Table 1. Results of FLUXNET sites ([1]) monthly mean latent heat evaluation for the models evaluated in the paper. This evaluation was performed with the ILAMB evaluation framework ([2,3]) as an addition to the published evaluation in Schellekens et al. (2017) [4]. ILAMB provides a scoring system to relate modelled results to reference datasets. In the ILAMB system multiple performance metrics are calculated, and additionally these metrics are converted to scores ranging between 0 and 1 to facilitate comparison and averaging. In this exercise three performance metrics are calculated for latent heat: total bias, root mean square error (RMSE) and phase difference (difference in months between peak values); furthermore a total of five 0–1 scores are calculated, for global bias, RMSE, seasonal cycle, spatial distribution and inter-annual variability, plus a 0–1 overall score that summarizes them. The metrics and scoring system are explained in detail in the ILAMB documentation (uploaded by eartH2Observe project, http://earth2observe.github.io/water-resource-reanalysis-v1/assets/pdf/ILAMB_metrics_document.pdf). Further details on the evaluation, including access to individual site metrics can be obtained via communication with the corresponding author.

Model	Period Mean [W·m^{-2}]	Bias [W·m^{-2}]	RMSE [W·m^{-2}]	Phase Shift [d]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [-]	Interannual Variability Score [1]	Overall score [1]
WaterGAP3	42.81	9.157	28.13	-24.3	0.594	0.509	0.874	0.773	0.633	0.649
SURFEX-TRIP	46.80	13.07	27.57	-9.40	0.593	0.509	0.931	0.848	0.614	0.667
ORCHIDEE	52.72	18.99	27.68	-10.3	0.526	0.502	0.921	0.868	0.71	0.672
HTESSEL-CaMa	49.95	16.22	26.28	-6.17	0.55	0.524	0.92	0.869	0.683	0.678
SWBM	43.70	9.965	23.94	-15.0	0.66	0.545	0.921	0.826	0.641	0.69
W3RA	44.61	10.88	24.48	-15.3	0.644	0.541	0.915	0.867	0.665	0.695
HBV-SIMREG	44.15	10.42	24.35	-25.3	0.656	0.548	0.871	0.885	0.747	0.709
LISFLOOD	40.22	6.484	21.32	-17.3	0.688	0.591	0.926	0.849	0.716	0.727
JULES	44.31	10.57	20.32	-3.82	0.646	0.601	0.924	0.879	0.778	0.738
PCR-GLOBWB	39.26	5.523	19.18	5.562	0.69	0.629	0.939	0.863	0.742	0.748

Table references:

1. Lasslop, G.; Reichstein, M.; Papale, D.; Richardson, A.D.; Arneth, A.; Barr, A.; Stoy, S.; Wohlfahrt, Georg. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global Change Biol*. **2009**, *16*, 1, 187–208.
2. Luo, Y. Q.; Randerson, J.T.; Abramowitz, G.; Bacour, C.; Blyth, E.; Carvalhais, N.; Ciais, P.; Dalmonech, D.; Fisher, J. B.; Fisher, R.; et al. A framework for benchmarking land models. *Biogeoscience*s **2012**, *9*, 3857–3874.
3. Collier, N.; Hoffman, F.M. ; Mu, M.; Randerson, J.T.; Riley,W.J.; Koven, C.D.; Keppel-Aleks, G.; Lawrence, D.M. International Land model Benchmarking (ILAMB) Package v002.00. Available online: https://doi.org/10.18139/ILAMB.v002.00/1251621 (Accessed on 19 February 2019).
4. Schellekens, J.; Dutra, E.; Martínez-de la Torre, A.; Balsamo, G.; van Dijk, A.; Sperna Weiland, F.; Minvielle, M.; Calvet, J.-C.; Decharme, B.; Eisner, S.; et al. A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset. *Earth Syst. Sci. Data* **2017**, *9*, 389–413.

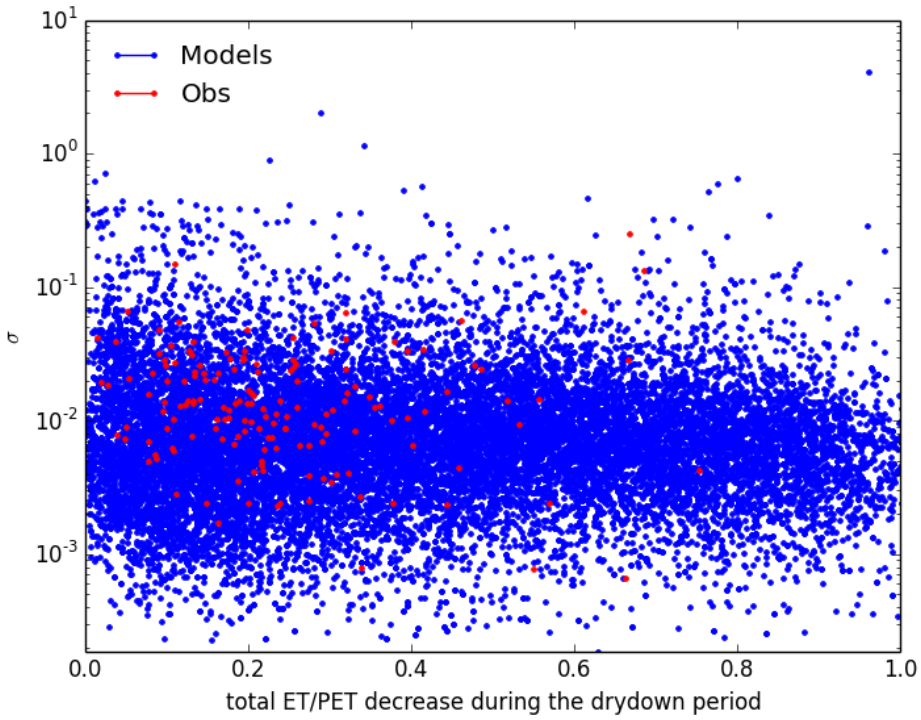

Figure 1. Standard deviation error metric (σ) of the exponential curve fittings for all drydowns identified as 'dry events' (sites in Figure 4) against the actual decrease of ET/PET ratio during the event duration. Red dots are σ for the site observations and blue dots are σ for the WRR1 models at those sites.

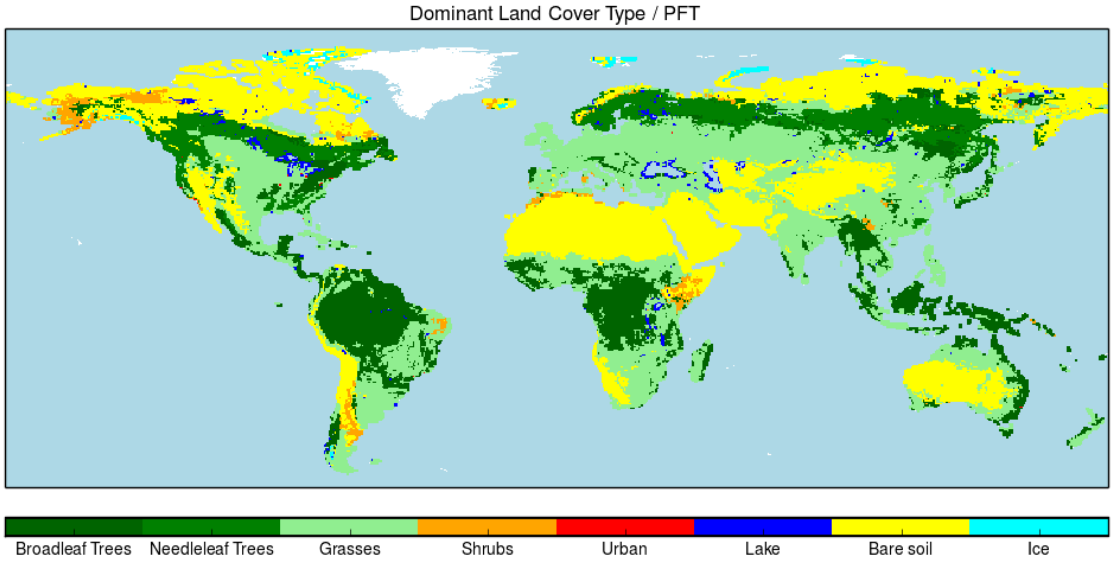

Figure S2. Global distribution of dominant land cover or plant functional type (PFT) at the 0.5° resolution, as highest fractional cover in the grid cell, in the land cover data used by the JULES model (from data derived from the International Geosphere-Biosphere Programme: http://www.igbp.net/).

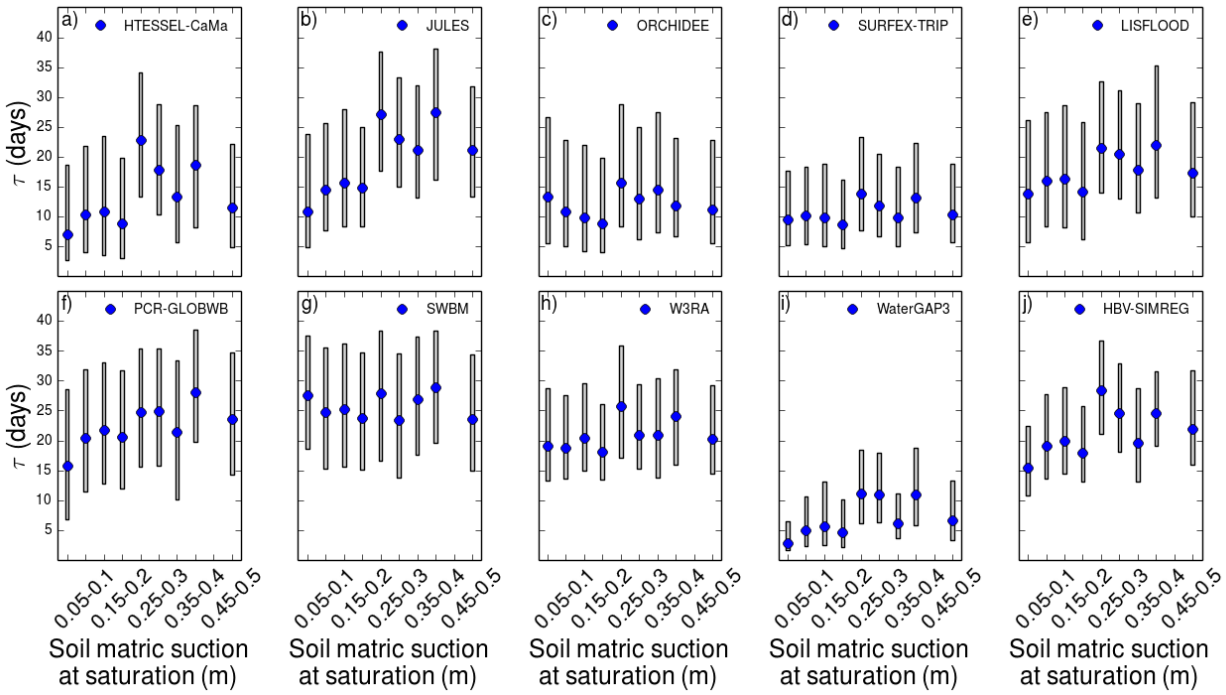

Figure S3. Drydown rates grouped by soil matric suction at saturation for the WRR1 models (blue dots). The soil water suction data comes from the Harmonized World Soil Database (HWSD; FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012). Dots indicate the median τ from identified drydown periods; bottom and top of grey boxplots indicate the 25th and 75th percentiles, respectively.