The paper presents an important work in the evaluation of terrestrial biosphere models (TBMs) for savanna ecosystems. A unique data set of fife measurement sites (meteorology and eddy covariance) from the North Australian Tropical Transect (NATT) along a precipitation gradient was utilized to evaluate the capability of six widely used TBMs to predict GPP and LE for savanna type ecosystems. The authors divided the data into wet and dry season to identify which TBMs perform better under which conditions. They used the results to interpret which factors (e.g. evaporation of trees, grass, soil, or rooting depth) make different models perform better than others in the dry or wet season.

The presented work shows nicely that savanna ecosystems are not well described by TBMs and that they need to be developed further to better predicted GPP and LE for savanna ecosystems.

In the following you find minor comments:

Page 19014 line 9 -11 'For dry season LE, half the models (BIOS2, MAESPA, and SPA) were able to consistently outperform the emp2 benchmark, and come close to meeting the same number of metrics as the emp3 benchmark particularly at the drier sites.'
I don't see that MAESPA fits into the described. Three out of fife sites are very close to emp1 and only two are slightly better than emp2. Did you maybe mean BESS? Or MAESPA for GPP? Please clarify.

Graphics
A figure including the seasonal cycle of the drivers used for the benchmarking would be nice. Sturt Plains is most likely a water limited site also during the wet season while Howard springs and Adalaide River might also be energy limited especially during the wet season. This could help to understand under which driver-conditions the BTMs are actually predicting better.

Figure 2 please use the same y-axis limits on all plots. Details of differences might get lost but the amplitude of the seasonality and its prediction becomes clearer between the sites.

Figure 4 correlation plots: most of the correlations are clearly not linear, as far as it can be identified from the plot. The color scheme with the faded / pale colors makes it hard to distinguish individual sites from each other. To me it seems that there are mostly individual clusters and large scatter around it. Maybe you can redo the figure with brighter colors.

Language:

Please write LPJGUESS or LPG-GUESS throughout the text including the figures.

Page 19006 Line 15 TBM is introduced but not explained.
Page 19006 Line 28 ESMs is introduced but not explained.
Page 19008 line 8-9 should be Donohue et al. (2009) not Donohue et al. (Donohue et al., 2009)
Page 19014 line 14 '....(LPJGUESS, BESS.....' instead of '...(LPJGUESS. BESS....'
Page 19016 line 27 please add 'to' between 'expected' and 'perform'
Page 19017 line 12 should be...(adjustments of stomatal ....
Page 19018 line 6 maybe better to write ... observations of understory LE...
Page 19021 line 4 DGVMs first time introduced please explain
Page 19021 line 16-17 should be: Such formulations are therefore not mechanistic, do not respond to actual season dynamics....