Review of

"Improving the LPJmL4-SPITFIRE vegetation-fire model for South America using satellite data"

Drüke et al. GMD-2019-92

The paper utilises a genetic optimization algorithm and a revised fire danger index to improve the representation of burnt area and biomass in the LPJmL4-SPITFIRE model compared to satellite-derived datasets, optimised against those same datasets. The authors also benchmarked the fractional cover of one PFT and claimed improvements to PFT distribution and temporal dynamics both (inter-annual variability and seasonal patterns). They also advocate the use of such methods for improving fire-vegetation models in general.

Investigating alternatives to the Nesterov Index in SPITFIRE (and other global fire models) and using optimisation algorithms to develop DGVMs and fire models are laudable aims and this work makes useful contributions in these directions. Simultaneously using both burnt area and biomass observations to constrain the model parameters, and the application of rigorous benchmarking metrics are also to be commended. Many parameters in SPITFIRE are very poorly constrained, so this is a promising approach to improve the model.

However, I do have substantial reservations regarding the presentation and, to some extent, the methodology, which I believe need to be addressed prior to publication. I first list my main concerns, and then a series of comments to the text. I feel confident that these concerns can be addressed in a revised version of the manuscript, perhaps with some additional analysis.

Main concerns

1. Whilst the optimisation procedure produces very reasonable results in the case of the VPD FDI, the Nesterov Index results are not so clear cut and cast some doubt on the efficacy of the method. Yes, the summary metrics for spatial BA do get better (at least the NSME does, the Willmott coefficient goes down, which I assume means worsening agreement?), the temporal metrics do improve drastically for Caatinga but worsen for the Cerrado, and the biomass results are basically unchanged. So that is a mixed bag. But, most critically, a visual inspection of the BA produced shows a massive reduction in fire and almost complete spatial mis-match compared to the observations, not the preferred behaviour of a fire model! There is much to discuss here which is missing from the manuscript. Benchmarking/optimising burnt area is hard due to the large amount of zero values and then high peaks, and so getting a fire peak wrong by one or two gridcells is heavily penalised. Thus an optimisation will tend towards a conservative 'no fire strategy'. This appears to be what is happening here, but is not discussed. This obviously raises questions about whether or not BA can effectively be used in such a context when it produces results which objectively (in terms of metrics) are perhaps better, but somewhat subjectively may not actually produce a more useful model.

2. The optimisation to both BA and biomass is definitely a good idea, and as far as I can tell combining the two KGE metrics is reasonable. However, as part of the paper is to

demonstrate this approach, I think there must be more discussion and analysis of this method. In particular, can the authors disentangle the relative constraints of each dataset in the method? I think this is important information for such a method. If all else fails, perhaps simply running the optimisation for BA and biomass individually would be an option.

3. No specific information on how the gridcells used in the optimisation were selected. It seems to have been done just by 'picking some'. By the authors' own admission this may bias the optimisation. Could they justify their choice a little better? Furthermore, could it be possible to run with random gridcells every time? Or gridcells close to the meteorological stations used in the preparation of the climate data? A more concrete method for select the gridcells, or at least a clearer justification, is required.

4. There is no discussion of what the optimised parameters mean in terms of process understanding or what the newly introduced 'alpha' for the VPD FDI really means. Many of the existing parameters move very little (perhaps a little surprising but also perhaps reassuringly), but the rCKs are very interesting. For NI_{optim} these converge to very similar values and move away strongly from their initial values. Having similar crown kill probability for raingreen and evergreen trees flies in the face of the assumptions in SPITFIRE so far. But for VPD_{optim} the story is somewhat different, with rCH for TrBE remaining very high, but rCK for TrBR also increasing. Please discuss these results, including some ecological context. And regarding the new 'alpha', what does this really mean? The very different value for TrBE compared to TrBR and TrH definitely deserves some discussion as it appears to be integrating some new factor into the FDI which the NI does not include and is not adequately represented in the other SPITFIRE PFT-specific parameters. Some discussion, even if it is a little speculative, is necessary here. In generally I can see no problem in tuning process-based models with 'black box' optimisation procedures and somewhat unphysical variables, but there must be at least some attempt to interpret and relate the results back to the processes.

5. Again, relating to the process-understanding, plots of the fire intensity resulting from the methods should be shown (possibly in an appendix if necessary). The "fuel moisture -> combustion completeness -> fire intensity -> mortality" link is a critical pathway in these results, it should be discussed explicitly but is not.

6. There is no benchmarking of the PFTs that we expect to be effected by fire! The inclusion of TrBE PFT FPC is great, but what about TrBR and TrH? These should be at least plotted, and ideally benchmarked. If the ESA CCI dataset does not have useful classifications in this regard, at least MODIS VCF MOD44B Tree-Nontree-Bare would provide some reference data for the Caatinga and Cerrado.

Specific comments to the text

Abstract

'partly poor'
-rephrase

‘as a starting point’
– rephrase, this is not the first work to improve fire in DGVMs

‘improves simulation of … plan functional type’
– is that really demonstrated?

Introduction

P2 *‘recent decline in global burnt area*’
– now contested (indeed by one of the authors)

P2 *‘Especially in South America, tropical forests, woodlands and other ecosystems are vulnerable to increasing fire danger and land use change’*
– reference?

Material and Methods

P6 *‘SPITFIRE further includes a surface intensity threshold’*
– please state this threshold here. I realise that this is in the Table 2 but the units are not given.

P7 *‘The fire danger index is scaled by a PFT-dependent constant, α_i, over the number of PFTs n (Thonicke et al., 2010)’*
- in the original Thonicke et al. 2010 implementation, the α varied over fuel classes (ie. 1hr, 10hr, 100hr, 1000hr and live grass fuels) *not* PFTs. Please explain and justify this change. Also, there no mention of live grass fuels. Are they parameterised as in original SPITFIRE?

P8 *‘and a monthly mean for R to avoid unrealistic high flammability fluctuations in time steps with isolated events of very low R’*
– can the authors justify this further? I know it is stated in the Pechnoy and Shindell paper, but it is not immediately clear why flammability fluctuations due to rainfall events should be ‘unrealistic’. Perhaps with their experience with this method, the authors can provide a more convincing argument.

P8 *‘Hence, we scaled the VPD up with a PFT-dependent scaling factor α_i’*
– since this has a very different physical meaning than the α_i above, I strongly suggest using a different symbol.

P8 ‘*The general behaviour of the two indices as modelled by LPJmL in dependence of relative humidity and temperature is shown in Fig. 3’*
– Fig 3 is a nice plot, but please explain in a little more detail how the panels are comparable , as in how was the effect of vegetation taken in to account in the lower panel for VPD FDI?

P8 *‘We regridded and aggregated the data set to the LPJmL resolution of 0.5 ◦ × 0.5 ◦ and to a daily time step’*
– normally climate data is the limiting factor when it comes to spatial resolution in DGVMs. Is there any reason that the authors chose to aggregate this rather that use 0.25 degree? Especially when the evaluation data sets are available at 0.25 degree or finer. It seems like throwing away information.

P10 *'The optimization was performed for 40 grid-cells in South America to represent a variety of fire regimes (Fig. 2). Most of them were selected in active fire regions, especially in the Cerrado and Caatinga. In addition a few pixels with no or almost no fire occurrence (e.g. central Brazilian Amazon) were chosen.'*
– this is a rather vague description of what may be a very important choice in the optimisation procedure! See my main concern above. Please give more details in the logic here.

P10 –Despite being important the FDI_{NI} and the Rothermal equations, and being poorly constrained, moisture of extinction was not mentioned as a possible parameter for optimisation. Could the authors discuss this?

P11 *'NMSE'*
– can the authors justify their choice of NMSE over NME?

P11 '*Willmott coefficient'*
– please explain its range and meaning, as is done for NMSE.

Results

P12 '*mainly by shifting much of the simulated burnt area from the sparsely vegetated Caatinga towards the Cerrado region'*
– this is true to some extent, but it also much is moved into Amazonia in regions where very little fire is observed in reality. In order to back up this statement, a table with the burnt area in each region for each simulation should be provided. The over-estimation of fire in Amazonia should be discussed in the Discussion section.

P13 *Figure 5*
- There is some fire in the Amazonia region, both in the data *and* in the simulations. Therefore, this region should be included in Figure 5 and Table 1, and discussed.

P13 *'Here, the TrBE showed the largest value (22.41), ca. 20 times as large as the TrBR (1.21) and TrH (1.13) (Tab. 2)*'
– there is no discussion of what this actually means in the Discussions section, please include an interpretation.

P17 *'... but also here we got an even larger improvement, when only the fire-prone regions Cerrado or Caatinga are considered (Tab. 3)'*
- Caatinga results are not shown in Tab. 3, although I think they should be. Possibly also results for Amazonia (see above)

P 18 *Figure 7*
– difference plots are great and I can see the logic behind including the difference relative to the original model version to show improvements (as you have done) but please show the absolute values too (as in Figure 4).

Discussion

P19 *'Another result of the optimizing procedure, using FDI_VPD , was the improvement of the PFT distribution..'*
– I am not sure that statement is justified given the very small improvement in TrBE and no demonstrated improvement in the other PFTS.

P19 '*it emphasizes that three parameter sets determining PFT distribution*'
– what three parameter sets? You mean three PFTs? Or something else? Please clarify.

P20 *'Limitations during the optimization process'*
– this heading is somewhat confusing and maybe should better be 'Limitations of the optimization process'

P20 'As shown in Fig. 8, the modelled PFT coverage showed an equal distribution of tropical raingreen and evergreen PFTs throughout wide parts of central-northern South America'
– Fig. 8 shows no such thing, it only shows the FPC of the evergreen PFT. Of course, it may simply be that the caption is incorrect somehow, but otherwise the distribution of the raingreen PFT must be shown to demonstrate this.

P20 *'By choosing a large amount of optimization cells in the, by NI orig , strongly overestimated Caatinga region, the burned area decreased there significantly after the optimization'*
– this (slightly confusing statement) would appear to indicate that the authors acknowledge that their results depend heavily on the choice of gridcells for the optimisation (see above)

P20 *'In the Cerrado and especially the Caatinga, however, trees suffer from water stress in the dry season and should shed their leaves to avoid mortality related to drought or growth efficiency. The resulting dominance of the TrBR PFT has a very different effect on fire spread and is more fire-tolerant (different fuel characteristics and resulting fire intensity), thus has a lower fire-related mortality.'*
– whilst this a reasonable enough statement (in fact pretty much inherent in the construction of DGVMs and SPITFIRE) it is hard to see what it has to do with the limitations of the optimisations process.

P21 – *'Nonetheless, we were able to improve the interannual variability and hence, the model performance during extreme years for the Cerrado and Caatinga regions (e.g. for 2007/2008, Fig. 5).The optimized SPITFIRE is now able to model accurately the climate dependent seasonal and interannual variability as well as the spatial extent of fire on natural land throughout the fire-prone woodlands of South America.'*
– yes and no. In the Cerrado the results from Fig 5. are not significantly different between VPD and Original, and whilst the results are better in the Caatinga for VPD, most of this comes down to the overall normalisation, it is hard to see if VPD really catches between IAV and seasonal dynamics. In fact, the R^2 (which is insensitive to the normalisation) actually gets worse going from Original to VPD. So these statements need much more nuance. And a plot of the normalised time series (equivalent to Fig 5., at least for the Caatinga) might be a more effective way showing improvements in IAV and seasonal dynamics.

P21 *entire section titled 'Outlook*
- the way ahead in improving fire modules in DGVMs' – this text does not really fit the title. Much of it refers specifically SPITFIRE or LPJml, specifically their current limitations. Please reconsider/revise/re-title this section.

P21 The statements *'it would be possible to use an even more comprehensive fire danger index (e.g. Canadian Fire Weather Index; Wagner et al., 1987) or different fire danger indices for different biomes'* and *'In a global modelling approach, however, we need to find one fire danger index'* seem to contradict each other, please resolve!

Conclusions

P21 *'We have demonstrated a major improvement of the fire representation within LPJmL4-SPITFIRE by implementing a new fire danger index and applying a model-data integration setup to optimize fire-related parameters.'*
- whilst there are tangible improvements, they are only tested and in the Caatinga and Cerrado, the region for which the optimisation was done (which you do mention in the next sentence). I would suggest toning this down slightly.

P21 *'We improved the seasonal and interannual variability'*
– I have yet to be convinced of this, especially as the R^2 for the time series are not improved with VPD. And I am not sure how to interpret the Willmott coefficient as this is not described.

P21 *'A realistic representation of fire is also crucial for fire-vegetation-climate feedbacks and is hence necessary for DGVMs coupled within and comprehensive Earth system model.'*
– I think you can drop that sentence, as it attempts to summarise and justify fire modelling in general rather than this work. The penultimate sentence is fine to end with.