuanced way to benchmark land models in the future. Ultimately, whilst I feel that this paper should be published, I think the text requires quite a lot of clarification and redrafting.

I found the introduction & methods text a bit disjointed. There is very little in the way of text to explain or set the “why is drought a big problem” argument. To me, there was a strange focus on considering the proportion of transpiration to ET. The authors spoke a lot about uncertainty, but it wasn’t very clear if their focus was data (flux measurements) or model world. In short, the authors could do a better job of framing the broader problem before they get to the hydraulic and non-stomatal limitation text. In the methods, you often jump or introduce new concepts with little back story and it becomes quite distracting.

I also think the analysis of the results could be more incisive. I strongly feel the authors are doing themselves a disservice in terms of likely citations by not picking through their results a bit further (see comments).

Finally, whilst I shouldn’t have to, it still true that it isn’t the norm to share code, so I applaud the authors. I suspect it is likely to lead to their work being more widely used and potentially improved upon.

We would like to thank Dr. De Kauwe for this very in-depth and thorough review. As authors it is always welcome to get such thoughtful outside opinions to further strengthen and clarify a manuscript. We hope the revised manuscript has a much better flow which allows the reader to follow our line of thinking and limit the distractions, as well as the expanded analysis and discussion. Following are the detailed responses which we hope address these overall concerns.

Introduction ————

1. I don’t immediately follow why you’ve introduced VPD into equation 2? Surely your estimation of GPP and ET have both already accounted explicitly for a VPD dependence? Then on line Pg 1, line 19 you say “more consistent” … more consistent with what? This is probably simply my ignorance, but I’d like to follow the logic here because you then use VPD in eqn 7 and 8.

We agree with the reviewer that this concept may have been inadequately introduced. It may have been rather distracting to introduce both the uWUE and iWUE equations, and as only the uWUE was use in the calculations of the DWCI and the Katul model. As such, the iWUE references and equations have been eliminated. Furthermore, we have tried to highlight that the \( \sqrt{VPD} \) allows for a more stable metric which can be compared across timescales, which becomes important when comparing diurnal GPP and
ET in calculating the DWCI. Therefore we have rewritten this section, which also incorporates comment
3 from reviewer 3. This subsection now reads:

Classically, vegetation water and carbon fluxes are linked by stomates, where an open stomate allows CO2 to enter the leaf and, consequentially, water is lost. From this, most theoretical frameworks make some form of assumption that carbon assimilation (A) and water losses (T) are both contingent primarily on leaf stomatal conductance (gs). This assumed relationship allows us to pass between the realms of carbon and water, based on the assumption that at any given time both A and T are proportional to the stomatal conductance multiplied by the difference in internal and external CO2 and water vapor concentrations. More specifically,

\[ A = g_s \cdot \Delta c \quad \text{and} \quad T = 1.6 \cdot g_s \cdot \Delta v \]

where \(\Delta c\) and \(\Delta v\) are the differences in inner and outer stomatal cavity concentrations of CO2 and water vapor, respectively. These diffusion equations lead to the relatively consistent carbon:water ratio, generally expressed as a water use efficiency \(\left( WUE = A/T \right)\). At the ecosystem level where direct measurements of A and T are not available, WUE is simply calculated as the ratio of gross primary productivity (GPP) to total evapotranspiration (ET) \([7]\). These carbon:water links are fundamental to understanding how stomata are regulated and underly key functioning in mechanistic plant and ecosystem models. One such set of models are those based on optimality theory which posit that plants tend to optimize carbon gains to water losses, such as those described by Katul et al. \([5]\) and Katul, Palmroth, and Oren \([4]\). These concepts from Katul, which carry the assumptions of RuBISCO (light) limitation, were built upon by Zhou et al. \([14]\) and Zhou et al. \([13]\) to give the equation,

\[ uWUE = \frac{GPP \cdot \sqrt{VPD}}{ET} \]

where the \(\sqrt{VPD}\) accounts for the stomatal response to vapor pressure deficit (VPD). Accounting for the VPD response allows for a more stable metric of WUE that is temporally more stable and physiologically more meaningful, such as when comparing the diurnal cycles of carbon and water. As ET is the sum of both T and non-biological evaporation (e.g. soil and intercepted evaporation), often periods during and shortly after rain events are excluded from WUE estimates to minimize the influence of non-plant evaporation. Ultimately, calculations of WUE provide a simple summary of the cost in water per carbon gain and becomes an indicator for how plants have and will adapt to the physical limitations of their changing environments \([6, 11]\).
2. Pg 2, line 2: “propagate errors …” I assume you mean in terms of a model? As actually measured fluxes would account for any drought signature? Please clarify.

Indeed this wording was unnecessarily vague. Here we were referring to assumptions on carbon and water in the models which then affects carbon and water estimates. To clarify this point the text has been changed to:

This failure to capture the effects of drought is not only disconcerting as water limited conditions are when ecosystems are most at risk, but an incomplete framework tends to propagates errors and uncertainties from models into estimates of the water and carbon cycles.

3. Pg 2: The arguments about the uncertainty of T as a proportion of ET … do we really think that this the chief uncertainty here is drought? To me it feels like an odd framing of the argument simply because I wouldn’t expect water stress to dominate the water cycle and the uncertainty range quoted is large. I think this line of argument would be improved by simply talking about the need to understand the carbon and water cycles during water stress. I’d argue for removing all of this text.

This section came in part from the fact that this research was motivated from a transpiration angle, and also to highlight that gaps in understanding of drought responses can be reflected in both the water and carbon cycle communities. However, as other reviewers have also noted that the introduction could be shortened, and this section has maybe become less relevant, it has been removed.

Methods ———

4. I think it would be helpful to explain why PET was calculated. The text just jumps to we calculated PET...Also, there is a brief mention of why the approach was adopted, but it should be expanded upon. Similarly with the CSWI, you just suddenly jump to explaining it without any back story for the reader.

We agree with the reviewer that the reasoning for calculating these parameters was not made clear, instead we simply jumped into the calculations. The PET was used in calculating evaporative fraction, which gives a consistent measure of ecosystem dryness across sites with which to compare the metrics (basically a daily aridity index). This concept is now introduced in the text as:

In order to provide a consistent measure of ecosystem dryness that can be utilized across sites, the ratio of water evaporated to potential water evaporated was calculated as evaporative fraction (EF), or the fraction of actual ET to Potential ET (PET).

Likewise, the CSWI was used as a replacement to the time after rain methods for screening periods with wet surfaces and likely higher evaporation. We now mention this when introducing the concept:

In an effort to minimize contributions of evaporation, the conservative soil wetness index
(CSWI) was employed which was designed to estimate whether the ecosystem is likely to have "dry" surfaces and therefore ET is likely to be dominated by transpiration. This approach requires a certain amount of evaporation to occur after a rain event before the surface is considered to be "dry" and can be contrasted to the method of removing a set time period after rain employed in previous studies [9, 1, 6].

Hopefully by introducing these topics before jumping into the details of their calculation this section with flow better and not be so jarring.

5. The screening of data to remove contributions from the soil is potentially problematic. I’ve seen that other authors have used 48 time slots after rain (see Medlyn et al 2017, New Phytologist and references they cite). The authors have taken a different approach, but screening GPP < 5 g m-2 d-1 seems quite high? Presumably as you get a drought, GPP drops and this may remove some of the signal you seek to explore? Similarly an air temp of 15 deg C. Whilst admittedly not “warm”, doesn’t it depend where you are? There are many locations with variable day-to-day temp, even in summer. Did the authors explore any sensitivities to these assumptions?

The reviewer makes a good point that these thresholds deserve further examination. As outlined in the previous response, the CSWI was designed to be an improvement to the time after rain method, as this does not account for differences in evaporation, say if there is little radiation in the 48 hours after the rain event. However, they are difficult to compare, as they are not immediately comparable. Upon further analysis, the median time period for the CSWI to go from fully wet (CSWI=5) to “dry” (CSWI<=0) was 3.5 days in summer (summer being periods when daily potential radiation above median daily potential radiation for that site), so more conservative than 1-2 days after rain, but on the same order of magnitude. This comparison has been added to the manuscript:

This method was used over the more standard method of removing 1-5 days after a rain event, as it does not make the assumption that the surface will dry in a fixed amount of time, instead relying on a minimum amount of evaporation. As a comparison, the median time period for the CSWI to go from fully wet (CSWI=5) to “dry” (CSWI<=0) was 3.5 days across all sites in summer, where summer was defined as the period when daily potential radiation above median daily potential radiation for each site.

As to the GPP and air temperature limits, we have added a sensitivity analysis showing the response of DWCI and $C_{\text{LE}}^{*}$ to EF within various ranges of GPP and air temperature, Figure S2 (here R1). Based on this analysis, we looked for thresholds which would show a strong signal in the metrics under increasingly dry conditions (lower EF). In regards to temperature, our initial threshold does seem to be a bit high, as temperature thresholds above 5° C showed good metric response with reasonable variability under unstressed conditions particularly with the DWCI. As to GPP, it seems we had a problem with the units.
Figure R1: Figure S1. Sensitivity analysis of DWCI and $C^*_ET$ to evaporative fraction (EF=ET/potential ET) under different GPP and air temperature (Tair) values, as well as the sensitivity of frequency of decoupling (DWCI < threshold) and morning ($C^*_ET$ < threshold) using various thresholds. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^*_ET$ of -1-1 indicate one hour morning shifted-one hour afternoon shifted ET. Bins for GPP and Tair based on percentiles to give an equal number of points per bin, with the exception of GPP<1 and Tair<0, which are approximately double the size of the other bins and represent periods of low plant activity such as winter. Vertical bars represent interquartile range in all cases.
where the threshold was reported in grams $CO_2 \cdot m^{-2} \cdot d^{-1}$ instead of $gC \cdot m^{-2} \cdot d^{-1}$. With the unit conversion, this would correspond to a threshold of $1.36 \ gC \cdot m^{-2} \cdot d^{-1}$. We would like to thank the reviewer for spotting the error which has been corrected in the text as well as in the figures. Based on the sensitivity analysis in figure S2 (R1), we found that a threshold of $1 \ gC \cdot m^{-2} \cdot d^{-1}$ works well in minimizing unnecessary variation in the metrics, as evidence by the smaller variability in the metrics during wet conditions. The issue is now discussed in the Data section in the Methods and Materials:

Growing season was defined as all days where $GPP > 1 \ gC \cdot m^{-2} \cdot d^{-1}$ and daily mean air temperature $> 5 \ ^\circ C$. These threshold were shown to give good response in the proposed metrics while minimizing variability due to low diurnal signals, a sensitivity analysis of which can be found in supplementary Figure S2.

6. Similarly the assumption about precipitation and gap filling. What about filling it with reanalysis data? Assuming a gap corresponds to a 5 mm precipitation event strikes me as quite a big assumption? What happens if you simply assumed a gap = no rain? How important is this assumption for your results?

As precipitation data can be quite hard to predict, and we did not have immediate access to the reanalysis data for this dataset, we found it prudent to simply remove those days. In this way the data screening is conservative, erring on the side of caution by removing periods if we are not sure if it has rained or not. This conservative take is possible due to the large size of the dataset. Perhaps a further analysis with the FLUXNET 2015 database would yield more data points as some of these issues have been address in the new dataset, such as inclusion of reanalysis data gap-filling. However, to make this point clear in the manuscript, we have amended the text to read:

Any gaps in the precipitation data were assumed to be a precipitation event of 5 mm in order to prevent any unmeasured precipitation event from biasing the results by inadvertently including rainy days.

7. Page 5, line 24. Is there any evidence of this shift? I’m not arguing it isn’t true, but the authors don’t cite any supporting literature. Later on in the text the authors cite Wilson, but are there any other citations? It would be good to support this point. Figure 1 is nice and useful for demonstrating the authors point.

Indeed there has been work on shifts in ET fluxes in the literature. We have added to this section to highlight some of the relevant work, as well as further highlight the work of Wilson and it now reads:

As soils dry, it becomes more difficult to transport stem and root zone moisture to the leaf, potentially causing hydraulic limitations for the plant to transport water. This shift was seen in eddy covariance data in a study by Wilson et al. [12], who examined the shift of latent compared to sensible heat, which suggested that a shift in water fluxes towards dawn can be indicative
of afternoon stomatal closure. Shifts were further explored in a modeling study by Matheny et al. [8] which found that the morning shift was not well captured by models and attributed the errors to inadequate hydraulic limitations in the models. The daily cycle of wetting and drying acts as a capacitor in the hydraulic circuit, allowing water stores to be more easily transported in the morning and depleting in the afternoon. As bulk soil moisture declines, this effect may be strong enough to shift the diurnal cycle of ET significantly toward the morning. Quantifying diurnal shifts in EC data using the diurnal centroid was first explored by Wilson et al. [12]: defined as the flux weighted mean hour, or

\[ C_{flux} = \frac{\sum flux_x \cdot t}{\sum flux_t} \]

where \( t \) is a regular, sub-daily time interval.

8. Page 8: similarly to where I’ve made this point before, you really need to introduce things better. Suddenly the text jumps to the “Katul” and then the “Boese” models, with little or no back story. To this point I’ve found this paper really interesting, but these jumps honestly make it hard to follow and are quite distracting, so I hope I’m being constructive here.

Please see the response to the next comment.

• I don’t really follow the benchmarking models? As to get ET, you use GPP derived from flux data and then measured VPD and Rg? Why do you need a benchmarking model? To me, you simply need to apply the method to the raw data?

We find the reviewers comments very constructive and indicative of the high level of attention given to the review. We agree this has not been made clear and, as a consequence, has made subsequent parts of the manuscript unclear. The metrics were always calculated from the original flux data. The point of the model exercise was to benchmark whether the metrics were capturing information that the models were unable to predict, thus indicating that we are indeed providing novel information with these metrics and not something the models already capture. The idea is that if the models are unable to capture the variability of WUE, but the metrics are, then this indicates that the metrics are indeed providing some new and useful information that the models could benefit from. Here the Katul model would represent something closer to theory with some underlying assumptions, and the random forest would be at the opposite end of the spectrum, being wholly empirical with no assumptions. Therefore, if the metrics are capturing variability that the Katul model is not, it could indicate a problem with our assumptions in implementing the Katul model. However, if the metric captures variability that the random forest is not able to capture, it could indicate we are indeed adding new information that not even empirical methods are capturing. As this seems to be a key point that was causing confusion, in addition to the expanded section in the Introduction as discussed in comment 1, this section has been revised to outline
the motivation for using the models and now starts:

In order to benchmark whether these metrics are capturing information that is possibly not being captured in modern model frameworks, three simple models were used to estimate WUE (GPP/ET) for each day at each site and compared to actual flux data. The purpose of the exercise was to evaluate if bias in the model predictions were associated with decoupled or morning shifted days, thus indicating that the metrics are corresponding to information that the models are unable to capture. Here we utilize three models to provide a spectrum of theoretical to empirical basis.

Results ———

9. Why when DWCI < 10 is it reasonable that you have decoupling? That is stated as a fact without any support? Ditto CET < -0.5. Can you not demonstrate this for a case study i.e. the 2003 summer data from Europe, or similar?

Based on the reviewers suggestion, we have added a case study of 6 sites during the 2003 heatwave, the results of which can be seen in the new Figure 3 (here R2). This new section shows that both monthly median values and interquartile range of DWCI and $C_{ET}^*$ did respond during the heat wave years, although the response varied both between the sites and in timing that sites showed a response. The results of this case study have been summarize in the manuscript with the following addition:

Figure R 2: Monthly median diurnal water carbon index (DWCI, lower panels) and diurnal centroids ($C_{ET}^*$, upper panels) for 6 sites in Europe. Data from all years available (black) is compared to 2003 (red) during which a drought event resulted in high temperatures and low precipitation throughout the summer. Note DWCI of 0–100 indicate lowest–highest probability of diurnal carbon:water coupling and $C_{ET}^*$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Vertical bars represent interquartile range. Sites from 4 plant functional types: evergreen broadleaf (EBF), deciduous broadleaf (DBF) and evergreen needleleaf (ENF) forests, as well as grasslands (GRA). Ecosystems show tendencies of morning shifts (e.g. DK-Sor and NL-Loo) and carbon:water decoupling (e.g. ES-ES1 and HU-Bug) during the drought year.

As a case study, $C_{ET}^*$ and DWCI time-courses for eight sites from Europe are shown in Figure 3, with an emphasis on 2003 when the continent was struck by a heatwave that was shown
to effect both the carbon and water cycles [2, 10, 3]. For DWCI, forest sites showed high water-carbon coupling throughout the growing season, with the exception of Puechabon (FR-Pue) which showed a regular seasonal cycle of decoupling. The grassland site (HU-Bg) showed a higher variability in DWCI compared to the forest sites (all others). All sites showed either a decrease in median DWCI or an increase in variability during 2003, generally in July or August, particularly at Hainich (DE-Hai), Bugacpuszta (HU-Bug), and El Saler (ES-ES1). This increase in decoupling during 2003 is consistent with the hypothesis of non-stomatal limitations being expressed in hot, dry conditions. Median diurnal centroid values across all years varied in absolute magnitude, but were generally near or above zero, i.e. the water cycle showed no shift or an afternoon shift. One exception would be the Mediterranean oak forest of Puechabon, which shows a slight seasonal cycle of morning shifts going from a slight afternoon shift to a slight morning shift during June, July, and August. During drought years, sites that showed distinctive morning shifts were Puechabon (FR-Pue), Soroe (DK-Sor), and Loobos (NL-Loo). The framework that morning shifts are associated with water stress from soil moisture depletion would be supported by the increase in morning shifts during 2003, though factors such as species composition and access to soil water would play a significant factor and could account for the differences among sites. All sites which had significantly different (p<0.05, Wilcoxon rank-sum test) DWCI values between 2003 and all other years except Puechabon, whereas with $C^*_\text{ET}$ only Puechabon, Soroe, and Loobos showed significant differences.

From the case study, we saw that monthly median values rarely reached below the original thresholds presented, though interquartile ranges did at some sites. Therefore, a further sensitivity of the thresholds was carried out (Figure S2 (R1)), which show that across all sites the thresholds designating decoupling or morning shifted days tended to change the magnitude of frequency, but not the response in frequency to EF. However, these values were likely low, as the response of frequency to EF was rather dampened with the DWCI<10 and $C^*_\text{ET}<-0.5$ values. Therefore, we have increase the threshold of DWCI and $C^*_\text{ET}$ to 25 and -0.25 respectively, which we would stress did not fundamentally change the other figures, though it does amplify their patterns. It seems, especially in the case of this large scale study, the patterns are preserved across a large range of thresholds. This point is now discussed in the manuscript as follows:

The thresholds designating decoupling and morning shifts were 25 and -0.25 for DWCI and $C^*_\text{ET}$ respectively. These thresholds were chosen to highlight frequency differences between sites and were shown to have large metric responses under dry conditions while having low frequencies under wetter conditions (see sensitivity analysis in supplementary figure S2). Furthermore, these thresholds results in a similar median frequency of uncoupled and morning
shifted days between all site-years being 8.7% and 9.4% of days respectively. The similarity in median frequencies across site-years allowed for easier inter-comparison between the two metrics. The frequency of decoupling and morning shifts using these thresholds for each site can be found in the map found in File S1.

10. The 7 to 8% of all points being decoupled at all sites. Does that make sense? Wouldn't you hypothesize differences based on the vegetation? Rather than expect to find a universal value? I realize you have large uncertainty bounds, but I wonder what the implication of that finding is? Does it imply anything about the method at all? I don’t have an immediate suggestion, I’m simply surprised.

   Indeed this statement was rather misleading, as we did not mean to indicate that 7-8% of points being decoupled at all sites. The number of points varied by site, as well as by climate and PFT as seen in Figure 3 (now Figure 4 in the updated manuscript). We have changed this section to hopefully clarify this point, as well and reference the map in File S1, which also gives information of the decoupling and morning shifts on a per site basis. However, we would again like to note that these thresholds are flexible, as the patterns are robust across thresholds from ranges of about 5-75 for DWCI and -1.0-0.0 for $C^*_{ET}$. As seen in the previous comment, these points have been added to the manuscript.

11. It might be interesting to see figure 3 expressed in a more informative way. Perhaps by mean annual precipitation, or spring/summer precipitation and/or an aridity index? It would also be interesting to see how variable individual years are? You clearly have this information, but it is compressed in your presentation of Fig 3 and arguably this information is very interesting and I’d argue that you’re selling your paper short by not exploring it. For example, how variable was 2003/2010 vs other years for European sites?

   The reviewer makes a good point that by beginning the results with a figure that was meant to give a broad overview across all 189 sites does obscure to the reader the seasonal and interannual dynamics. Therefore we have added the case study of the European sites (now Figure 3 (R2) in the updated manuscript) which shows both the regular seasonal cycles in the dry subterranean site of Puechabon forest, and the response of the metrics in 2003 across the European sites (as discussed in the response to comment 9). Furthermore, to focus on the differences in response of the metrics between tree and grass ecosystems, we have added Figure 5 (R3) which shows that tree ecosystems show a heightened response to diurnal centroid under lower evaporative fraction levels, while grasses and savannas responded more with the DWCI. This new analysis helps support the discussion on tree and grass responses (in the Discussion section). Hopefully this added analysis, along with the original Figures 3-4 (now Figures 4 and 6) will give the reader a better idea of how the metrics respond across environmental conditions and ecosystems.

12. Similarly, do you see a shift in the centroid related to specific times in the year? Which sites shift earlier? What physically can you tie this to?
Figure R3: Median diurnal water carbon index (DWCI, upper panel) and diurnal centroid ($C_{ET}^*$, lower panel) of plant functional types binned by evaporative Fraction (EF, low values indicate dry conditions). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C_{ET}^*$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Evergreen needleleaf (ENF), deciduous broadleaf (DBF), and evergreen broadleaf (EBF) forests show increased morning shifts (low $C_{ET}^*$) with decreasing EF when compared to grassland (GRA) sites which tended to have decreased carbon:water decoupling (low DWCI) with decreasing EF. Savanna ecosystems (SAV) show a high degree of decoupling and intermediate levels of morning shifts. Vertical bars represent interquartile range.
As seen in the new Figure 3 (R2) on the European drought, some sites such as the Mediterranean oak forest Puechabon show a regular seasonal morning shift. More dramatic shifts can be seen at some sites during the 2003 drought, such as at Loobos forest in the Netherlands. The diversity of responses would hopefully indicate that the ecosystems are responding not just to atmospheric conditions but also soil water availability to the ecosystem, which we did not have the relevant data to properly analyse. However, Figure 4 (Figure 6 in the new manuscript, here R4) shows that the cleanest pattern in diurnal centroid is associated with periods of high net radiation and low latent energy, which is indicative water limited conditions. As for the DWCI, we see in the sensitivity analysis to air temperature (Figure S2 (R1)) that the decoupling tends to happen under wetter conditions at higher temperatures, so there may be a temperature component as well. We hope the added analysis focusing on particular sites during the European droughts give a better picture of how the metrics respond on different seasonal timescales.

13. I don’t find figure 4 all that informative. Again I wonder if you are exploiting the interesting findings to their fullest? Which sites are most decoupled? Which vegetation types? Does it make sense to exclude the well coupled days, you’re not really interested in these days?

Indeed as Figure 4 (here R4, now Figure 6 in the new manuscript) was intended to show that the response of the metrics could be seen across all sites, indicating that the metrics are indeed universally useful. However, this broad analysis then masks the dynamics seen from site to site and at different times of year. As mentioned in the previous comment, we see from the new Figure 5 (R3) that grassland and savanna ecosystems tend to decouple more and at higher evaporative fractions than tree ecosystems, possibly due to a lack of hydraulic architecture as discussed in the discussion sub-section “trees, grass, and drought stress”. Regarding excluding the coupled days with regards to Figure 4 (R4, now Figure 6), as we are hoping to show the universality of the response, we want to be transparent and not mask any “false positives” one would see, as is the case of the noisiness of the response of DWCI in Figure 4a ((R4b) now Figure 6b). To highlight this point, the Figure is now referenced in the manuscript as:

The response of both variables to drought stress is further observed in Figure 6, where low mean values of both DWCI and $C^*_{ET}$ are associated with conditions of high net radiation and low latent energy, indicative of drought. As this figure includes all days from all sites which meet the filtering outlined in the Data subsection of the Methods, i.e. dry periods in the growing season, these figures exhibit the universality of the metrics across climates, ecosystems, and time periods.

Note the new subfigures (R4a,d), which were a response to the comments of reviewer 3 (comment 2)

14. I won’t really comment on Fig 5 because I don’t follow the motivation. Partly because of my question about VPD and partly because I don’t see why the metrics which are data driven, require a benchmark like this? I’m not totally sold on
Figure R 4: Mean DWCI (upper panels) and $C^*_\text{ET}$ (lower panels) with respect to evaporative fraction (EF) by vapor pressure deficit VPD (a,d), latent energy (LE) by Rn (b,e) and LE by GPP (c,g). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^*_\text{ET}$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Points with high Rn and low LE are associated with both low DWCI and $C^*_\text{ET}$, indicating that both metrics are related to water limitations. Though both metrics are associated with low EF, DWCI shows a much higher response to atmospheric demand as measured by VPD, with $C^*_\text{ET}$ showing very limited response. Both metrics, and DWCI in particular, show low values with high ET and low Rn, though these points are also associated with over closed energy balances (LE+H>Rn). Both metrics are associated with low GPP, but the $C^*_\text{ET}$ is restricted to both low GPP and ET, indicating water and carbon can decouple over a wider range of water stress. This also holds when points with energy balance over-closer are excluded (data not shown).
this being an objective means to test the approach, but appreciate why the authors have taken this approach. This is simply my opinion and I’m sure others would disagree. My first point of my discussion text below would be the way I would have been tempted to proceed.

The idea behind Figure 5 (now Figure 7) was to really get at the question of whether the metrics are giving some new information on carbon water dynamics that a model couldn’t capture, as discussed in the response to comment 8. As it is clear by the responses of all the reviewers, this point was not originally made clear to the reader. In an attempt to clarify this point, we have added text to highlight this point when outlining Figure 5(now 7) which also refers back to the expanded explanation in the materials and methods:

Figure 7 shows the difference between expected and observed WUE from the Katul, Boese, and random forest (RF) models, with respect to conditions of drought as characterized by low evaporative fraction (EF<0.2), C:W decoupling (DWCI<25), and morning shifts ($C_{E/T}<-0.25$). This exercise was designed to test whether the metrics were associated with bias in the models, indicating that the metrics are able to capture information that the models are not (as further outlined in Methods and Materials subsection “models and parameter estimation”).

Discussion ———-

15. How do we know the method works? What would be the best test of the method? Even if the authors don’t have access to the necessary data, could they set a challenge to the community? For example, if groups had sapflux or information on non-stomatal limitations at any flux sites, do the authors have thoughts how these data could be used? How should the community push such an approach forward?

We agree with the reviewer that though we have demonstrated the phenomenon, this broader approach would be supplemented by a site level investigation. With the added figures that show the response of the metrics during the heatwave of 2003, we hope that we give readers confidence that the metrics are indeed showing ecosystem stress. Furthermore, in an effort to make a call to the community we have added the following section to the discussion:

Given the broad nature of the analysis here, the metrics and hypothesis presented would benefit from site specific validations such as looking to see if the morning shits and decoupling are indeed associated with lower soil moisture levels, leaf water potentials, and/or decreases in sap flux. Sap flux in particular could give some interesting insights, as the diurnal patterns in sap flux velocity will also have an offset to incoming radiation related to tree capacitance, therefore relating sap flow diurnal centroids to the ET diurnal centroid could give some information on changes in plant water recharge. Furthermore, the diurnal centroid base metrics complement
the hysteresis quantification methods such as those employed by Zhou et al. [14] and Matheny et al. [8], with the advantage of $C_{ET}^*$ being compensation for cloudy conditions and possibly less influence of noise, though an intercomparison would be useful to explore the strengths and weaknesses of the different approaches. By providing both the equations and related code of the metrics, we the authors hope the metrics will be used by the community for both validation and to further ecophysiological understanding.

15. I think the discussion of trees vs grasses is interesting and welcomed, but I wonder if the authors looked at exploring a bit more within a functional group (i.e. by aridity etc), whether they might find something else too. Up to the authors of course.

   Based on the reviewers suggestion, we have added Figure 5 (R3) which shows a more in depth view of the tree vs grass responses based on binned evaporative fraction. The figure further shows the distinct patterns with tree and grass responses, which we have highlighted in this section of the discussion.

16. I'd argue that the authors could set aside some text to suggesting how their approach could be used in terms of benchmarking land models during drought? I'm not saying this paper has to do such a comparison, but it might be advantageous to lay the ground work. I'm guessing that the authors see modellers as potential users of their metrics? and if so, it is worth them making a case. Assessing models for responses to drought is very complicated and so their approach is welcomed.

   We agree with the reviewer that this would be a useful addition and have added the following text to the discussion:

   Though the models used here are relatively simple and lack the complexities and feedbacks found in more vigorous ecosystem models, Matheny et al. [8] also demonstrated the fundamental inability of 9 different land-surface models with 4 different stomatal conductance schemes to capture diurnal variability which the authors attributed to inadequate representation of how water gets from the soil to the leaf. Given the demonstrated phenomenon of morning shifts and decoupling across sites under dry conditions, the metrics here provide a benchmarking tool for mechanistic models to test their ability to replicate these patterns, suggesting that the models are capable of expressing hydraulic and non-stomatal limitations. Furthermore, in the case of machine learning approaches, the metrics may provide a useful input parameter which summarizes these diurnal effects, as is evidence by the bias in RF modeled WUE associated with morning shifts. As such, by demonstrating the utility of the metrics, and providing code and explanations for calculation, we hope they become useful to the community at large.
References


Response to interactive comment by Anonymous Referee #2 on “Water stress induced breakdown of carbon-water relations: indicators from diurnal FLUXNET patterns”

This study proposes two data-driven indicators using eddy covariance data to examine water stress induced breakdown of carbon-water relations. These results are scientifically interesting. Sharing code is also a good practice and should be praised, although the calculations seem to be straightforward. Given the problems with the manuscript, I recommend that it be reconsidered after a major revision.

1. The biggest problem of this manuscript is that the Results section is very weak. It only contains three relatively short paragraphs, which is far from being sufficient for a research article. The authors should substantially strengthen this section.

As all of the reviewers have made this point, we have expanded this section considerably. First we have included a new section focused on a case study of the European heatwave of 2003, in which one can see morning shifts and decoupling of water and carbon at 6 sites across Europe (now Figure 3 in the new manuscript, here Figure (R1)). The following text has been added to the results in relation to this figure:

![Figure R1](image)

Figure R1: Monthly median diurnal water carbon index (DWCI, lower panels) and diurnal centroids ($C^{*}_{ET}$, upper panels) for 6 sites in Europe. Data from all years available (black) is compared to 2003 (red) during which a drought event resulted in high temperatures and low precipitation throughout the summer. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^{*}_{ET}$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Vertical bars represent interquartile range. Sites from 4 plant functional types: evergreen broadleaf (EBF), deciduous broadleaf (DBF) and evergreen needleleaf (ENF) forests, as well as grasslands (GRA). Ecosystems show tendencies of morning shifts (e.g. DK-Sor and NL-Loo) and carbon:water decoupling (e.g. ES-ES1 and HU-Bug) during the drought year.

As a case study, $C^{*}_{ET}$ and DWCI time-courses for eight sites from Europe are shown in Figure 3, with an emphasis on 2003 when the continent was struck by a heatwave that was shown to effect
both the carbon and water cycles [1, 4, 2]. For DWCI, forest sties showed high water:carbon coupling throughout the growing season, with the exception of Peuchebon (FR-Pue) which showed a regular seasonal cycle of decoupling. The grassland site (HU-Bg) showed a higher variability in DWCI compared to the forest sites (all others). All sites showed either a decrease in median DWCI or an increase in variability during 2003, generally in July or August, particularly at Hainich (DE-Hai), Bugacpuszta (HU-Bug), and El Saler (ES-ES1). This increase in decoupling during 2003 is consistent with the hypothesis of non-stomatal limitations being expressed in hot, dry conditions. Median diurnal centroid values across all years varied in absolute magnitude, but were generally near or above zero, i.e. the water cycle showed no shift or an afternoon shift. One exception would be the Mediterranean oak forest of Puechabon, which shows a slight seasonal cycle of morning shifts going from a slight afternoon shift to a slight morning shift during June, July, and August. During drought years, sites that showed distinctive morning shifts were Puechabon (FR-Pue), Soroe (DK-Sor), and Loobos (NL-Loo). The framework that morning shifts are associated with water stress from soil moisture depletion would be supported by the increase in morning shifts during 2003, though factors such as species composition and access to soil water would play a significant factor and could account for the differences among sites. All sites which had significantly different (p<0.05, Wilcoxon rank-sum test) DWCI values between 2003 and all other years except Puechabon, whereas with $C^*_ET$ only Puechabon, Soroe, and Loobos showed significant differences.

Furthermore, to highlight the divergent responses of the metrics between tree and grass dominated ecosystems, we have added a figure which shows the response of forest, savanna, and grassland $C^*_ET$ and DWCI values binned by evaporative fraction (now Figure 5, here R2). In the figure, one can see the tendency for grassland sites to decouple under low evaporative fraction, while forest sites tend to show morning shifts, which ties into the discussion on tree vs grass responses and isohydricity.

Finally, we have added two new subplots to Figure 3 (now Figure 6 in the new manuscript, here (R3)), which in addition to current subplots showing the DWCI and $C^*_ET$ response to LE by Rn and LE by GPP, shows the response to VPD by evaporative fraction (EF). These new subplots further differentiate the responses of the two metrics, as DWCI shows a combined effect of VPD and EF but $C^*_ET$ only shows a response to EF and not VPD.

We hope these new plots, as well as the associated references to them in the results, give the reader a better understanding of the dynamics of the metrics and further support our subsequent Discussion section.

2. The authors claimed that they proposed two indicators. Are these indicators new and have not been used in the literature?
Figure R 2: Median diurnal water carbon index (DWCI, upper panel) and diurnal centroid ($C_{ET}^*$, lower panel) of plant functional types binned by evaporative Fraction (EF, low values indicate dry conditions). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C_{ET}^*$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Evergreen needleleaf (ENF), deciduous broadleaf (DBF), and evergreen broadleaf (EBF) forests show increased morning shifts (low $C_{ET}^*$) with decreasing EF when compared to grassland (GRA) sites which tended to have decreased carbon:water decoupling (low DWCI) with decreasing EF. Savanna ecosystems (SAV) show a high degree of decoupling and intermediate levels of morning shifts. Vertical bars represent interquartile range.
Figure R 3: Mean DWCI (upper panels) and $C^*_ET$ (lower panels) with respect to evaporative fraction (EF) by vapor pressure deficit VPD (a,d), latent energy (LE) by Rn (b,e) and LE by GPP (c,g). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^*_ET$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Points with high Rn and low LE are associated with both low DWCI and $C^*_ET$, indicating that both metrics are related to water limitations. Though both metrics are associated with low EF, DWCI shows a much higher response to atmospheric demand as measured by VPD, with $C^*_ET$ showing very limited response. Both metrics, and DWCI in particular, show low values with high ET and low Rn, though these points are also associated with over closed energy balances (LE+H>Rn). Both metrics are associated with low GPP, but the $C^*_ET$ is restricted to both low GPP and ET, indicating water and carbon can decouple over a wider range of water stress. This also holds when points with energy balance over-closer are excluded (data not shown).
If they have been used in the literature, then they are not new and the authors should rephrase this statement. Take diurnal centroid as an example, it was used back in 2003 (Wilson et al. 2003), and what is new with the relative diurnal centroid?

The reviewer makes a good point that the relative diurnal centroid builds on the work of Wilson et al. [5] and should be properly credited. To this end the objectives have been amended to clarify that we propose these metrics particularly as indicators of water stress and now reads:

To this end, we propose two data-driven indicators of water stress, the diurnal water:carbon index (DWCI) and the relative diurnal centroid in LE ($C_{ET}^*$).

We have also made the reference to Wilson et al. [5] more explicit, as well as highlighting the work of Matheny et al. [3] when introducing the metric:

As soils dry, it becomes more difficult to transport stem and root zone moisture to the leaf, potentially causing hydraulic limitations for the plant to transport water. This shift was seen in eddy covariance data in a study by Wilson et al. [5], who examined the shift of latent compared to sensible heat, which suggested that a shift in water fluxes towards dawn can be indicative of afternoon stomatal closure. Shifts were further explored in a modeling study by Matheny et al. [3] which found that the morning shift was not well captured by models and attributed the errors to inadequate hydraulic limitations in the models. The daily cycle of wetting and drying acts as a capacitor in the hydraulic circuit, allowing water stores to be more easily transported in the morning and depleting in the afternoon. As bulk soil moisture declines, this effect may be strong enough to shift the diurnal cycle of ET significantly toward the morning. Quantifying diurnal shifts in EC data using the diurnal centroid was first explored by Wilson et al. [5]: defined as the flux weighted mean hour, or

$$C_{flu} = \frac{\sum \text{flux}_t \cdot t}{\sum \text{flux}_t}$$

where t is a regular, sub-daily time interval.

Hopefully these changes give better context to the manuscript and proper credit to previous works.

3. The abstract should be rewritten to contain less introduction and more results.

The abstract has been rewritten to remove some introduction as well as to highlight the added analysis from the European heatwave in 2003:

Understanding of terrestrial carbon and water cycles is currently hampered by an uncertainty in how to capture the large variety of plant responses to drought. In FLUXNET, the global network of CO2 and H2O flux observations, many sites do not uniformly report the ancillary variables
needed to study drought response physiology. To this end, we outline two data-driven indicators based on diurnal energy, water, and carbon flux patterns derived directly from the eddy covariance data and based on theorized physiological responses to hydraulic and non-stomatal limitations. Hydraulic limitations (i.e. intra-plant limitations to water movement) are proxied using the relative diurnal centroid ($C^*_{ET}$), which measures the degree to which the flux of evapotranspiration (ET) is shifted toward the morning. Non-stomatal limitations (e.g. inhibitions of biochemical reactions, Rubisco activity, and/or mesophyll conductance) are characterized by the Diurnal Water:Carbon Index (DWCI), which measures the degree of coupling between ET and gross primary productivity (GPP) within each day. As a proof of concept, the metrics indicated morning shifts and decoupling effects at 6 European sites during the 2003 heatwave event. Globally, we found indications of hydraulic limitations in the form of significantly high frequencies of morning shifted days in dry/Mediterranean climates and savanna/evergreen plant functional types (PFT), whereas high frequencies of decoupling were dominated by dry climates and grassland/savanna PFTs indicating a prevalence of non-stomatal limitations in these ecosystems. Overall, both the diurnal centroid and DWCI were associated with high net radiation and low latent energy typical of drought. Using three water use efficiency (WUE) models, we found the mean differences between expected and observed WUE to be -0.09 to 0.44 umol/mmol and -0.29 to -0.40 umol/mmol for decoupled and morning shifted days respectively compared to mean differences -1.41 to -1.42 umol/mmol in dry conditions. These results suggest that morning shifts/hydraulic responses are associated with an increase in WUE whereas decoupling/non-stomatal limitations are not.

4. The Introduction section is organized by sub-sections. The three levels of organization is pretty unusual for scientific papers. I suggest that the authors remove sub-section titles and rewrite it as a regular Introduction. If the authors intend to provide more background material, it is better to add a Background section right after the Introduction.

Based on the reviewers suggestion, the sub-sections have been removed and the Introductions has been further shortened to be more focused on the subject at hand, such as removing the focus on transpiration estimates as suggested by reviewer 1 (comment 3), as well as removing discussion of the iWUE metrics which hopefully alleviates some unnecessary confusion.

5. It is unclear which FLUXNET database (LaThuile or 2015) is used in this study. Details should be provided. The Baldocchi et al. 2008 paper does not seem to be the proper citation for the database used. The supplementary figure can be moved to the manuscript given the relatively small number of illustrations.

We agree that this was not made entirely clear and the reference was not the most appropriate one. We have updated our references to include the citation of the dataset pointing to the actual download location.

6
Further, as we have expanded the results section based on the reviewers suggestions, we have increased the number of figures from 5 to 7, with an additional supplementary figure showing a sensitivity analysis of the metrics to GPP and air temperature, as well as sensitivity of the frequency decoupled and morning shifted days based on different thresholds.

6. Line 30, page 4: FileS1 is a nice way to present study sites. Adding some technical details about how this kind of file was created will be helpful to the audience who might want to produce this kind of illustration as well.

   We are glad the reviewer likes our presentation of the sites. The file is created using the Bokeh package in Python based on the following technical example: https://bokeh.pydata.org/en/latest/docs/user_guide/geo.html. A reference to this example has been added to File S1, hopefully allowing these types of presentations to be more commonplace in literature.

7. Figure 2: An overall title should be added for the figure. Moreover, “upper” and “lower” should be changed to something like “Upper panels:” and “Lower panels:”, respectively.

   We have given this figure the overall title of “Theoretical overview of diurnal water carbon index” and the upper and lower designations have been changed as the reviewer suggested. This designation was also used in the new figures (now figures 3 (R1) and 5 (R2))

8. Line 8, page 3: “are GPP” should be changed to “GPP are”.

   This has been corrected.

9. Line 25, page 8: remove “be”.

   This has been corrected.

References


Response to enteractive comment by Anonymous Referee #3 on “Water stress induced breakdown of carbon-water relations: indicators from diurnal FLUXNET patterns”

Overall, I am supportive of the goals of this study. I agree that the asymmetry of the diurnal cycle of ET, and the correlation between GPP and ET, likely contain meaningful information about ecosystem response to drought. I also applaud the author's efforts to link these metrics to insights informed by mechanistic theory.

1. The analysis presented here is very broad (i.e. results are synthesized across many sites, and often many PFTs). Cross-site syntheses like this are invaluable for understanding broad patterns in vegetation functioning; however, the objectives of this study would benefit from a more in-depth analysis of the results at at least few sites, preferably sites where there exist independent estimates of plant water relations during periods of hydrologic stress (i.e. from gas exchange, sap flux, isotope analysis of tree cores, etc). This sort of analysis would give the readers confidence that the C_ET and DWCI metrics are really reflecting stomatal and non-stomatal limitations to transpiration and carbon uptake, and aren’t unduly contaminated by the many sources of uncertainty in using eddy covariance measurements to infer GPP and water use efficiency (e.g. Knauer et al. 2017). Or, to put it differently the new metrics introduced in this manuscript merit some “proof of concept” before they are applied broadly.

As the reviewer has pointed out, the presented analysis is indeed takes a broad approach to ecosystem physiology. This broad approach was undertaken in part due to the lack of a congruent dataset of independent estimates of plant water relations with which to compare the FLUXNET dataset. While sites do exists with measurements such as sap flux and isotope composition, they are currently not well homogenized and synchronizes in a way that is conducive to do an analysis across ecosystems, though some are definitely in the pipeline such as the SAPFLUXNET initiative. However, in an effort to give the presented manuscript a more in-depth analysis, we have included a new section in the manuscript which focuses on 6 European sites and compares average monthly values to the monthly values in 2003 when the continent was hit by a heat wave and drought. As seen in Figure 3 (here R1), the sites show diverse responses in the summer of 2003, with some exhibiting distinct morning shifts and some decoupling, but each of the sites show some response to the high temperatures and low water conditions of the heatwave. This new figure is accompanied by the following text in the Results section:

As a case study, $C_{ET}^*$ and DWCI time-courses for eight sites from Europe are shown in Figure 3, with an emphasis on 2003 when the continent was struck by a heatwave that was shown to effect both the carbon and water cycles [1, 9, 2]. For DWCI, forest sties showed high water:carbon coupling throughout the growing season, with the exception of Peuchebon (FR-Pue) which showed
Figure R 1: Monthly median diurnal water carbon index (DWCI, lower panels) and diurnal centroids ($C_{ET}^*$, upper panels) for 6 sites in Europe. Data from all years available (black) is compared to 2003 (red) during which a drought event resulted in high temperatures and low precipitation throughout the summer. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C_{ET}^*$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Vertical bars represent interquartile range. Sites from 4 plant functional types: evergreen broadleaf (EBF), deciduous broadleaf (DBF) and evergreen needleleaf (ENF) forests, as well as grasslands (GRA). Ecosystems show tendencies of morning shifts (e.g. DK-Sor and NL-Loo) and carbon:water decoupling (e.g. ES-ES1 and HU-Bug) during the drought year.

A regular seasonal cycle of decoupling. The grassland site (HU-Bg) showed a higher variability in DWCI compared to the forest sites (all others). All sites showed either a decrease in median DWCI or an increase in variability during 2003, generally in July or August, particularly at Hainich (DE-Hai), Bugacpuszta (HU-Bug), and El Saler (ES-ES1). This increase in decoupling during 2003 is consistent with the hypothesis of non-stomatal limitations being expressed in hot, dry conditions. Median diurnal centroid values across all years varied in absolute magnitude, but were generally near or above zero, i.e. the water cycle showed no shift or an afternoon shift. One exception would be the Mediterranean oak forest of Puechabon, which shows a slight seasonal cycle of morning shifts going from a slight afternoon shift to a slight morning shift during June, July, and August. During drought years, sites that showed distinctive morning shifts were Puechabon (FR-Pue), Soroe (DK-Sor), and Loobos (NL-Loo). The framework that morning shifts are associated with water stress from soil moisture depletion would be supported by the increase in morning shifts during 2003, though factors such as species composition and access to soil water would play a significant factor and could account for the differences among sites. All sites which had significantly different ($p<0.05$, Wilcoxon rank-sum test) DWCI values between 2003 and all other years except Puechabon, whereas with $C_{ET}^*$ only Puechabon, Soroe, and Loobos showed significant differences.

Furthermore, we hope that by demonstrating the response of the metrics across a large number of sites, and providing the equations and associated code, we can provide the tools to those with both the access...
and expertise of these particular sites with independent measurements. In this regard, we have taken
the suggestion given by reviewer #1 to give a direct call for such an analysis, with the following text being
added to the Discussion section “Looking beyond sums and means”:

Given the broad nature of the analysis here, the metrics and hypothesis presented would benefit
from site specific validations such as looking to see if the morning shifts and decoupling are
indeed associated with lower soil moisture levels, leaf water potentials, and/or decreases in sap
flux. Sap flux in particular could give some interesting insights, as the diurnal patterns in sap
flux velocity will also have an offset to incoming radiation related to tree capacitance, therefore
relating sap flow diurnal centroids to the ET diurnal centroid could give some information on
changes in plant water recharge. Furthermore, the diurnal centroid base metrics complement
the hysteresis quantification methods such as those employed by Zhou et al. [13] and Matheny
et al. [7], with the advantage of $C_{ET}$ being compensation for cloudy conditions and possibly
less influence of noise, though an intercomparison would be useful to explore the strengths and
weaknesses of the different approaches. By providing both the equations and related code of
the metrics, we the authors hope the metrics will be used by the community for both validation
and to further ecophysiological understanding.

Finally, we have added Figure 5 (R2) highlighting the tree vs. grass responses based on evaporative frac-
tion, which shows both the effect of decoupling and morning shifts in forest, savanna, and grassland
sites with lower evaporative fraction, as well as showing the divergent responses. We hope these added
analyses will give the reader confidence that the metrics are reflecting water stress responses.

2. I also had a few concerns about the presentation and interpretation of the water use efficiency theory. First, the authors
attributed the afternoon decline in ET to “hydraulic limitation” driven specifically by challenges of moving liquid water
from roots to the leaves as soil dries (e.g. Lines 24-25). While I agree that hydraulics are an important control on stomatal
functioning, stomates may also close directly in response to rising VPD even if soil moisture is unchanged (as discussed
at length in the stomatal optimization literature), and the mechanisms responsible for the VPD response are still not yet
clear. Thus, it may be more appropriate to describe the afternoon decline in ET as simply “stomatal limitations.”

The reviewer makes a very good point and highlights the complexities of the problem. Indeed, the key
control mechanism a plant has in the soil-plant-atmosphere continuum is via the stomates, especially
at the diurnal scale. Likewise, we agree that the stomates have been shown to have a primary response
to VPD, indeed we introduce the $VPD^{-0.5}$ term when calculating the DWCI in an effort to mitigate
the VPD effects on decoupling diurnal GPP and ET. However, the closure of stomates due to VPD would
cause an afternoon decrease in GPP but not necessarily in ET as VPD is also a driver of ET. In other words,
as the stomate is closing in response to VPD, VPD is also pulling harder at the water in the leaf. So the
Figure R 2: Median diurnal water carbon index (DWCI, upper panel) and diurnal centroid ($C^*_ET$, lower panel) of plant functional types binned by evaporative Fraction (EF, low values indicate dry conditions). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^*_ET$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Evergreen needleleaf (ENF), deciduous broadleaf (DBF), and evergreen broadleaf (EBF) forests show increased morning shifts (low $C^*_ET$) with decreasing EF when compared to grassland (GRA) sites which tended to have decreased carbon:water decoupling (low DWCI) with decreasing EF. Savanna ecosystems (SAV) show a high degree of decoupling and intermediate levels of morning shifts. Vertical bars represent interquartile range.
hypothesis driving the diurnal centroid is then that the plant is responding to some form of hydraulic limitation of moving water, with the assumption that if the plant had unlimited access to water, and the ability to move it to the leaves, it would continue to keep the stomates open and take up carbon (the carbon cost of water would be zero as water is infinite). Though the reviewer makes a good point and this is simply a hypothesis for the metrics. As a further test, we have added additional subplots to Figure 4 (Figure 6a,d in the new manuscript, here R3a,d) which shows the response (as color) along the axes of evaporative fraction (EF=ET/ETpot) and VPD.

Figure R 3: Mean DWCI (upper panels) and $C^{*}_{ET}$ (lower panels) with respect to evaporative fraction (EF) by vapor pressure deficit VPD (a,d), latent energy (LE) by Rn (b,e) and LE by GPP (c,g). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C^{*}_{ET}$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Points with high Rn and low LE are associated with both low DWCI and $C^{*}_{ET}$, indicating that both metrics are related to water limitations. Though both metrics are associated with low EF, DWCI shows a much higher response to atmospheric demand as measured by VPD, with $C^{*}_{ET}$ showing very limited response. Both metrics, and DWCI in particular, show low values with high ET and low Rn, though these points are also associated with over closed energy balances (LE+H>Rn). Both metrics are associated with low GPP, but the $C^{*}_{ET}$ is restricted to both low GPP and ET, indicating water and carbon can decouple over a wider range of water stress. This also holds when points with energy balance over-closer are excluded (data not shown).

From the new subplot (Figure 6a,d (R3a,d)), one can see that the diurnal centroid in fact is not very responsive to mean daily VPD, instead being almost entirely responsive to EF. This would support our hypothesis that the morning shift is significantly driven by hydraulic limitations rather then simply the VPD response. In contrast, the same subplots shows that the DWCI is much more responsive to high VPD, indicating that decoupling is happening during high atmospheric demand. This is highlighted in the Results section with the text:

Apart from the response to periods of high LE and low Rn, the metrics showed diverging response when looking at EF (ET/PET which is similar to LE/Rn) and VPD, with DWCI showing a much stronger response to VPD and $C^{*}_{ET}$ showing a much stronger response to EF (Figure
This difference in response would indicate that DWCI is more responsive to atmospheric demand (estimated via VPD) and $C^*_ET$ is more responsive to water limitations.

So while the reviewer is correct in that this morning shift likely would ultimately be a stomatal control, we hope to highlight this underlying hypothesis which differentiates what we are looking for from the stomatal limitations often describing VPD effects on carbon uptake. To clarify this point in the introduction, we have added the following text:

Under this hydraulic limitation framework a plant will be reacting to the inability to transport water, even though the key control mechanism for a plant is via the stomata, possibly expressed as an increase in sensitivity. Such assumptions are consistent with the mechanisms encoded in some land surface and ecosystem models, which account for water limitations by scaling the water to carbon ratio in relation to available soil moisture.

3. Next, the function $ET = iGPP\sqrt{VPD}$, proposed by Zhou et al. (2015), is referred to in this manuscript as the “Katul” model; presumably this nomenclature originates from the theory presented in Katul et al. (2010), which presents arguments leading to the equation: $ET = GPP \times \sqrt{VPD} / \sqrt{1.6 \lambda ca}$ The parameter lambda is the so-called “marginal water use efficiency” and $ca$ is atmospheric CO2 concentration. This is similar to the Eq. 8 in the present manuscript: $ET = i \times GPP \times \sqrt{VPD}$ If $i = 1/\sqrt{1.6 \lambda ca}$ I appreciate that the authors have attributed the model to Katul et al. (2010), who presented the theory on which the equation is based. However, before the authors attribute the model’s “inability to make accurate predictions” to “be a result of a failure of their underlying assumptions,” (Page 9, lines 10-11), care should be taken to make sure the underlying assumptions are properly stated and considered. The Katul et al. (2010) result relies on an assumption of Rubisco-limited photosynthesis (which generates a linear A-Ci curve). . .this assumption is not likely to hold in dense forests where understory vegetation is often light-limited. Second, and perhaps more importantly, stomatal optimization theory assumes that the parameter lambda should hold constant over timescales of hours, but varies over longer timescales (days to weeks) as other slowly-evolving boundary conditions change (Manzoni et al. 2013, Palmroth et al. 2013). So in that regard, I disagree with the author’s assessment on Page 8, Line 5, that the “Katul” model “makes the assumption that the WUE is constant if corrected by the effect of VPD.” The potential for lambda to vary may also help to explain the tendency of the Katul (and other) models to underestimate WUE during dry conditions (i.e. Figure 5), especially if the ‘i’ parameter is determined using observations from well-watered conditions.

We would like to thank the reviewer for so succinctly summarizing what is not always made clear in many manuscripts, including this one in present state. In regard to the initial comments, the reference to Katul 2010 is one that should be cited in this context and this reference has now been included in the appropriate places. Furthermore, the text has been changed so that the model is now initially introduced as:

These carbon:water links are fundamental to understanding how stomata are regulated and...
underly key functioning in mechanistic plant and ecosystem models. One such set of models are those based on optimality theory which posit that plants tend to optimize carbon gains to water losses, such as those described by Katul et al. [4] and Katul, Palmroth, and Oren [3]. These concepts from Katul, which carry the assumptions of RuBISCO limitation, were built upon by Zhou et al. [13] and Zhou et al. [11] to give the equation,

$$uWUE = \frac{GPP \cdot \sqrt{VPD}}{ET}$$

where the $\sqrt{VPD}$ accounts for the stomatal response to vapor pressure deficit (VPD). Accounting for the VPD response allows for a more stable metric of WUE that is temporally more stable and physiologically more meaningful, such as when comparing the diurnal cycles of carbon and water.

The point that the carbon cost of water is likely to change is further highlighted in the Materials and Methods where the subsection “Models and parameter estimation” has been expanded to say,

The “Katul” model, as defined and used in calculation of the DWCI, is based in stomatal optimization theory [4, 3, 11], which makes the assumption that the WUE is constant if corrected by the effect of VPD, using an inverse square root as the assumed relationship. Though the constant nature of $uWUE$ may not be correct, with the optimal carbon cost of water changing over day or weeks [5, 8], a yearly parameter of $uWUE$ was estimated which is consistent with other modeling exercises [12].

Note that an explanation of how these yearly parameters are calculated follows this statement, which hasn’t changed from the original submission. Finally, as the reviewer has pointed out, our statement that biases of the Katul model are a failure of underlying assumptions becomes misleading as we do not reference what assumptions we have made. As such, this has been amended to say:

Both the Katul and Boese models are theoretically based and here implemented have the underlying assumptions of RuBiSCO-limited conditions and constant carbon cost of water throughout the season which may not reflect reality.

4. Finally, when using the shape of the diurnal pattern of ET to infer stomatal limitation, I wondered why the authors focused only on the shift in the peak, and not the overall degree of asymmetry between morning and afternoon periods (for example, if the ET data from hours 0-12 are reflected about the solar noon axis, that is the area between the reflected and actual ET data).

This idea was actually explored during the course of the analysis while looking for a metric to quantify
the asymmetry of the diurnal patterns, and would lead in the direction of hysteresis analysis that other researchers had done such as Zhou et al. [13] and Matheny et al. [7]. One issue with this type of analysis is that one tends to compare individual half hours of the morning to half hours in the evening which can be problematic. For example, in the case of the example the reviewer suggests, a cloud passing in afternoon would be indistinguishable to a physiological response to water limitation as both would reduce the water flux. So by using the diurnal centroid of ET vs Rg we are able to measure the asymmetry while correcting for changes in incoming energy. Previous experiences with hysteresis approaches showed that the quantifications could be very sensitive to noise, especially at the daily scale and the diurnal centroid may be a more robust metric. However, the reviewer makes a good point that hysteresis quantification and diurnal centroid based metrics are two approaches to explore the same effect, a point which we have added to the discussion:

Furthermore, the diurnal centroid base metrics complement the hysteresis quantification methods such as those employed by Zhou et al. [13] and Matheny et al. [7], with the advantage of $C_{ET}^*$ being compensation for cloudy conditions and possibly less influence of noise, though an intercomparison would be useful to explore the strengths and weaknesses of the different approaches.

A few other comments include:

5. Page 2, lines 15-24. The presentation of the iWUE models would benefit from a more general explanation of what each of the three metrics actually describes and/or is sensitive too. E.g. WUE is useful for understanding broad patterns of ecosystem water use and carbon uptake, but is sensitive to non-biological drivers (e.g. VPD). The iWUE attempts to correct for the direct effect of VPD on transpiration, and is thus a more biologically relevant metric. The uWUE further attempts to correct for stomatal closure under high VPD, and therefore may be more closely linked to the “non-stomatal” limitations to gas exchange during drought.

In an effort to make this section more clean and concise, we have removed the reference to Beer et al. 2008 and iWUE as it was not further referenced in the manuscript. We have also taken the reviewers advice and expanded the section to give some context to the VPD term. The text now reads:

These carbon:water links are fundamental to understanding how stomata are regulated and underly key functioning in mechanistic plant and ecosystem models. One such set of models are those based on optimality theory which posit that plants tend to optimize carbon gains to water losses, such as those described by Katul et al. [4] and Katul, Palmroth, and Oren [3]. These concepts from Katul, which carry the assumptions of RuBISCO limitation, were built upon by Zhou et al. [13] and Zhou et al. [11] to give the equation,
\[ uWUE = \frac{GPP \cdot \sqrt{VPD}}{ET} \]

where the \( \sqrt{VPD} \) accounts for the stomatal response to vapor pressure deficit (VPD). Accounting for the VPD response allows for a more stable metric of WUE that is temporally more stable and physiologically more meaningful, such as when comparing the diurnal cycles of carbon and water.

6. Page 4, Line 17: “the use of EC measured diurnal patterns of carbon, water, and energy fluxes to derive clues on ecosystem drought responses at a daily resolution could prove valuable, if nothing less than a benchmark to test current hypotheses.” As far as rationale for the work goes, I found this to be rather weak. Perhaps the authors could give specific examples of hypotheses that could be tested with these metrics.

We agree that this statement was rather vague, it has now been changed to be more explicit:

In this sense, the use of EC measured diurnal patterns of carbon, water, and energy fluxes to derive clues on ecosystem drought responses at a daily resolution could prove valuable both as a means to identify potential periods of ecosystem stress, inform machine learning algorithms on ecophysiological conditions not found in environmental variables, as well as benchmarking a models ability to capture sub-daily dynamics.

7. Page 9, line 20: “sites under water stress tended to have C_E < -0.50.” How do the authors know that the sites were under water stress? This gets back to my original point about validating the metrics against independent observations of plant function.

The reviewer makes a good point, one which was also brought up by reviewer 1 (comment 9). As discussed in comment 1 of this review, evidence of metric response is seen in the case study of the European heatwave of 2003. Furthermore, to address the appropriateness of these thresholds, we have included a new sensitivity analysis which shows the relationship of frequency of uncoupled and morning shifted days based on different threshold levels (Figure S2. here R4). This analysis indicates that indeed these levels may have been slightly too strict and have been changed to DWCI<25 and \( C^*_ET < -0.25 \), which gives a stronger response while does not change the patterns. The unchanged patterns demonstrate both the robustness of the patterns, as well as that the absolute threshold levels are flexible with acceptable ranges of about 5-75 for DWCI and -1.0-0.0 for \( C^*_ET \). This analysis is now discussed in the results section with the following text:

These thresholds were chosen to highlight frequency differences between sites and were shown to have large metric responses under dry conditions while having low frequencies under wetter conditions (see sensitivity analysis in supplementary figure S2). Furthermore, these thresholds
results in a similar median frequency of uncoupled and morning shifted days between all site-years being 8.7% and 9.4% of days respectively. The similarity in median frequencies across site-years allowed for easier inter-comparison between the two metrics. The frequency of decoupling and morning shifts using these thresholds for each site can be found in the map found in File S1.

8. Figure 3: The text is small and hard to read.

The text size has been increase to improve readability (Figure R5).

9. Page 13, Lines 15-25: I found this discussion of the links between C_ET and isohydricity to be highly speculative, notably because isohydricity tends to describe plant response to declining soil moisture, yet the afternoon stomatal closure may be largely caused by increasing VPD.

To further the points discussed above, the VPD responses are likely tied to the concept of isohydricity. The review by Martínez-Vilalta and García-Forner[6] makes the point that VPD and transpiration dynamics feed back into the rate of soil moisture depletion and how fast a plant will reach the point of hydraulic failure. Discussion of these points have been added to highlight these feedbacks, and the section now reads:

In this way, it seems that though $C^*_{ET}$ is less noisy as a drought indicator (see Fig. 6), it may only be of use in tree systems that are more prone to hydraulic stress. However, this does put the metric in a rather unique position in that it could be used as a global scale hydraulic indicator, having potential application in exploring ecosystem level isohydricity[6]. Isohydricity is intrinsically a concept that relates to an individual plant, as dynamics of rooting depth, hydraulic conductances, and sensitivities to VPD can vary within individuals of the same species at the same location. However, these factors are all interrelated, as hydraulic and stomatal conductances drive transpiration dynamics which control the rate of depletion of root zone water which can then feed back to stomatal sensitivity, such as via ABA signaling[10].

References

Figure 4: Figure S1. Sensitivity analysis of DWCI and $C_{ET}$ to evaporative fraction (EF=ET/potential ET) under different GPP and air temperature (Tair) values, as well as the sensitivity of frequency of decoupling (DWCI < threshold) and morning ($C_{ET}$ < threshold) using various thresholds. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and $C_{ET}$ of -1-1 indicate one hour morning shifted-one hour afternoon shifted ET. Bins for GPP and Tair based on percentiles to give an equal number of points per bin, with the exception of GPP<1 and Tair<0, which are approximately double the size of the other bins and represent periods of low plant activity such as winter. Vertical bars represent interquartile range in all cases.
Figure R 5: The frequency of morning-shifted Diurnal Centroids ($C^*_ET < -0.25$ hours, upper panels a,b) and low diurnal water carbon index (DWCI<25, lower panels c,d) for 690 fluxnet site-years/192 sites, grouped by climate group (left panels a,c) and plant functional type (right panels b,d). Group labels on x-axis indicate the number of site-years/sites (n=site-years/sites) for each category. Dashed line is the median for all site-years. Color shade indicates level of significance, with light colors and dark colors having p-values <0.10 and <0.05 respectively (Wilcoxon–Mann–Whitney two-sample rank-sum test), red and blue colors indicate distributions higher and lower respectively compared to data from all sites excluding the group. Only sites-years with at least 20 data points and groups with more than 5 site-years were included.


Water stress induced breakdown of carbon-water relations: indicators from diurnal FLUXNET patterns

Jacob A. Nelson, Nuno Carvalhais, Mirco Migliavacca, Markus Reichstein, Martin Jung

Abstract

Understanding of terrestrial carbon and water cycles is currently hampered by an uncertainty in how to capture the large variety of plant responses to drought across climates, ecological strategies, and environments. In FLUXNET, the global network of CO2 and H2O flux observations, many sites do not uniformly report the ancillary variables needed to study drought response physiology such as soil moisture, sap flux, or species composition. In this sense, the use of To this end, we outline two data-driven indicators based on diurnal energy, water, and carbon flux patterns to derive clues on ecosystem water limitation responses at a daily resolution could prove valuable, if nothing less than a benchmark to test current hypotheses. To this end, we propose two data-driven indicators derived directly from the eddy covariance data and based on theorized physiological responses to hydraulic and non-stomatal limitations. Hydraulic limitations (i.e. intra-plant limitations to water movement) are proxied using the relative diurnal centroid ($C^{\ast}_{ET}$), which measures the degree to which the flux of evapotranspiration (ET) is shifted toward the morning. Non-stomatal limitations (e.g. inhibitions of biochemical reactions, Rubisco ($\text{RuBisCO}$) activity, and/or mesophyll conductance) are characterized by the Diurnal Water:Carbon Index (DWCI), which measures the degree of coupling between ET and gross primary productivity (GPP) within each day. As a proof of concept we show the response of the metrics at 6 European sites during the 2003 heatwave event, showing varied response of morning shifts and decoupling. Globally, we found indications of hydraulic limitations in the form of significantly high frequencies of morning shifted days in dry/Mediterranean climates and savanna/evergreen plant functional types (PFT), whereas high frequencies of decoupling were dominated by dry climates and grassland/savanna PFTs indicating a prevalence of non-stomatal limitations in these ecosystems. Overall, both the diurnal centroid and DWCI were associated with high net radiation and low latent energy typical of drought. Using three water use efficiency (WUE) models, we found the mean differences between expected and observed WUE to be $-0.59 \text{ to } -0.44 \text{ to } -0.44 \text{ umol/mmol}$ and $-0.56 \text{ to } -0.69 \text{ to } -0.40 \text{ umol/mmol}$ for decoupled and morning shifted days respectively compared to mean differences $-1.4 \text{ to } -1.71 \text{ to } -1.41 \text{ umol/mmol}$ in dry conditions. These results suggest, suggesting that morning shifts/hydraulic responses are associated with an increase in WUE whereas decoupling/non-stomatal limitations are not.
Introduction

Processes such as photosynthesis and transpiration are so intimately linked that knowledge and assumptions about one process are needed to accurately understand the other. Unfortunately, the relationship between carbon and water cycles is not fully understood [48], passing the biases and uncertainties caused by an incomplete carbon:water assumption framework back onto flux estimates specifically and global water and carbon cycle interactions and dynamics in general [18, 47, 15]. One source of uncertainty that is increasingly being identified is the diverse responses of plants to water limitation [56, 9, 45], which hampers the understanding and predictability of water and carbon cycles during drought. Here we outline potential causes of uncertainty in carbon:water dynamics in an effort to outline data-derived inductors based on current theory.

Stomata, WUE, and the carbon:water relationship

Classically, vegetation water and carbon fluxes are linked by stomates, where an open stomate allows CO2 to enter the leaf and, consequentially, water is lost. From this, most theoretical frameworks make some form of assumption that carbon assimilation (A) and water losses (T) are both contingent primarily on leaf stomatal conductance (gs). This assumed relationship allows us to pass between the realms of carbon and water, based on the assumption that at any given time both A and T are proportional to the stomatal conductance multiplied by the difference in internal and external CO2 and water vapor concentrations. More specifically,

$$A = g_s \cdot \Delta c \quad \text{and} \quad T = 1.6 \cdot g_s \cdot \Delta v$$

where $\Delta c$ and $\Delta v$ are the differences in inner and outer stomatal cavity concentrations of CO2 and water vapor, respectively.

By making some assumptions on how stomates react to environmental variables and how these changes affect water and carbon concentrations, we can come to a-These diffusion equations lead to the relatively consistent carbon:water ratio, generally expressed as a water use efficiency ($WUE = A/T$). At the ecosystem level where direct measurements of A and T are not available, WUE is simply calculated as the ratio of gross primary productivity (GPP) to total evapotranspiration (ET) [22]. Further, the effect of These carbon:water links are fundamental to understanding how stomata are regulated and underly key functioning in mechanistic plant and ecosystem models. One such set of models are those based on optimality theory which posit that plants tend to optimize carbon gains to water losses, such
as those described by Katul et al. [17] and Katul, Palmroth, and Oren [16]. These concepts from Katul, which carry the assumptions of RuBisCO (light) limitation, were built upon by Zhou et al. [55] and Zhou et al. [53] to give the equation,

$$uWUE = \frac{GPP \cdot \sqrt{VPD}}{ET}$$

where the $\sqrt{VPD}$ accounts for the stomatal response to vapor pressure deficit (VPD) on stomatal response can be taken into account giving formulations such as,

$$iWUE = \frac{GPP \cdot VPD}{ET} \quad \text{and} \quad uWUE = \frac{GPP \cdot \sqrt{VPD}}{ET}$$

for the intrinsic (iWUE) and underlying water use efficiencies (uWUE) respectively, both of which tend to be more constant than WUE assuming stomatal response optimizes carbon gain to water losses. Accounting for the VPD response allows for a more stable metric of WUE that is temporally more stable and physiologically more meaningful, such as when comparing the diurnal cycles of carbon and water. As ET is the sum of both T and non-biological evaporation, often periods during and shortly after rain events are excluded from WUE estimates to minimize the influence of non-plant evaporation. Ultimately, calculations of WUE provide a simple summary of the cost in water per carbon gain and becomes an indicator for how plants have and will adapt to the physical limitations of their changing environments [18, 48].

**Uncertainty during drought**

Though assuming a rigid carbon:water relationship works well in conditions when ecosystems are moderately wet, conditions associated with the majority of carbon and water fluxes, an inflexible carbon:water assumption is unsatisfactory in that these assumptions may breakdown as plants shift from light to water limitations. Indeed, in a review of leaf level stomatal conductance models, Damour et al. [7] concluded that the majority of stomatal models fail to adequately capture the effects of drought. This failure to capture the effects of drought is not only disconcerting as this is water limited conditions are when ecosystems are most at risk, but this an incomplete framework tends to propagates errors and uncertainties from models into estimates of the water and carbon cycles. For instance, these uncertainties are reflected in one recent meta-analysis of T to ET partitioning studies which included EC, sap flow and isotopic methods, which puts the state of the art estimated global T/ET at a notably uncertain 61% ±15% s.d.. Similarly, using estimates from models based on global remote-sensing products, Miralles et al. [53] found global T/ET estimates varied from 24-76%. This high uncertainty is disconcerting, as transpiration is the largest and most dynamic component of the global water
cycle, which likely controls the variability in global ET.

In addition to the uncertainty in water cycles, estimates of are GPP also sensitive to carbon:water assumptions. In outlining a road map for improved modeling of photosynthesis, Rogers et al. [45] noted as key recommendations both improving information about water:carbon relations (in the form of the stomatal slope parameter $g_1$) as well as improving understanding of the response of carbon assimilation to drought. Similarly, in an analysis of parameter uncertainties for a terrestrial biosphere model, Dietze et al. [9] found that two of the top five parameters contributing to the predictive uncertainty of net primary productivity were associated with plant water regulation. This uncertainty is reflected in the stomatal conductance parameterization exercise from Knauer, Werner, and Zaehle [19], where the authors were able to improve model performance in predicting EC measured GPP and ET by including atmospheric effects (in the form of VPD) on stomatal conductance, but concluded that further improvement required global understanding of water limitation response variation across plant functional traits and growing conditions, which is currently unavailable.

**Sources of uncertainty**

Two ideas to account for the errors in carbon:water assumptions under dry conditions have begun to emerge: that hydraulic limitations in transporting water from root to leaf change stomatal responses and thus limit transpiration under high demand, or that changes in the intra-leaf processes of carbon transport and fixation under drought conditions result in non-stomatal limitations that impact carbon assimilation independently of water fluxes [35].

**Hydraulic limitation**

As soil water potentials in the root zone become increasingly negative, the long-term plant strategy may turn from optimizing carbon fixation to preventing damage to hydraulic architecture [49]. As such, stomata and transpiration are likely to increasingly respond not just to atmospheric conditions, but also soil moisture. Under this hydraulic limitation framework, a plant will be reacting to the inability to transport water, even though the key control mechanism for a plant is via the stomata, possibly expressed as an increase in sensitivity. Such assumptions are consistent with the mechanisms encoded in some land surface and ecosystem models, which account for water limitations in LSMs, by scaling the water to carbon ratio in relation to available soil moisture. Though this method should link the leaf physiology to the soil and thus capture some hydraulic limitation, it has been criticized for not capturing the variety of drought responses found in different plant species and ecosystems [8]. This diversity in plant responses has been pointed to as a key point of uncertainty in earth system models [9].

**Non-stomatal limitation**
Though ecosystem water and carbon fluxes are predominantly controlled by stomates, non-stomatal or bio/photocatalytic inhibitions to carbon assimilation are worth considering as they have the capacity to decouple the water-carbon exchange. This decoupling could include conditions where the stomates are transpiring water but intra-leaf factors are slowing carbon fixation, changing the intrinsic water use efficiency directly. Intra-leaf factors could include effects such as production of reactive oxygen species [23]; environmental limitations to the photosynthetic pathways, such as leaf temperature [31]; or declines in mesophyll conductance [11]. Non-stomatal limitations have been observed at ecosystem scale [43, 32], though the exact mechanism is difficult to elucidate [41]. These effects likely vary between species, as well as with the rate of onset of drought, access to water, and other environmental conditions.

Objectives

There seems to be a collective conclusion that the breakdown of carbon:water assumptions needs to be better characterized in general, and specifically for implementation in modeling frameworks [De Kauwe et al. [8]; Manzoni [25]; Zhou et al. [56]; Flexas et al. [11]; Egea, Verhoef, and Vidale [10]]. Though the problem is becoming clear, the way forward is hampered by an uncertainty in how to capture the large variety in the response to drought across climates, strategies, and species. In this sense, the use of EC measured diurnal patterns of carbon, water, and energy fluxes to derive clues on ecosystem drought responses at a daily resolution could prove valuable, if nothing less than a benchmark to test current hypotheses both as a means to identify potential periods of ecosystem stress, inform machine learning algorithms on ecophysiological conditions not found in environmental variables, as well as benchmarking a models ability to capture sub-daily dynamics. To this end, we propose two data-driven indicators of water stress, the diurnal water:carbon index (DWCI) and the relative diurnal centroid in LE ($C_{ET}^*$). Both metrics are derived directly from the EC data and based on expected physiological responses to hydraulic and non-stomatal limitations. Using these data-driven indicators we then characterize the distribution of these limitations across a global spread of climate and vegetation types. Finally, we explore the ability of these indicators to detect the disagreements between modeled and observed water use efficiency, and explore how these biases may be attributed to hydraulic and non-stomatal limitations.

Methods and Materials

Data

Carbon, water, and energy fluxes measured with EC, as well as meteorological data, were obtained from the FLUXNET database [12]. Half-hourly latent heat and net ecosystem exchange (NEE) fluxes were collected and processed using standard QA/QC procedures [37], gap-filling and partitioning algorithms [42]. From the database, half-hourly gross primary productivity (GPP) and ET data (derived from latent heat
flux measurements) were downloaded and used for the following analysis. The list of selected sites and an interactive map of sites used can be found in File S1.

In order to provide a consistent measure of ecosystem dryness that can be utilized across sites, the ratio of water evaporated to potential water evaporated was calculated as evaporative fraction (EF), or the fraction of actual ET to Potential ET (PET). PET was calculated as the daily fraction between the measured ET and estimated ET via a Priestley-Taylor model [39] using site measured net radiation (Rn) and air temperature (T_{air}). The slope (alpha parameter) was fit for each site-year using 95th quantile regression [20] instead of using the original 1.26 value derived for a “well watered crop”. Evaporative fraction (EF) was calculated as the fraction of actual ET for PET [39].

In order to get high quality data and minimize the influence of abiotic evaporation (hereafter just evaporation), all data was filtered with the aim to include only non-gap filled data in the growing season with dry surface conditions. Specifically, days with GPP < 5 g·m^{-2}·d^{-1} or Growing season was defined as all days where GPP > 1 gC·m^{-2}·d^{-1} and daily mean air temperature $< 15 > 5 ^\circ$C were removed. These threshold were shown to give good response in the proposed metrics while minimizing variability due to low diurnal signals, a sensitivity analysis of which can be found in supplementary Figure S2. In an effort to minimize contributions of evaporation, the conservative soil wetness index (CSWI) was employed which was designed to force a positive water storage for any time step with any amount of precipitation. So by calculating estimate whether the ecosystem is likely to have “dry” surfaces and therefore ET is likely to be dominated by transpiration. This approach requires a certain amount of evaporation to occur after a rain event before the surface is considered to be “dry” and can be contrasted to the method of removing a set time period after rain employed in previous studies [30, 2, 18]. CSWI is calculating by first quantifying the storage at time t ($S_t$) as,

$$
S_t = \min (S_{t-1} + P_t - ET_t, S_o)
$$

$$
S_t = \min (S_{t-1} + P_t - ET_t, S_o)
$$

(3)

where $ET_t$ and $P_t$ are the ET and precipitation at time-step t respectively, $S_t$ is effectively capped at a maximum storage value of $S_o$, which was set to 5 mm. Furthermore, to make the metric conservative in regards to assumed water inputs, any precipitation event will refill the storage from 0 mm,

$$
CSWI = \max (S_t, \min (P_t, S_o))
$$

$$
CSWI = \max (S_t, \min (P_t, S_o))
$$

(4)

which has the effect of requiring all precipitation up to 5 mm to be evaporated from the system before negative storage
can occur. Any gaps in the precipitation data were assumed to be a precipitation event of 5 mm in order to prevent any unmeasured precipitation from biasing the results by inadvertently including rainy days. Code and further outline of the algorithm can be found in file S2 File S3 as well as at Nelson [34]. Evaporation was assumed to be negligible when CSWI<0. This method was used over the more standard method of removing 1-5 days after a rain event, as it does not make the assumption that the surface will dry in a fixed amount of time, instead relying on a minimum amount of evaporation ET. As a comparison, the median time period for the CSWI to go from fully wet (CSWI=5) to “dry” (CSWI<=0) was 3.5 days across all sites in summer, where summer was defined as the period when daily potential radiation above median daily potential radiation for each site.

The data filtering as outlined in this section was designed to isolate periods firmly in the growing season when plants are active and the signal of ET is most likely to be dominated by plant controls.

Relative diurnal centroid (\(C_{ET}^r\))

As soils dry, it becomes more difficult to transport stem and root zone moisture to the leaf, potentially causing hydraulic limitations. This for the plant to transport water. This shift was seen in eddy covariance data in a study by Wilson et al. [51], who examined the shift of latent compared to sensible heat, which suggested that a shift in water fluxes towards dawn can be indicative of afternoon stomatal closure. Shifts were further explored in a modeling study by Matheny et al. [29] which found that the morning shift was not well captured by models and attributed the errors to inadequate hydraulic limitations in the models. The daily cycle of wetting and drying acts as a capacitor in the hydraulic circuit, allowing water stores to be more easily transported in the morning and depleting in the afternoon. As bulk soil moisture declines, this effect may be strong enough to shift the diurnal cycle of ET significantly toward the morning. This phenomenon can be quantified. Quantifying diurnal shifts in ET data using the diurnal centroid was first explored by Wilson et al. [51]: defined as the flux weighted mean hour, or

\[
C_{flux} = \frac{\sum \text{flux}_i \cdot t}{\sum \text{flux}_i}
\]

\[
C_{flux} = \frac{\sum \text{flux}_i \cdot t}{\sum \text{flux}_i}
\]

(5)

where t is a regular, sub-daily time interval (here t measures as decimal hour at half-hourly time-step). The resulting \(C_{flux}\) is the weighted mean hour of the diurnal cycle of that particular flux for that particular day. For example, if a calculated \(C_{ET}\) for a given day (using measurements at an hourly interval) equals 12 of decimal hour) equals 12.25, this would entail that the weighted mean for that day is at 15 minutes past noon. Figure 1 shows an example of the shifts in
the monthly average cycle from a wet month to a dry month. As analyzed in Wilson et al. [51], a shift in ET towards dawn can be indicative of afternoon stomatal closure. In order to isolate a shift, we then had to control for variations in global radiation (Rg), both fluctuations due to clouds and differences in the timing of solar noon. Therefore, the difference between the diurnal centroids of ET ($C_{ET}$) and Rg ($C_{Rg}$) was calculated as

\[
C^*_{ET} = C_{Rg} - C_{ET}
\]

\[
C^*_{ET} = C_{Rg} - C_{ET}
\]

(6)

giving $C^*_{ET}$ as the diurnal centroid of ET relative to Rg. The resulting values of $C^*_{ET}$ are not tied to the carbon cycle, which can be affected by non-stomatal limitations and generally shows a more prominent midday depression. Annotated code for this CSWI calculation can be found in file S3, File S4 as well as at Nelson [34]. Though a diurnal centroid can be calculated for any diurnal cycle, basing a metric on the morning shift of ET relative to Rg has the advantage of targeting the non-atmospheric drivers of the water flux, of which there are few ancillary variables.

Figure 1: One month average cycle (solid lines) and accompanying diurnal centroid (vertical dashed lines) of incoming shortwave radiation (Rg), evapotranspiration (ET), and gross primary productivity (GPP) at the Peuchabon, France (‘FR-Pue’) site during 2003. May is relatively wet (32 mm rainfall, left) and July is relatively dry (0 mm rainfall, right). While ET and Rg correspond well in the wet month, the dry month shows a distinct phase shift in both GPP and ET fluxes towards the morning, as well as a midday depression in GPP.
Diurnal water carbon index (DWCI)

If transpiration and carbon assimilation are predominantly controlled by stomatal conductance, it follows that their diurnal cycles should be largely in sync. In other words, regardless of a plant’s maximum T or A, if the stomates start to close, both rates should decrease by a similar percentage. On the other hand, non-stomatal limitations that inhibit carbon assimilation independent of water have the capability to alter the diurnal cycle on just one flux, causing them to decouple. In an effort to quantify the degree of carbon:water coupling for an individual day, we examined the relationship of GPP and ET, where,

\[
\text{ET} \propto GPP \cdot \sqrt{VPD}
\]

or,

\[
\text{ET} = i \cdot GPP \cdot \sqrt{VPD}
\]

This relationship incorporates the assumption that, at least over short time scales, the amount of carbon that enters the leaf is proportional to the amount of water that leaves, and also incorporates the non-linear response of stomates to VPD [17, 16, 55]. This model, though simple, has been shown to work well across a variety of EC sites [53]. Figure 2 (upper panels a,b) shows a comparison between the daily cycles in a wet and dry month. By calculating a daily correlation between the normalized daily cycles of ET and \( GPP \cdot \sqrt{VPD} \), we come to a correlation coefficient for each day (see Figure 2, lower panels c,d). For well watered days in the growing season the two signals tend to be well correlated \((\rho > 0.9)\), but tends not to be less correlated in periods of stress, a comparison of which can be seen in Figure 2 (lower).

As it is, this daily correlation coefficient is dependent on the signal strength, or magnitude, of the flux. Low correlation values could just as easily be from carbon:water decoupling as to a low signal to noise ratio. Therefore, to produce a more robust metric and account for these statistical decreases in correlation, we turned the daily correlation coefficient into an index based on its rank in a distribution of correlation coefficients from artificial datasets. These artificial datasets are constructed using the diurnal signal from potential radiation, with Gaussian noise \( \mathcal{N}(0, \sigma) \) added according to the standard deviation random uncertainty of the ET and NEE fluxes, or
\[ f_{\text{flux } \text{artificial}} = \frac{R_{\text{pot}}}{R_{\text{pot}}} \cdot f_{\text{lux}} + N(0, \sigma_{\text{flux}}^2) \]

\[ L_{\text{E } \text{artificial}} = \frac{R_{\text{pot}}}{R_{\text{pot}}} \cdot L_{\text{E}} + N(0, \sigma_{\text{LE}|\text{NEE}}^2) \]

(9)

and

\[ N_{\text{EE } \text{artificial}} = \frac{R_{\text{pot}}}{R_{\text{pot}}} \cdot N_{\text{EE}} + N(0, \sigma_{\text{NEE}|\text{LE}}^2) \]

(10)

Uncertainties of the NEE and ET fluxes were estimated from the gap filling procedure of \textit{reichstein_separation_2} with the uncertainty equal to the standard deviation of flux measurements within a time window and similar meteorological conditions. As GPP is calculated from gap-filled values of NEE, the uncertainty from NEE was used for GPP. Furthermore, the correlation structure between the noises in LE and and NEE was preserved in the artificial dataset.

In essence, by using the underlying signal from potential radiation, both the artificial ET and GPP \( \sqrt{\text{V PDT}} \) are perfectly correlated when no noise is added. Adding noise then isolates the decoupling effect of signal to noise ratio. An artificial correlation coefficient can then be calculated from the two artificial datasets in the same manner as from the real dataset, and this experiment is repeated 100 times for each day, giving a daily distribution of artificial correlation coefficients. The rank of the real correlation coefficient in the distribution from the artificial set gives a probability that the carbon and water signals are actually coupled. The resulting index has a range of 0-100, with 100 indicating that the real correlation coefficient was greater than the entire artificial set, and therefore it is very likely that carbon and water are coupled. From this index we can now quantify if the water and carbon signals are coupled for any given day, and therefore shed light onto whether the two fluxes are only controlled by the opening and closing of stomates. Annotated code for this calculation can be found in file S4 as well as at Nelson [34].

Models and parameter estimation

To benchmark whether these metrics are capturing information that is possibly not being captured in modern model frameworks, three simple models were used as benchmarks of physiological understanding. The three models were used to estimate WUE (GPP/ET) for each day at each site and compared to actual flux data. The purpose of the exercise was to evaluate if bias in the model predictions were associated with decoupled or morning shifted days, thus indicating that the metrics are corresponding to information that the models are unable to capture. Here we utilize three models to provide a spectrum of theoretical to empirical basis. The "Katul Katul-Zhou" model, as defined and used in calculation of...
Figure 2: **Theoretical overview of diurnal water carbon index** upper panels: One month average diurnal cycle of incoming shortwave radiation (Rg), evapotranspiration (ET), vapor pressure deficit (VPD), gross primary productivity (GPP), and underlying water use efficiency (uWUE, $GPP \ast VPD^{-0.5}$) at the Peuchabon Forest, France ('FR-Pue') site during 2003. Discrepancies between uWUE $GPP \ast VPD^{-0.5}$ and ET increase from the relatively wet May (32 mm rainfall, left) to the relatively dry July (0 mm rainfall, right). lower panels: These discrepancies are reflected in the daily correlation values between uWUE $GPP \ast VPD^{-0.5}$ and ET, giving an indication of the appropriateness of the uWUE model for each day, as well as the degree of coupling between water and carbon signals.
the DWCI, is based in stomatal optimization theory [17, 16, 53], which makes the assumption that the WUE is constant if
corrected by the effect of VPD, using an inverse square root as the assumed relationship. Though the constant nature of
uWUE may not be correct, with the optimal carbon cost of water changing over day or weeks [26, 36], a yearly parameter
of uWUE was estimated which is consistent with other modeling exercises [54]. One step away from a theoretical basis
is a revision of this model by Boese et al. [3], the “Boese” model, where an additional radiation term was added such that,

\[
ET = i \cdot GPP \cdot \sqrt{VPD} + r \cdot Rg
\]

(11)

where i and r are parameters fit to each site-year. This relationship with Rg was shown to have a better predictive per-
formance for EC data from 115 sites [3]. The interpretation of this extra radiation term is not clear and is difficult to
reconcile with the current understanding of physiology. It is possible the term could be related to biophysical effects,
e.g. VPD at leaf surface vs the measured ambient VPD. Nevertheless, the Boese model is an empirical and ecosystem
scale model that complements the theoretical and originally leaf-level model from Katul-Zhou.

Parameters of these models were estimated for each site-year. The Boese model parameters were fit using trimmed
least squares regression (TLS) which minimizes the 90th percentile of SSE to prevent influence of large outliers [46,
44]. As the error in both ET and GPP are assumed to be of similar magnitude, the i parameter in the Katul-Katul-Zhou
model was calculated using geometric mean regression, where the final slope was calculated as the geometric mean of
the parameters from

\[
\begin{align*}
ET &= i_{GPP} \cdot GPP \cdot \sqrt{VPD} \\
\text{and } GPP \cdot \sqrt{VPD} &= \frac{ET}{i_{ET}}
\end{align*}
\]

(12)

Because both the Katul–Both the Katul-Zhou and Boese models are anchored in a theoretical understanding of leaf
physiology, their inability to make accurate predictions may be a result of a failure of their underlying assumptions
theoretically based and here implemented have the underlying assumptions of RuBisCO-limited conditions and constant
carbon cost of water throughout the season which may not reflect reality. Therefore a fully empirical and highly non-
linear model can give insight into how much information is actually stored in the data while minimizing any assump-
tions. As a fully empirical model, a random forest regression (RandomForestRegressor from Pedregosa et al. [38] based
on Breiman [4]) was fit to half-hourly ET data for each site using Rg, VPD, Tair, GPP and year as input parameters.
Values were estimated using 50 trees with predictions made using out-of-bag estimates to prevent over-fitted model predictions.

**Results**

![Graph showing monthly median diurnal water carbon index (DWCI, lower panels) and diurnal centroids (C*T, upper panels) for 6 sites in Europe. Data from all years available (black) is compared to 2003 (red) during which a drought event resulted in high temperatures and low precipitation throughout the summer. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and C*T of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Vertical bars represent interquartile range. Sites from 4 plant functional types: evergreen broadleaf (EBF), deciduous broadleaf (DBF) and evergreen needleleaf (ENF) forests, as well as grasslands (GRA). Ecosystems show tendencies of morning shifts (e.g. DK-Sor and NL-Loo) and carbon:water decoupling (e.g. ES-ES1 and HU-Bug) during the drought year.](image)

Figure 3: Monthly median diurnal water carbon index (DWCI, lower panels) and diurnal centoids (C*T, upper panels) for 6 sites in Europe. Data from all years available (black) is compared to 2003 (red) during which a drought event resulted in high temperatures and low precipitation throughout the summer. Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon:water coupling and C*T of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Vertical bars represent interquartile range. Sites from 4 plant functional types: evergreen broadleaf (EBF), deciduous broadleaf (DBF) and evergreen needleleaf (ENF) forests, as well as grasslands (GRA). Ecosystems show tendencies of morning shifts (e.g. DK-Sor and NL-Loo) and carbon:water decoupling (e.g. ES-ES1 and HU-Bug) during the drought year.

As a case study, C*T and DWCI time-courses for six sites from Europe are shown in Figure 3, with an emphasis on 2003 when the continent was struck by a heatwave that was shown to effect both the carbon and water cycles [6, 40, 13]. For DWCI, forest sites showed high water:carbon coupling throughout the growing season, with the exception of Puechabon (FR-Pue) which showed a regular seasonal cycle of decoupling. The grassland site (HU-Bg) showed a higher variability in DWCI compared to the forest sites (all others). All sites showed either a decrease in median DWCI or an increase in variability during 2003, generally in July or August, particularly at Hainich (DE-Hai), Bugacpuszta (HU-Bug), and El Saler (ES-ES1). This increase in decoupling during 2003 is consistent with the hypothesis of non-stomatal limitations being expressed in hot, dry conditions which can affect carbon fixing mechanisms. Median diurnal centroid values across all years varied in absolute magnitude, but were generally near or above zero, i.e. the water cycle showed no shift or an afternoon shift. One exception would be the Mediterranean oak forest of Puechabon, which shows a slight seasonal cycle of morning shifts going from a slight afternoon shift to a slight morning shift during June, July, and August. During drought years, sites that showed distinctive morning shifts were Puechabon (FR-Pue), Soroe (DK-Sor), and Loobos (NL-Loo). The framework that morning shifts are associated with water stress from soil moisture depletion would be
supported by the increase in morning shifts during 2003, though factors such as species composition and access to soil water would play a significant role and could account for the differences among sites. All sites had significantly different (p<0.05, Wilcoxon rank-sum test) DWCI values between 2003 and all other years except Puechabon, whereas with $C_{ET}^*$ only Puechabon, Soroe, and Loobos showed significant differences.

Distribution of data driven indicators by vegetation type and climate

Figure 4: The frequency of morning-shifted Diurnal Centroids ($C_{ET}^*$< -0.25 hours, upper panels a,b) and low diurnal water carbon index (DWCI<25, lower panels c,d) for 690 Fluxnet site-years/192 sites, grouped by climate group (left panels a,c) and plant functional type (right panels b,d). Group labels on x-axis indicate the number of site-years/sites (n=site-years/sites) for each category. Dashed line is the median for all site-years. Color shade indicates level of significance, with light colors and dark colors having p-values <0.10 and <0.05 respectively (Wilcoxon–Mann–Whitney two-sample rank-sum test), red and blue colors indicate distributions higher and lower respectively compared to data from all sites excluding the group. Only site-years with at least 20 data points and groups with more than 5 site-years were included.

The frequency of low values of diurnal centroid and DWCI across climate groups and plant functional types is shown in Figure 9. The thresholds of -0.5 and -0.25. The thresholds designating decoupling and morning shifts were 25 and -0.25.
for DWCI and $C_{ET}^*$ respectively were chosen because: when DWCI < 10 it is reasonably likely that carbon and water are
decoupled, and sites under water stress tended to have $C_{ET}^* < 0.5$. Using these thresholds we find that 7% and 8% of
data points across all sites are decoupled and morning shifted respectively. These thresholds were chosen to highlight
frequency differences between sites and were shown to have large metric responses under dry conditions while having low
frequencies under wetter conditions (see sensitivity analysis in supplementary Figure S2). Furthermore, these thresholds
result in a similar median frequency of uncoupled and morning shifted days between all site-years being 8.7% and 9.4%
of days respectively. The similarity in median frequencies across site-years allowed for easier inter-comparison between
the two metrics. The frequency of decoupling and morning shifts using these thresholds for each site can be found in
the map found in File S1. Though there is a fairly large variance across climate groups and plant functional types, low
values of both DWCI and $C_{ET}^*$ occur at higher frequencies in savanna ecosystems and dry or Mediterranean climates.
Conversely, lower frequencies of both metrics are seen in tropical, boreal, and temperate-continental climates. Strikingly,
the arid and semi-arid climate group seems to be associated with the majority of low DWCI occurrences, with a median
frequency of about 20% of days being uncoupled between site-years. Overall, frequencies were highly variable within
plant functional types. Interestingly, $C_{ET}^*$ seems to be more variable in moderately dry ecosystems with potentially deep
roots, favoring woodier savannas and evergreen broad-leaf–needle-leaf forests over grasslands and open shrub lands.

In contrast, DWCI shows similarly high frequencies from savannas and grasslands. Strikingly, the arid and semi-arid
climate group seems to be associated with the majority of low DWCI occurrences, with a mean frequency of about 15%
of days being uncoupled. The differing responses between tree and grass dominated ecosystems can be further seen in
Figure 5, where savanna and grassland ecosystems show a distinct decrease in DWCI under conditions of low EF, in
contrast to the forested sites which show a higher degree of carbon–water coupling, though still a slight decrease. Forested
ecosystems show a higher degree of morning shift under low EF conditions when compared to grasslands, with savannas
being somewhere between the two.

The response of both variables to drought stress is further observed in Figure 46, where low mean values of both DWCI
and $C_{ET}^*$ are associated with conditions of high net radiation and low latent energy, indicative of drought. As this figure
includes all days from all sites which meet the filtering outlined in the Data subsection of the Methods, i.e. dry periods
in the growing season, these figures exhibit the universality of the metrics across climates, ecosystems, and time periods.

This pattern is much cleaner with the diurnal centroid than with DWCI, though mean values are generally above 50 for
most bins, indicating that most days are well coupled. Low values of both indicators are also seen under conditions with
low Rn and high latent energy (as seen by the dark streak at the top edge of the left plate in Figure 4 in Figures 6c,e),
which is generally not associated with drought stress. Further analysis showed that these points are also associated with
energy balance over closure, where the sum of latent and sensible heat is greater than net radiation (ET+H>Rn, see Figure
S8) and therefore likely represent a data problem rather than a physiological response. Removing all days where the
Figure 5: The frequency of morning-shifted Diurnal Centroids (CET< -0.5 hours DWCI, upper figure panel) and low diurnal water carbon correlation centroid (DWCI<10°CET, lower figure) for 651 fluxnet site-years/189 sites, grouped by climate group (left panel) and of plant functional type functional types binned by evaporative Fraction (right EF, low values indicate dry conditions). Group labels on x-axis Note DWCI of 0-100 indicate the number lowest-highest probability of site-years/sites diurnal carbon:water coupling and CET of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Evergreen needleleaf (n=site-years/sites ENF) for each category. Dashed line is the median for all site-years combined. Color shade indicates level of significance, with light colors and dark colors having p-values <0.10 and <0.05 respectively deciduous broadleaf (Wilcoxon–Mann–Whitney two sample rank sum test DBF), red and blue colors indicate distributions higher and lower respectively evergreen boradleaf (EBF) forests show increased morning shifts (low CET) with decreasing EF when compared to data from all sites excluding the group. Only-grassland (GRA) sites which tended to have decreased carbon:water decoupling (low DWCI) with at least 20 data points decreasing EF, Savanna ecosystems (SAV) show a high degree of decoupling and groups with more than
energy balance is over closed did not alter the patterns associated with drought. Apart from the response to periods of high LE and low Rn, the metrics showed diverging response when looking at EF (ET/PET which is similar to LE/Rn) and VPD, with DWCI showing a much stronger response to VPD and $C_{ET}^*$ showing a much stronger response to EF (Figure 6a,d). This difference in response would indicate that DWCI is more responsive to atmospheric demand (estimated via VPD) and $C_{ET}^*$ is more responsive to water limitations. Both DWCI and $C_{ET}^*$ also show a trend with low GPP, although in the case of the diurnal centroid the effect is limited to both low GPP and ET (Figure 6c,g).

Figure 6: Mean DWCI (upper panels) and $C_{ET}^*$ (lower panels) with respect to evaporative fraction (EF) by vapor pressure deficit VPD (a,d), latent energy (LE) by Rn (left b,e) and LE by GPP (right c,g). Note DWCI of 0-100 indicate lowest-highest probability of diurnal carbon-water coupling and $C_{ET}^*$ of -1 to 1 indicate one hour morning shifted to one hour afternoon shifted ET. Points with high Rn and low LE are associated with both low DWCI and $C_{ET}^*$, indicating that both metrics are related to water limitations. Though both metrics are associated with low EF, DWCI shows a much higher response to atmospheric demand as measured by VPD, with $C_{ET}^*$ showing very limited response. Both metrics, and DWCI in particular, show low values with high ET and low Rn, though these points are also associated with over closed energy balances (LE+H>Rn-GRn). Both metrics are associated with low GPP, but the $C_{ET}^*$ is restricted to both low GPP and ET, indicating water and carbon can decouple over a wider range of water stress. This also holds when points with energy balance over-closer are excluded (data not shown).

**Difference between modeled and actual WUE**

Figure 5-7 shows the difference between expected and observed WUE from the Katul-Katul-Zhou, Boese, and random forest (RF) models, with respect to conditions of drought as characterized by low evaporative fraction (EF<0.2), C:W decoupling (DWCI<0.25), and morning shifts ($C_{ET}^*$<0.25). This exercise was designed to test whether the metrics were associated with bias in the models, indicating that the metrics are able to capture information that the models are not (as further outlined in Methods and Materials subsection “models and parameter estimation”). For all models, the dry days show the largest average shift between expected and observed WUE, followed by morning shifted days. Uncoupled
Figure 7: Difference in modeled and measured WUE for Katul-Zhou (top), Boese (middle), and random forest (bottom) models. The random forest model was fit using \( R_g \), VPD, \( T_{air} \), GPP, and year. Thresholds designating dry, morning shifted, and C:W uncoupled days were \( \text{EF} < 0.2 \), \( \frac{\text{EF}C\text{ET}}{C\text{ET}} < -0.5 \) to \(-0.25 \), and \( \text{DWCI} < 0 \) to \(-0.25 \) respectively for each day. The distributions span from the 10th to 95th percentiles, and the width of each gives an indication of the variance, which is larger in the sub groups compared to all points. Furthermore, the mean difference in WUE (black lines) tends to be shifted in dry and morning shifted days indicating a mean underestimation of WUE by the models mostly due to the long tails. Decoupled days show higher variance, but no clear pattern in under- or over-estimation. The percentage of days in each category are designated next to y-axis label in parenthesis.
days show the smallest shifts for all models, with an overestimation of WUE for the Katul model and a slight mean underestimation of WUE the more empirical models (Boese and random forest) Katul-Zhou and Boese models and no significant shift of WUE with the random forest model. As all models were calibrated within a site-year, the over or under estimation of WUE indicate an inability of the model to capture a change in the system. Cases of mean mis-estimation tended to be influenced by long tails in the distribution with median differences being less exaggerated. However, these long tails are indicative of major model error in periods where the ecosystem is likely under stress conditions.

Discussion

Looking beyond sums and means

The proposed metrics, DWCI and $C_{ET}^*$, depart from more traditional summarization methods to summarize from sub-daily to daily timescales such as sums and means. This departure is advantageous in that it extracts added information that may have been otherwise ignored by turning the focus from signal amplitude to the signal shape. However, these new metrics also come with their own set of caveats, most notably issues with data quality confounding interpretability. Both metrics are susceptible to noise, as one or two errant points within a day can be reflected as a decrease in correlation or a shift in diurnal centroid. This is evident from the existence of very afternoon shifted $C_{ET}^*$, sometimes by more than an hour, which the authors have no proposed explanation for other than noise in the data. However, attributing highly afternoon shifted points as poor data requires further investigation. Note here that the “resting” $C_{ET}^*$ seems to be slightly afternoon shifted, which could be caused by real physiological factors such as differences in the incoming SW radiation (Rg) used in the calculation and net radiation (Rn), higher atmospherics demand (VPD) in the afternoon driving higher ET, or increased convection throughout the day resulting in higher transport of water away from the canopy, and is likely a combination of all three. Differences in resting $C_{ET}^*$ between sites could also be from instrumental causes such as radiometric sensors which are not adequately leveled or dirty, though the consistent, slight afternoon shifts would suggest this is a real response. Despite the possible shortcomings, both metrics show a definite response to drought conditions across the broad array of sites, climates, and ecosystems contained in FLUXNET (see Fig. 4 and Fig. 6), and give valuable insight into the underlying physiology. Given the broad nature of the analysis here, the metrics and hypothesis presented would benefit from site specific validations such as looking to see if the morning shits and decoupling are indeed associated with lower soil moisture levels, leaf water potentials, and/or decreases in sap flux. Sap flux in particular could give some interesting insights, as the diurnal patterns in sap flux velocity will also have an offset to incoming radiation related to tree capacitance, therefore relating sap flow diurnal centroids to the ET diurnal centroid could give some information on changes in plant water recharge. Furthermore, the diurnal centroid base metrics complement the hysteresis quantification methods such as those employed by Zhou et al. [55] and
Matheny et al. [29], with the advantage of $C_{ET}^*$, being compensated for cloudy conditions and possibly comparatively less influence of noise, though an intercomparison would be useful to explore the strengths and weaknesses of the different approaches. By providing both the equations and related code of the metrics, we the authors hope the metrics will be used by the community for both validation and to further ecophysiological understanding.

**Trees, grass, and drought stress**

By comparing climate groups and PFTs with the frequent occurrence of low DWCI and $C_{ET}^*$ from Figure 4, we can note two striking differences: evergreen broad- and needle-leaf forests show high variability of morning shifted days but not uncoupled days, whereas grasslands show significantly high uncoupled but not morning shifted days. The pattern is further seen in Figure 5, where the distinct divergent responses of decoupling and morning shifts between tree and grass dominated systems. This disparity may indicate an interaction of $C_{ET}^*$ not only with drought, but hydraulic sensitivity. The association of morning shifts to hydraulic sensitivity is further strengthened by Figure 6a,d where $C_{ET}^*$ shows a much stronger response to EF rather than VPD, indicating that morning shifts of ET are not simply due to stomatal closure due to VPD but in fact a response to drought conditions. The shorter hydraulic system of grasses may not necessitate stomatal closure under high demands [14], thus causing less frequent phase shifts even under drought conditions. These In contrast, tree ecosystems may only exhibit higher hydraulic stresses, associated with both dryness and a more sensitive hydraulic strategy. Temperate-continental and tropical climates all showed a low frequency of morning shifted days, even though they are occupied by large trees with cavitation susceptible vascular systems [21], suggesting that these ecosystems show limited drought stress even with the hydraulic susceptibility. Similarly, the high degree of variability of morning shifted frequency between site-years in sub-tropical/Mediterranean and evergreen broad- and needle-leaf forests could either indicate variation in the response in hydraulic stress between sites, or that hydraulic stress is only expressed some years, leading to high and low frequencies within the same site.

In this way, it seems that though $C_{ET}^*$ is less noisy as a drought indicator (see Figure 6), it may only be of use in tree systems that are more prone to hydraulic stress. However, this does put the metric in a rather unique position in that it could be used as a global scale hydraulic indicator, having potential application in exploring ecosystem level isohydricity [27]. Current or the degree to which risks vascular system damage to continue to extract water. Isohydricity is intrinsically a concept that relates to an individual plant, as dynamics of rooting depth, hydraulic conductances, and sensitivities to VPD can vary within individuals of the same species at the same location. However, these factors are all interrelated, as hydraulic and stomatal conductances drive transpiration dynamics which control the rate of depletion of root zone water which can then feed back to stomatal sensitivity, such as via ABA signaling [50]. As such, current estimates of isohydricity require repeated quantification of leaf water potential plant level measurements, which are currently restrained to the individual scale, i.e. from actual leaf measurements [28] or to global scale, but only 0.5 degree
resolution estimates from radar [21]. This limitation of large and small scales leaves a knowledge gap at the size of an eddy covariance footprint, hindering the study of ecosystem response to drought. However, under the assumption that the morning shifts seen under low evaporative fraction are due to stomates closing throughout the day-related to increased stomatal sensitivity in response to root zone moisture depletion, it may be possible to compare the onset and speed with which the diurnal centroid shifts toward the mornings as ecosystems dry. In this way, one could infer the ecosystem response to soil moisture, without explicitly knowing the soil moisture. The resulting relationship could prove useful as a data derived ecosystem functional property, giving direct information on variations in water limitation response.

C:W decoupling and energy balance closure

In addition to error from single data points, both metrics, but especially the DWCI, show some relationship with energy balance over closure. Energy balance mismatch is a common phenomenon in EC measurements, with under closure (ET+H<Rn) being a more common concern [24, 52]. Issues with energy balance closure can be, among other causes, attributed to advection, where energy, water, and carbon are transported in and out of the tower footprint, complicating an absolute accounting of these quantities [1, 5, 51]. The apparent association of DWCI and over closure could be due to transfer of moist air from the surrounding landscape, causing the DWCI to be more contingent on the mixing of source air and less from plant controls. In this scheme, the over closure seen in Figure 6 could be caused by the mixing of outside moist air into the drier air from the EC site, causing an increase in latent energy. However, the infiltrating air sources could also have similar or drier moisture levels which would not necessarily be seen as over closure. In this scenario, this infiltrating air could contain varying carbon and water concentrations, again causing a carbon:water decoupling, but one that would not be associated with over closure. If this effect has no diurnal pattern, and thus does not generally influence the mean diurnal centroid in ET, it could explain why the patterns with dryness are much clearer with $C^*_{ET}$ compared to DWCI. This would have the implication that DWCI is then a mixture of advection and non-stomatal signals, complicating the biological interpretability. However, the association with dryness in both metrics gives credence that they do indeed reflect some physiology, if we assume EBC should not be influenced by dryness level. Furthermore, if potential stress conditions are removed, the DWCI could be useful as a metric of advection in the system, even when the energy balance is relatively well closed.

WUE shifts associated with metrics and not captured by models

Figure 5 demonstrates the strong tendency of the models to underestimation WUE in dry conditions. This is true even for the fully non-linear and empirical random forest model, indicating that the model under-performance is not necessarily due to an incomplete model framework, but due to a lack of information to constrain the problem.
the association of both metrics with drought (Fig. 6), one could expect that the models would underestimate WUE in uncoupled and morning shifted days. Though this is the case with morning shifted days, decoupling shows only small and inconsistent no underestimations of WUE, with even a mean overestimation in the case of the Katul model. Given the limitations outlined in the previous sections, one could blame noise for the lack of WUE shift, but this does not reconcile with the higher frequency of decoupling during dry days which should bias the WUE estimates. Furthermore, as the more empirical models (Boese and RF) reduce random forest model reduces the prediction variability, they leave a slight WUE underestimation, indicating that some of the overestimation from the Katul model may be tied to limitations of the underlying theory, assumptions, yet the distribution from the RF model still lacks the long tails of underestimation in the empirical models characteristic of the dry points. Extending these findings to the underlying hypotheses of the metrics, namely hydraulic and non-stomatal limitations, we could conclude that the hydraulic controls do impose a greater water use advantage than non-stomatal limitations. In other words, the findings suggest that days with water:carbon decoupling, and possibly non-stomatal limitations, do not improve WUE, whereas hydraulic responses can improve WUE. As WUE is a ratio, this does not shed any light onto the change in productivity, as low values of WUE may indicate that a plant is still productive, but at a higher water cost. However, solid conclusions would require further analysis with some site specific measurements of actual plant function.

Though the models used here are relatively simple and lack the complexities and feedbacks found in more vigorous ecosystem models, Matheny et al. [29] also demonstrated the fundamental inability of 9 different land-surface models with 4 different stomatal conductance schemes to capture diurnal variability which the authors attributed to inadequate representation of how water gets from the soil to the leaf. Given the demonstrated phenomenon of morning shifts and decoupling across sites under dry conditions, the metrics here provide a benchmarking tool for mechanistic models to test their ability to replicate these patterns, suggesting that the models are capable of expressing hydraulic and non-stomatal limitations. Furthermore, in the case of machine learning approaches, the metrics may provide a useful input parameter which summarizes these diurnal effects, as is evidence by the difference in response the bias in RF modeled WUE, i.e. while both metrics are associated with low EF, RF WUE was underestimated with morning shifted days but not decoupled days implying that two different strategies are being captured by the metrics. As such, by demonstrating the utility of the metrics, and providing code and explanations for calculation, we hope they become useful to the community at large.

Conclusion

Both the DWCI and the $C_{ET}^*$ demonstrate an ability to show consistent patterns with drought across a broad array of sites, climates, and ecosystems, with the added advantage of being tied to theoretical underpinnings. Particularly,
the demonstrated patterns give novel information about carbon water relations and hydrological dynamics that are not currently present at ecosystem scale across a database as large as FLUXNET. These metrics and their underlying theory provide a data derived example differentiating the hydrological response of tree and grass plant functional types, as well as give evidence for the presence and absence of a WUE advantage from hydraulic and stomatal limitations respectively. Going forward, these metrics can be used as a tool to further understand the diversity of ecosystem drought responses.

References


