We thank the reviewer for their constructive comments and we address their various concerns below.

In revision, we have now fixed the WUE bug (see below) and this now allows us to investigate both the carbon and the water simulations by CABLE. We now very clearly demonstrate that models need to implement different sensitivities to water stress at xeric sites, otherwise models will underestimate carbon and water fluxes during drought. Finally, we have restructured the results to make the text clearer for the reader.

De Kauwe and others explore drought parameterization in the CABLE model. An alternate drought formulation is found to improve modeled GPP and LE across five European flux sites in response to the 2003 drought.

The paper as written is interesting and complete but in many cases must be revised for clarity. The choice of sites is poorly described, as is the justification for the drought schemes chosen. The tendency to describe the gradient of sites as north/south rather than xeric/mesic is distracting. That being said, the results are logical with a simple and clear message that will benefit global model development. I recommend publication following (many) minor revisions.

Forested sites in Europe were selected from those available through the Protocol for the Analysis of Land Surface models (PALS; http://pals.unsw.edu.au; Abramowitz, 2012). These data have previously been pre-processed and quality controlled for use within the LSM community. We have clarified the text: "*To assess the performance of the CABLE model both with and without the new drought scheme, we selected a gradient of five forested Fluxnet (http://www.fluxdata.org/) sites across Europe (Table 2) from those available through the Protocol for the Analysis of Land Surface models (PALS; pals.unsw.edu.au; Abramowitz, 2012). These data have previously been pre-processed and quality controlled for use within the LSM community. Consequently, all site-years had near complete observations of key meteorological drivers (as opposed to significant gap-filled periods).*"

As for the description of sites, we have followed the reviewer's suggestion and replaced north/south with mesic/xeric. We have also changed the paper title to reflect this change as well.

The introduction is well-written and well-cited but could use improvement. The passage 'Our ability to model drought effect on vegetation function is currently limited' is vague. Some drought responses are simulated very well, others poorly, and the challenge remains to model drought response well, all the time.

We have clarified this sentence: "*Our ability to model drought effect on vegetation function (carbon and water fluxes) is currently limited (Galbraith et al. 2010; Egea et al. 2011; Powell et al. 2013).*"

The following paragraph discusses the Galbraith results, then the Powell results, then the Galbraith results again.

We have now combined both sections of text that referred to the Galbraith paper.

A good argument that PFTs are insufficient to capture the range in drought responses. It would be even better to give examples within PFTs that differ with respect to their isohydric or anisohydric behavior. In this case, might the behavior of species in a PFT average out or would all different species (or groups thereof) emerge to become important?

We thank the reviewer for this suggestion and have amended the text: "*Such an approach ignores experimental evidence of the range of sensitivities to drought among vegetation types, which would fall within a single PFT group (Choat et al. 2012; Limousin et al. 2013; Zhou et al. 2014; Mitchell et al., 2014; Mencuccini et al. 2015). For example, Turner et al. (1984) found contrasting responses in leaf water potential to increasing vapour pressure deficit, ranging from isohydric to anisohydric, among a group of woody and herbaceous species. Similarly, Zhou et al. (2014) found that in a dry-down experiment, European sapling species originating from more mesic environments were more sensitive to water stress (more rapid reduction of photosynthesis and stomatal conductance) than species from more xeric*

regions. However, it is not known whether observed differences in the response to soil moisture deficit among species are important in determining fluxes at large scales."

The need to test drought parameterizations across sites is described nicely. What was not described well is the justification for the hypothesis that drought sensitivity would increase as a function of latitude. First and foremost, latitude is only ever a correlate of something else like temperature or daylength. If this justification is improved, the manuscript would be more compelling.

We have now added additional text to clarify this: "*We hypothesised that drought sensitivity would increase as sites transitioned from xeric to mesic. We hypothesised that trees at more mesic sites, with a greater abundance of available water than at xeric sites, would be more vulnerable to shorter duration droughts, and thus have higher drought sensitivity (or lower resistance to drought). Therefore, accounting for this latitudinal gradient in drought sensitivity would improve the performance of CABLE.*"

It may be argued that the optimal stomatal function framework falls victim to the simultaneous need for plants to not succumb to hydraulic stress (e.g. Sperry 2004). That being said, optimization theory is important to consider in models although for the case of drought it might be superseded by hydraulic considerations, which are described nicely in equations 3-5. In other words, the model as written incorporates optimal stomatal behavior and conductance, but it is able to simulate tree death?

CABLE, similar to many other LSMs does not directly simulate tree mortality: instead only accounting for reductions in productivity.

It would be good to cite the work of Katul, Leuning, and Oren (2003) with respect to the coupling of hydraulic and photosynthetic parameters; I believe this is the original reference for this notion.

Much as we like the paper by Katul et al., in this paper we are not discussing the hydraulic constraints to transpiration so it does not seem relevant to cite this paper.

The three tested approaches were selected as plausible ways to appropriately weight soil water potential in a model. They were not intended to span a complexity gradient. Rather, we started with M1; finding it was too strongly weighted to the top soil we tried M2; finding that was too strongly weighted towards the lower soil we tried M3 which uses dynamic weighting. We have added further text, similar to that found in the discussion section 4.1, which should better explain the logic behind our selected approaches: "*We tested three potential approaches for weighting in this paper:*

(i) *Using the root-biomass weighted θ and converting this to Ψ_S using Eq. (8). Such an approach is often favoured by models, following experimental evidence that plants preferentially access regions in the root zone where water is most freely available (Green and Clotheir 1995; Huang et al. 1997).*

(ii) *Taking the integrated θ over the top 5 soil layers (1.7 m depth) and converting this to Ψ_S using Eq. (8). This method assumes the plant effectively has access to an entire "bucket" of soil water. This approach is often favoured by "simpler" forest productivity models (e.g. Landsberg and Waring, 1997).*

(iii) *Weighting the average Ψ_S for each of the six soil layers by the weighted soil-to-root conductance to water uptake of each layer, following Williams et al. (1996; 2001). The total conductance term depends the combination of a soil component (R_s) and a root component (R_r). R_s is defined as (Gardner, 1960):*

$$R_s = \frac{ln\left(\frac{r_s}{r_r}\right)}{2\pi l_r D G_{soil}} \tag{6}$$

where r_s is the mean distance between roots (m), r_r is the fine root radius (m), D is the depth of the soil layer, G_{soil} is the soil conductivity (mmol m^{-1} s^{-1} MPa^{-1}) which depends on soil texture and soil water content, l_r is the fine root density (mm^{-3}). R_r is defined as:

$$R_r = \frac{R_r^*}{FD} \qquad\qquad (7)$$

where R_r^* is the root resistivity (MPa s g mmol^{-1}), F is the root biomass per unit volume (g m^{-3}). This method weights Ψ_S to the upper soil layers when the soil is wet, but shifts towards layer lowers as the soil dries, due to the lower soil hydraulic conductance (e.g. Duursma et al. 2011)."

I like the honesty of section 2.3.1. That being said, is the problem simply and conveniently avoided in this case? How is a reader to know that it does not factor into the results?

Both reviewers have raised an issue with respect to our section outlining the water use efficiency (WUE) bug and the ensuing analysis. We acknowledge therefore that we were not clear enough in our original text.

The bug *only* affected the modelled fluxes of gross primary productivity during periods of drought: *all* water fluxes were correctly simulated (at all times). During periods where there was an inadequate soil water supply to meet the atmospheric demand for simulated transpiration, transpiration was reduced to the available amount of water in the root zone. To maintain a coupled carbon and water cycle, photosynthesis should also be reduced, but wasn't. This has the effect of producing erroneously high WUE during periods of extreme water limitation, i.e. free carbon in exchange for water might be another way to interpret this. It was for this reason we chose to focus *all* of our analysis on latent heat and transpiration and *not* on gross primary productivity. Nevertheless, we did originally show the gross primary productivity fluxes so that the readers could judge for themselves the erroneously increased water use efficiency during drought (i.e. the effect of the bug).

The bug has now been fixed. We thank Vanessa Haverd and Ying-Ping Wang of CSIRO for their assistance with this fix. Following the fix, we have removed all text referred to the bug from the manuscript. Hopefully this change should allay the reviewers' concerns about the impact of the bug on our results. We do note that part of the fix requires changing the method of root extraction in the standard version of

the model. Consequently, as well as reducing erroneously high GPP during drought, there were small changes in the latent heat fluxes as well.

Section 2.4 could use expansion to justify the choice of the 5 sites. Why were they chosen?

We have clarified the selection choice as stated in our comment above.

The results section is succinct. Note that RMSE has units.

We have added the missing units throughout.

Also, back to the question about why the three different drought parameterizations were chosen, were the first two straw men or are these common in LSMs for simulating drought?

No, these are not straw men; they are common approaches in LSMs. With the additional information added to the methods as per the reviewer's earlier comment, this should now be clearer.

Per the comments above regarding latitude, the first sentence of the discussion sounds more robust with mesic species exhibiting higher drought sensitivity than xeric ones for which one can assume that plants have adapted. That being said, there must be some good references for this basic concept. In the first paragraph of the discussion the authors move back to this north/south framework rather than the wet/dry framework, which is perhaps additionally surprising from a group from Australia.

As suggested we have replaced the latitude text and added appropriate references.

In section 4.1 I wouldn't say that pot moisture is necessarily uniform but rather the relationship between active root area and the moisture profile does not match what is commonly observed in the field.

The text has been modified to read: "*in which it is fair to assume that the soil moisture content is relatively uniform and fully explored by roots. In contrast, soil moisture content and rooting depth in the field typically have strong vertical profiles.*"

Whilst the TRY database does have a great deal of useful information on plant traits, it does not contain the necessary information to parameterise a model like the one considered in this manuscript. The sensitivity of gas exchange to drought is not one of the traits compiled in the database.

We have amended the text to: "*During droughts, plants are often observed to shed their leaves. This is a self-regulatory mechanism to reduce water losses (Tyree et al. 1993; Jonasson et al. 1997; Bréda et al. 2006).*"

The following sentence could use re-working: Overall however, there remains a tendency to trade mechanistic realism is often traded for present day accuracy,
We have removed this text.

From Table 1 the sites go at least as far maritime/continental as they do 'north/south'.
See comment below.

A relatively far northern site wasn't chosen. Just another reason to couch things in terms of water availability rather than latitude.

As suggested, we have changed all site descriptions to mesic/xeric, rather than north/south. We did not use a far northern site as the 2003 heatwave did not extend into the far north, as far as we are aware.

In figure 1 (and figures 3-7), how was transpiration measured?

In these figures transpiration was not measured (note there is no black line to indicate observations), instead only the observed LE fluxes are shown.

References Sperry J.S. (2004). Coordinating stomatal and xylem functioning: an evolutionary perspective. New Phytologist, 162, 568-570.