Supporting Information for

Description and Evaluation of the JULES-ES setup for ISIMIP2b

Camilla Mathison[1,2], Eleanor Burke[1], Andrew Hartley[1], Douglas I Kelley[4], Eddy Robertson[1], Chantelle Burton[1], Nic Gedney[1], Karina Williams[1,3], Andy Wiltshire[1,3], Richard J Ellis[4], Alistair A Sellar[1], Chris D. Jones [1]

[1] Met Office Hadley Centre, FitzRoy Road, Exeter, UK

[2] School of Earth and Environment, Institute for Climate and Atmospheric Science, University of Leeds, Leeds, UK

[3] Global Systems Institute, University of Exeter, Laver Building, North Park Road, Exeter, UK

[4] UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK

Corresponding author: Andrew Hartley (andrew.hartley@metoffice.gov.uk)

Contents of this file

Text S1 to S2
Figures S1 to S6
Tables S1 to S6

Introduction

The following supporting information provides additional detail on the following:

- Evidence of minimal impact of the disaggregator to convert daily driving data into 3-hourly timesteps (Figure S1)
- Post-processing steps involved in formatting the JULES raw output into the files available for download at https://data.isimip.org/search/query/jules-es-55/ (Text S1)
- ISIMIP driving data used in this study (Figure S2 and S3)
- iLAMB approach to model assessment (Text S2; Tables S1 to S6)
- More detailed evaluation of the simulations, including seasonal variation in surface fluxes (Figure S4), modelled vegetation with fire on (Figure S5) and comparison of hydrological impacts with fire on and off (Figure S6)

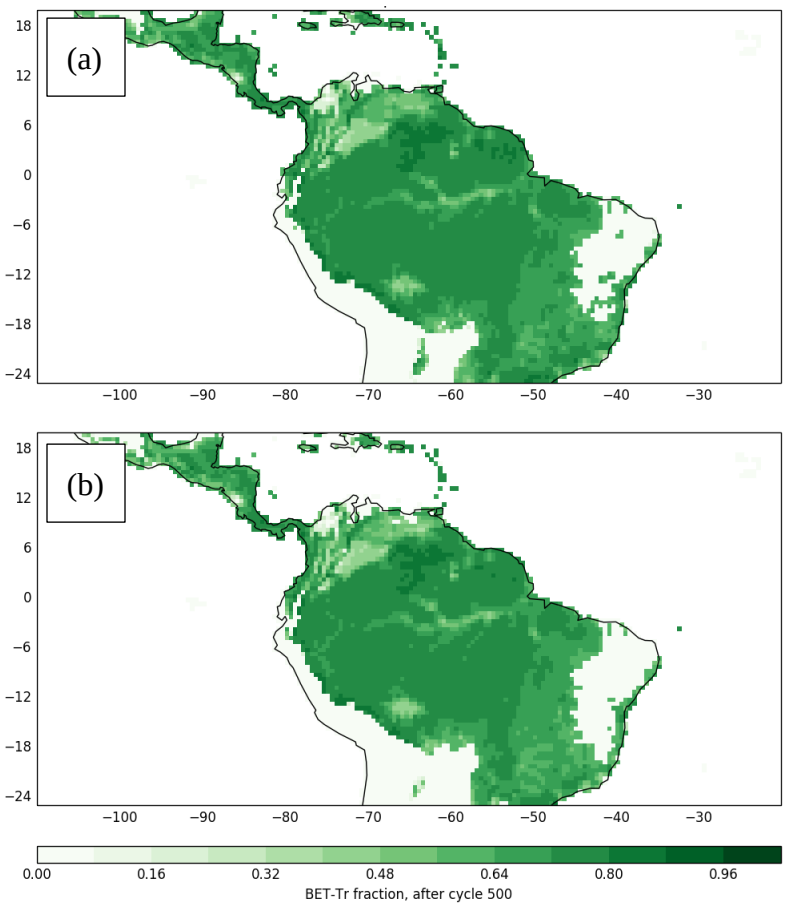

Figure S1: Tropical broadleaf evergreen tree fraction after 500 cycles of spinup from (a) JULES-ES-ISIMIP driven by 3-hour GSWP3 driving data, and (b) JULES-ES-ISIMIP with disaggregation.

Text S1: Converting data to ISIMIP Protocols

Code is available and is used here to post-process the raw JULES output data to ensure that it conforms to the ISIMIP protocols. It's main functions are to:

- Modify units
- Define special variables, for example those that need to be calculated from others e.g., Net Biome Production (NBP).

We also use the ISIMIP quality control python code available from github (https://github.com/ISI-MIP/isimip-qc). It checks the filenames against the protocol schemas and patterns. This code also checks variables, dimensions and global attributes ensuring consistency across all submitted data.

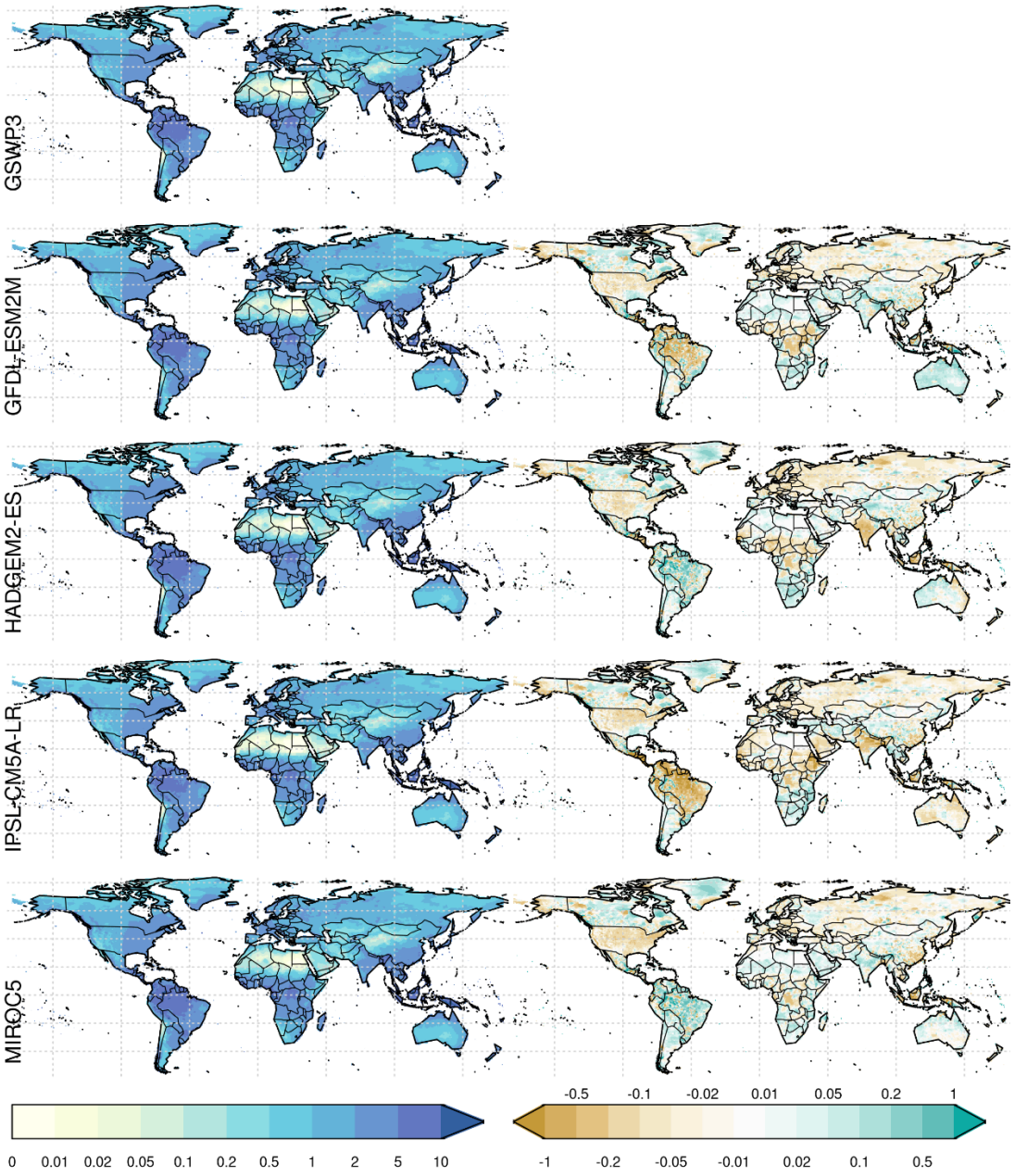

Figure S2: Annual average precipitation for 1979-2006 for (top to bottom) GSWP3 observations and each driving dataset. (right) the difference between driving data and observations.

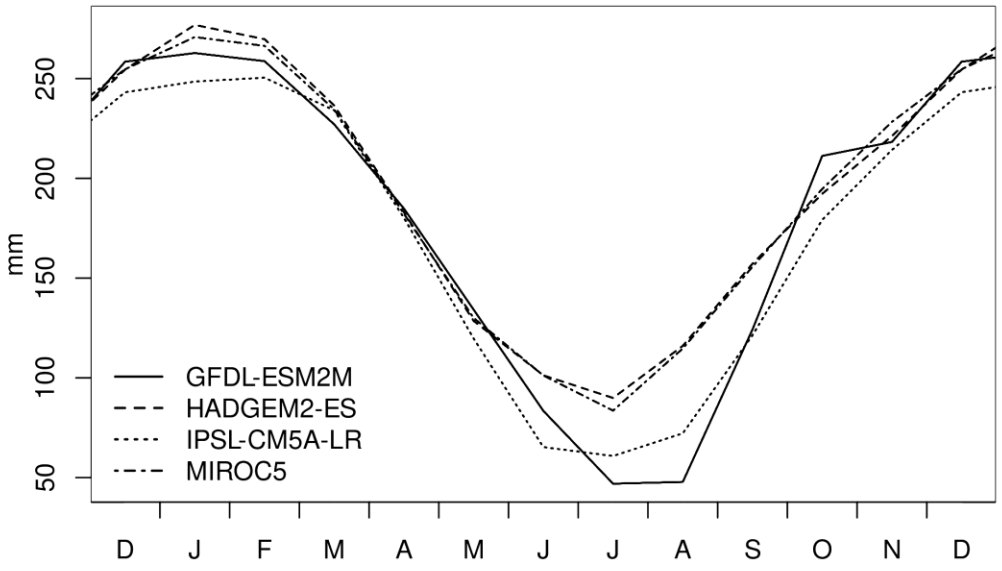

Figure S3: Climatological averaged precipitation over the Amazon basin (see Figure 1 and 2) between 1979-2003

Text S2: ILAMB

Evaluation of JULES using the International Land Model Benchmarking system (ILAMB; Collier et al. (2018)). For each model–observation comparison a series of error metrics are calculated, scores are then calculated as an exponential function of each error metric, scores increase from 0 to 1 with improvements in model performance. Metrics and scores for GPP, ET and Albedo are evaluated on a latitude-longitude grid and global mean scores are shown in tables S1, S4 and S5. The runoff metrics and scores are calculated for 50 major river basins and the global mean score is weighted by the area of the basin (Table S6). To calculate the bias score, bias is normalized by observed temporal variability before an exponential function is used to map the normalized bias to fit between 0 and 1. For some variables, including GPP and ET the global mean bias score is weighted by the absolute value of the observation in order to down-weight regions of low GPP or ET. Spatial distributions scores are based on a combination of the spatial correlation and spatial standard deviation of time-mean model data. Seasonal cycle scores are based on the error in the timing of the seasonal maximum. For vegetation cover, we used the Manhattan Metric (MM; Table S2) from Kelley et al. (2013), which is the area weighted mean of grid-cell summed absolute difference between different fractional items. We compared tree vs none-tree (which includes all other vegetation items and bare soil), shrub vs none-shrub, wood (tree and shrub) vs none-wood and grass vs none-grass. Each comparison excludes water bodies, ice and urban areas, with vegetation and soil rescaled to add up to 1, and area weighting adjusted accordingly.

	Global average NBP(PgC/yr)	Global average GPP (PgC/yr)	Bias (g m-2 d^{-1})	Bias Score	Seasonal Cycle Score	Spatial Distribution Score
Observed	1.0-2.8	120-140				
GFDL	0.93	135	0.39	0.49	0.75	0.95
HadGEM2	1.37	137	0.45	0.48	0.78	0.93
IPSL	1.25	134	0.37	0.47	0.76	0.94
MIROC	1.46	137	0.45	0.47	0.77	0.93
GFDL + fire	0.66	133	0.34	0.48	0.74	0.94
HadGEM2 + fire	1.16	135	0.40	0.45	0.78	0.91
IPSL + fire	0.87	132	0.32	0.45	0.75	0.92
MIROC + fire	1.34	135	0.40	0.45	0.76	0.91

Table S1: benchmark results for NBP and GPP versus Jung et al. (2011) upscaled Fluxnet GPP observation. Observed global total for GPP and NBP are from Global Carbon Budget (Friedlingstein et al., 2020). Cells shaded in global NBP and GPP columns to indicate where the simulation is within the range of the Global Carbon Budget.

	Tree		Shrub		Woody total		Grass	
	Bias (Mkm^2)	MM	Bias (Mkm^2)	MM	Bias (Mkm^2)	MM	Bias (Mkm^2)	MM
Observation (Global total)	34.86		5.02		39.88		36.87	
GFDL	5.31	0.24	1.28	0.12	6.59	0.44	-8.91	0.38
HadGEM2	5.08	0.24	1.31	0.12	6.39	0.44	-9.96	0.38
IPSL	4.97	0.24	1.22	0.12	6.19	0.43	-9.66	0.38
MIROC	5.31	0.24	1.22	0.12	6.53	0.44	-11.47	0.38
GFDL + fire	-0.49	0.27	-4.09	0.08	-3.60	0.30	-9.5	0.42
HadGEM2 + fire	1.55	0.26	-4.34	0.08	-2.79	0.32	-10.98	0.41
IPSL + fire	1.05	0.26	-4.41	0.08	-3.36	0.31	-10.69	0.41
MIROC + fire	1.86	0.26	-4.11	0.08	-2.25	0.33	-12.99	0.42

Table S2: benchmark results for vegetation fraction observations ESACCI Land cover v2.0.7 (Harper et al., 2022). Shades as per Table S1.

	Global total (Mkm^2)	Bias (monthly %)	Bias Score	Seasonal Cycle Score	Spatial Distribution Score
Observation	4.55	0.346			
GFDL + fire	4.43	-0.0431	0.74	0.83	0.74
HadGEM2 + fire	4.13	-0.0655	0.74	0.83	0.74
IPSL + fire	4.33	-0.0479	0.75	0.84	0.76
MIROC + fire	3.94	-0.0657	0.74	0.83	0.73

Table S3: benchmark results for burnt area versus GFED4s observations (Van Der Werf et al., 2017).

	Bias ($mm\ d^{-1}$)	Bias Score	Seasonal Cycle Score	Spatial Distribution Score
GFDL	0.13	0.62	0.82	0.98
HadGEM2	0.17	0.62	0.80	0.97
IPSL	0.13	0.62	0.81	0.97
MIROC	0.18	0.63	0.81	0.97
GFDL + fire	0.09	0.63	0.82	0.98
HadGEM2 + fire	0.14	0.64	0.80	0.97
IPSL + fire	0.10	0.64	0.81	0.98
MIROC + fire	0.15	0.64	0.81	0.97

Table S4: benchmark results for evapotranspiration (ET) versus GLEAM V2A dataset (Miralles et al., 2011). Note that the ILAMB comparison to MODIS MOD16A2 ET (Mu et al., 2011) produces consistent results, with mean biases being positive and slightly smaller when fire is included, and the other scores being nearly identical between the simulations with and without fire. Shades as per Table S1

	Bias	Bias Score	Seasonal Cycle Score	Spatial Distribution Score
GFDL	0.07	0.32	0.59	0.97
HadGEM2	0.07	0.33	0.60	0.97
IPSL	0.07	0.33	0.59	0.97
MIROC	0.07	0.32	0.60	0.97
GFDL + fire	0.07	0.33	0.57	0.96
HadGEM2 + fire	0.07	0.33	0.58	0.96
IPSL + fire	0.07	0.33	0.57	0.96
MIROC + fire	0.07	0.33	0.57	0.96

Table S5: benchmark results for albedo versus GEWEX SRB radiation observations (Stackhouse et al., 2011). Shades as per Table S1

	Bias (mm d^{-1})	Bias Score	Spatial Distribution Score
GFDL	0.02	0.80	0.97
HadGEM2	-0.01	0.80	0.97
IPSL	-0.04	0.78	0.94
MIROC	-0.01	0.80	0.97
GFDL + fire	0.05	0.78	0.96
HadGEM2 + fire	0.02	0.79	0.96
IPSL + fire	-0.01	0.77	0.94
MIROC + fire	0.02	0.78	0.96

Table S6: benchmark results for global runoff verus Dai and Trenberth (2002) derived GRDC Aiguo Runoff Dataset. We've shaded cells where metric scores indict whether the with or without fire runs perform better. Lighter shade indicates equal scores with and without fire.

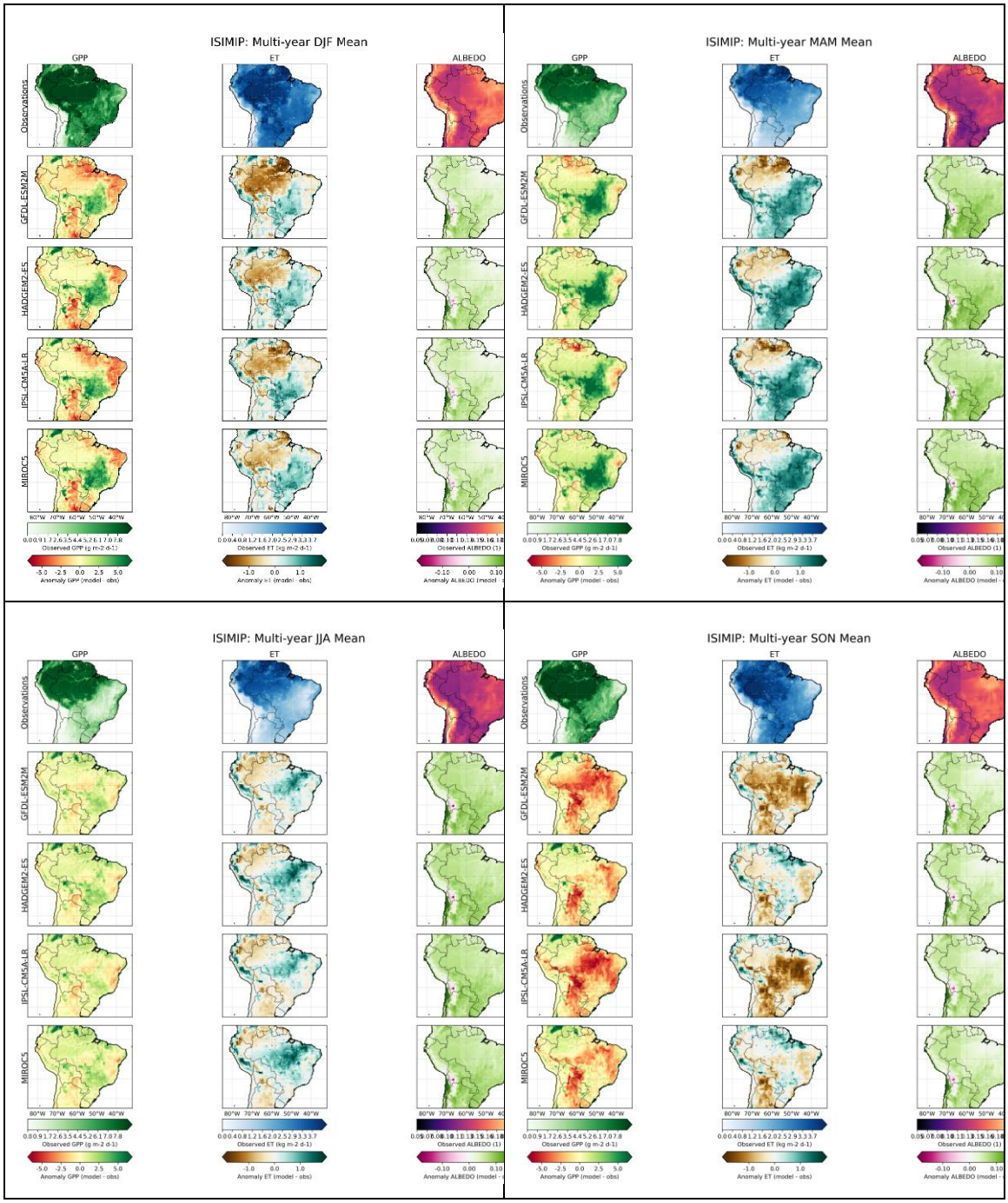

Figure S4: Seasonal variation in surface fluxes for tropical South America

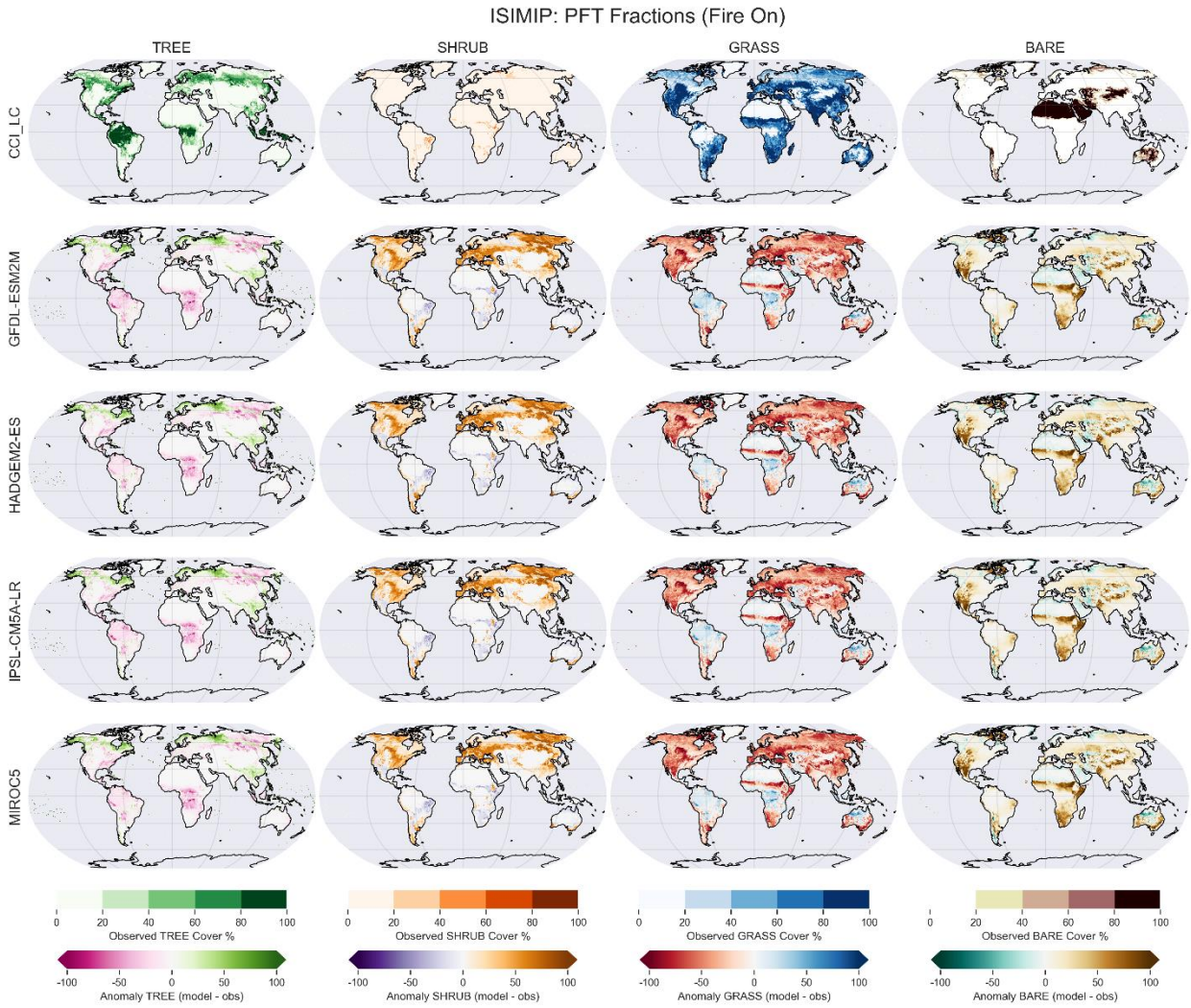

Figure S5: Comparison of modelled vegetation cover, with fire, vs observations. Split (left to right) by tree, shrub, grass and unvegetated (bare) fraction. (Top to bottom) Observations CCI_LC v2.0.7 (Harper et al., 2022), and the difference between model and observations for each set of driving data

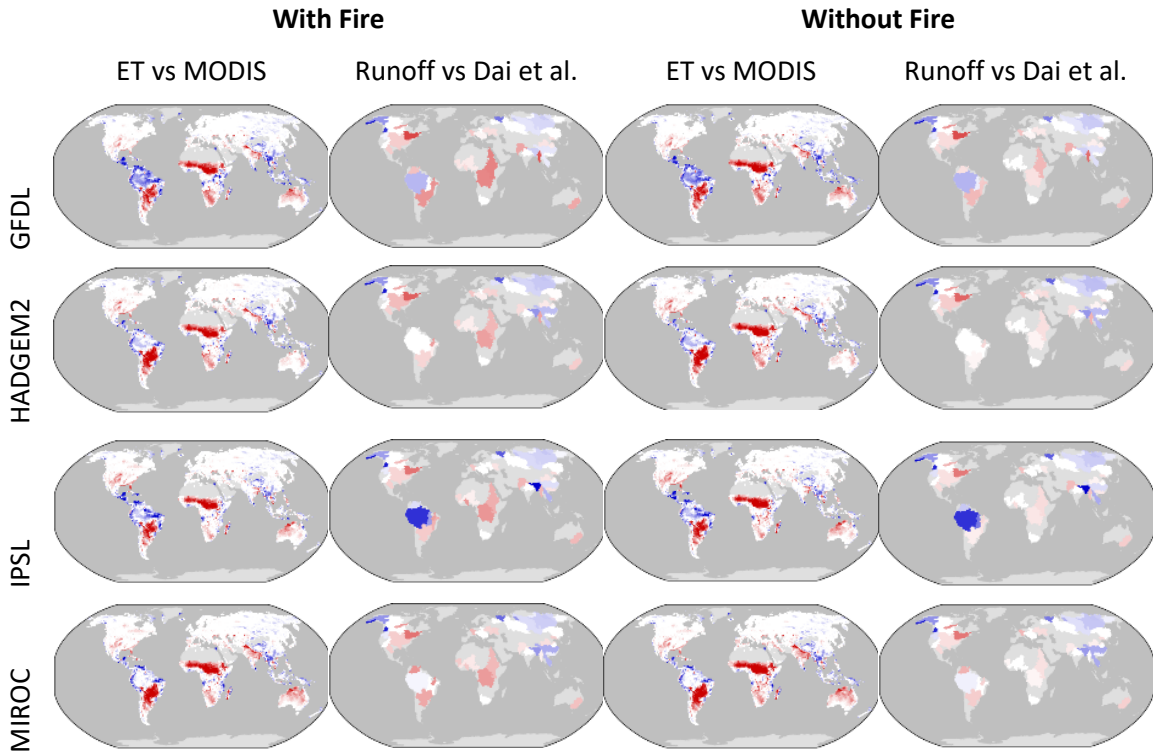

Figure S6: Multi-year mean bias of ET and catchment scale runoff simulated by JULES with and without fire, as driven by 4 sets of climate driving data. Modelled ET (1980-2011) is compared to GLEAM (Miralles et al., 2011). Runoff (1980-2014) is compared to Dai & Trenberth (2002). ISIMIP2b forcing data derived from 4 CMIP5 GCMs: GFDL-ESM2M; HadGEM2-ES; IPSL-CM5A-LR; MIROC5.

References

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Dai, A., & Trenberth, K. E. (2002). Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations. *Journal of Hydrometeorology*, *3*(6), 660–687. Retrieved from www.R-ArcticNET.sr.unh.edu

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global Carbon Budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. https://doi.org/10.5194/ESSD-12-3269-2020

Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., et al. (2022). A 29-year time series of annual 300-metre resolution plant functional type maps for climate models. *Earth System Science Data Discussions*, *Submitted*.

Jung, M., Reichstein, M., Margolis, H. a., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*, G00J07. https://doi.org/10.1029/2010JG001566

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., & Willis, K. O. (2013). A comprehensive benchmarking system for evaluating global vegetation models. *Biogeosciences*, *10*(5), 3313–3340. https://doi.org/10.5194/BG-10-3313-2013

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., & Dolman, A. J. (2011). Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences*, *15*(3), 967–981. https://doi.org/10.5194/HESS-15-967-2011

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., et al. (2017). Global fire emissions estimates during 1997-2016. *Earth System Science Data*, *9*(2), 697–720. https://doi.org/10.5194/ESSD-9-697-2017