Geophysical Research Letters

Supporting Information for

Revisiting the potential to narrow uncertainty in the projections of Arctic runoff

E. Dutot[1], H. Douville[1]

[1] Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France

Content of this file:

- Table S1 and S2 compare the performance of the L19 and KCC statistical methods at the basin scale using pseudo-observations derived from a single realization of one out of the CMIP6 models.

- Table S2 and S3 compare the performance of the KCC statistical method at the basin scale using pseudo-observations derived from single versus multiple realizations of one out of the CMIP6 models.

- Fig. S1 and S2 compare the skill of the L19 regression method to predict simulated anomalies depending on the considered variable (P-E instead of runoff).

- Fig. S3 and S4 (to be compared with Fig. 1) evaluate the sensitivity of the L19 method to the choice of the constrained variable (P-E instead of runoff) and of the model ensemble (CMIP5 instead of CMIP6).

- Fig. S5 and S6 (to be compared with Fig. 2) evaluate the sensitivity of the KCC method to the choice of the model ensemble (P-E instead of runoff) and of the model ensemble (CMIP5 instead of CMIP6).

- Fig. S7 and S8 compare the runoff sensitivity to temperature and precipitation at short versus long timescales and explain why the L19 results should be considered with caution.

Basin	Coverage probability (%)	Change in CRPS (%)	Spread reduction (%)
Columbia	69.7	9.06	-37.5
Kolyma	57.58	22.8	-36.74
Lena	69.7	7.74	-12.83
Mackenzie	78.79	58.06	-14.24

Table S1: Probabilistic scores of the L19 method using the ssp51 ensemble (a single realization of each CMIP6 model under the SSP5-8.5 high-emission scenario). For each river basin, CRPS change and spread reduction are averaged after using successively each CMIP6 model as pseudo-observations.

Basin	Coverage probability (%)	Change in CRPS (%)	Spread reduction (%)
Columbia	84.85	16.08	-13.44
Kolyma	87.88	0.17	-28.61
Lena	93.94	-6.25	-23.88
Mackenzie	84.85	-22.65	-23.29

Table S2: Probabilistic scores of the KCC method using the ssp51 ensemble (a single realization of each CMIP6 model under the SSP5-8.5 high-emission scenario) and two observational constraints. For each river basin, CRPS change and spread reduction are averaged after using successively each CMIP6 model as pseudo-observations.

Basin	Coverage probability (%)	Change in CRPS (%)	Spread reduction (%)
Columbia	82.76	4.86	-25.48
Kolyma	82.76	-18.75	-39.89
Lena	86.21	6.54	-30.64
Mackenzie	72.41	-10.31	-35.39

Table S3: Probabilistic scores of the KCC method using the ssp50 ensemble (multiple realizations of each CMIP6 model under the SSP5-8.5 high-emission scenario) and two observational constraints. For each river basin, CRPS change and spread reduction are averaged after using successively each CMIP6 model as pseudo-observations.

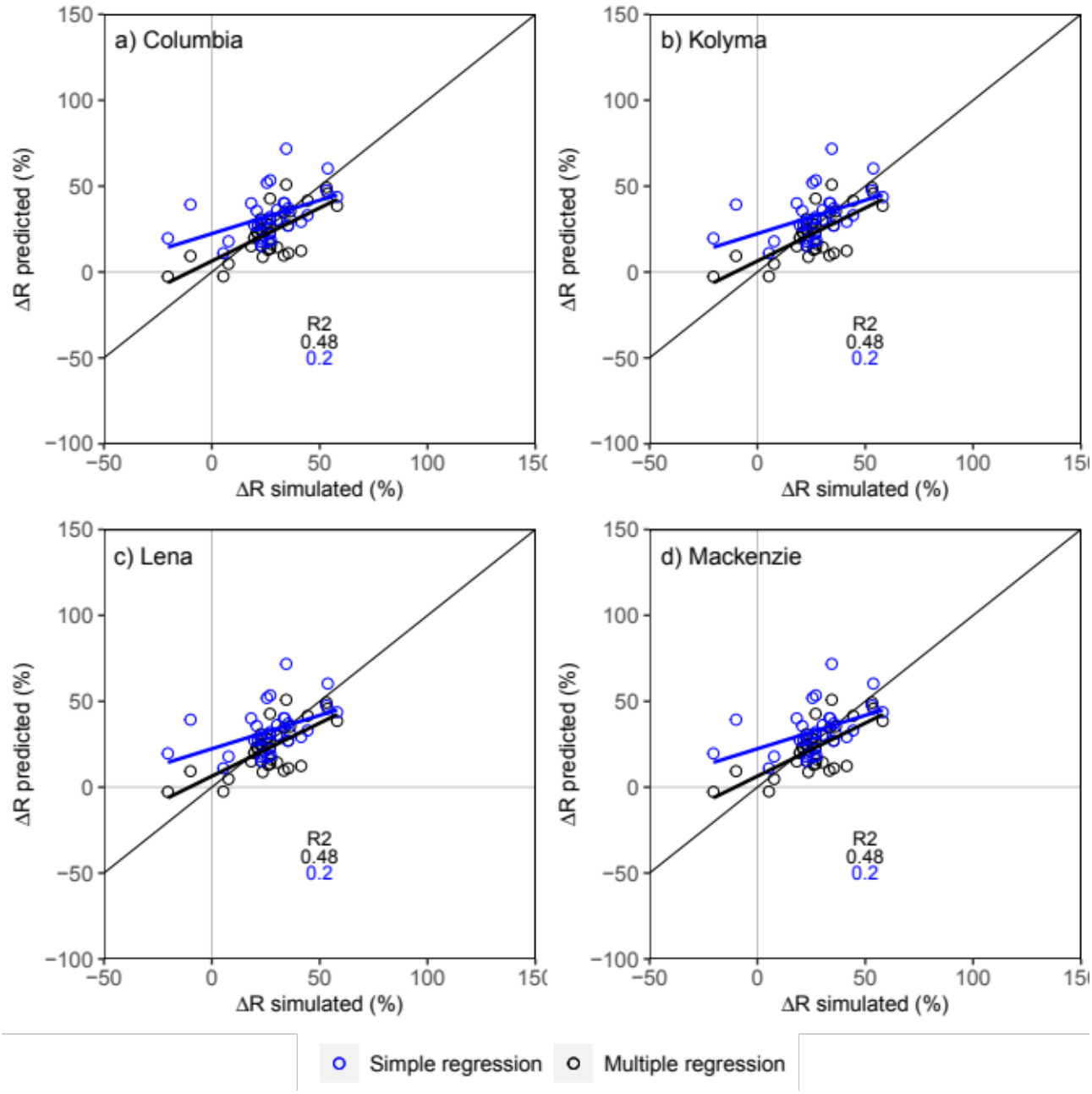

Figure S1: Scatterplots of predicted (L19) versus simulated (GCM) relative anomalies (%) of basin-scale water-year runoff in individual CMIP6 models under the SSP5-8.5 high-emission scenario: a) Columbia, b) Kolyma, c) Lena, d) Mackenzie. L19 runoff anomalies are computed from a simple (ΔR~ΔP) or multiple (ΔR~ΔP+ΔT) linear regression and plotted against the corresponding simulated anomalies. All anomalies are averaged over 2081-2100 relatively to the 1902-1930 baseline period. In each panel, R2 denotes coefficient of determination of the linear regression. The closer the regression line is from y=x (thin black solid line), the better the linear regression is. Not surprisingly, the multiple regression (black circles) is better than the simple regression (blues circles) at predicting the simulated anomalies.

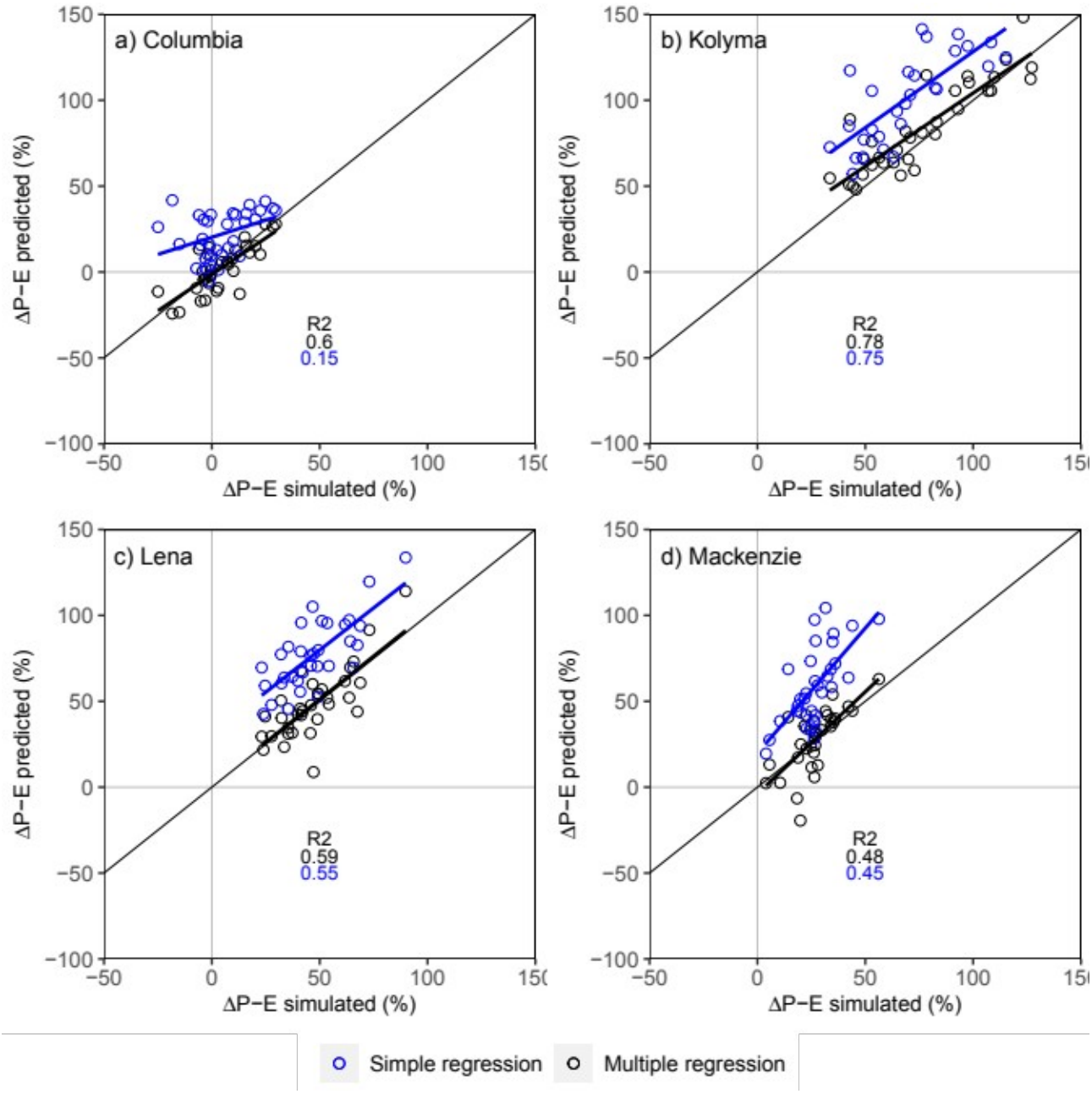

Figure S2: Same as Fig. S1 but using P-E (precipitation minus evapotranspiration) rather than runoff relative anomalies (%). Not surprisingly, it is easier to predict P-E rather than R anomalies with both simple (ΔP-ΔE~ΔP) and multiple (ΔP-ΔE~ΔP+ΔT) regressions given the strongly model-dependent simulated soil moisture anomalies in climate models. When temperature is not accounted for, the P-E simulated anomalies are systematically overestimated by the regression, thus highlighting the strong influence of global warming on surface evapotranspiration. Using P-E as a surrogate for water-year runoff therefore leads to overconfident projections.

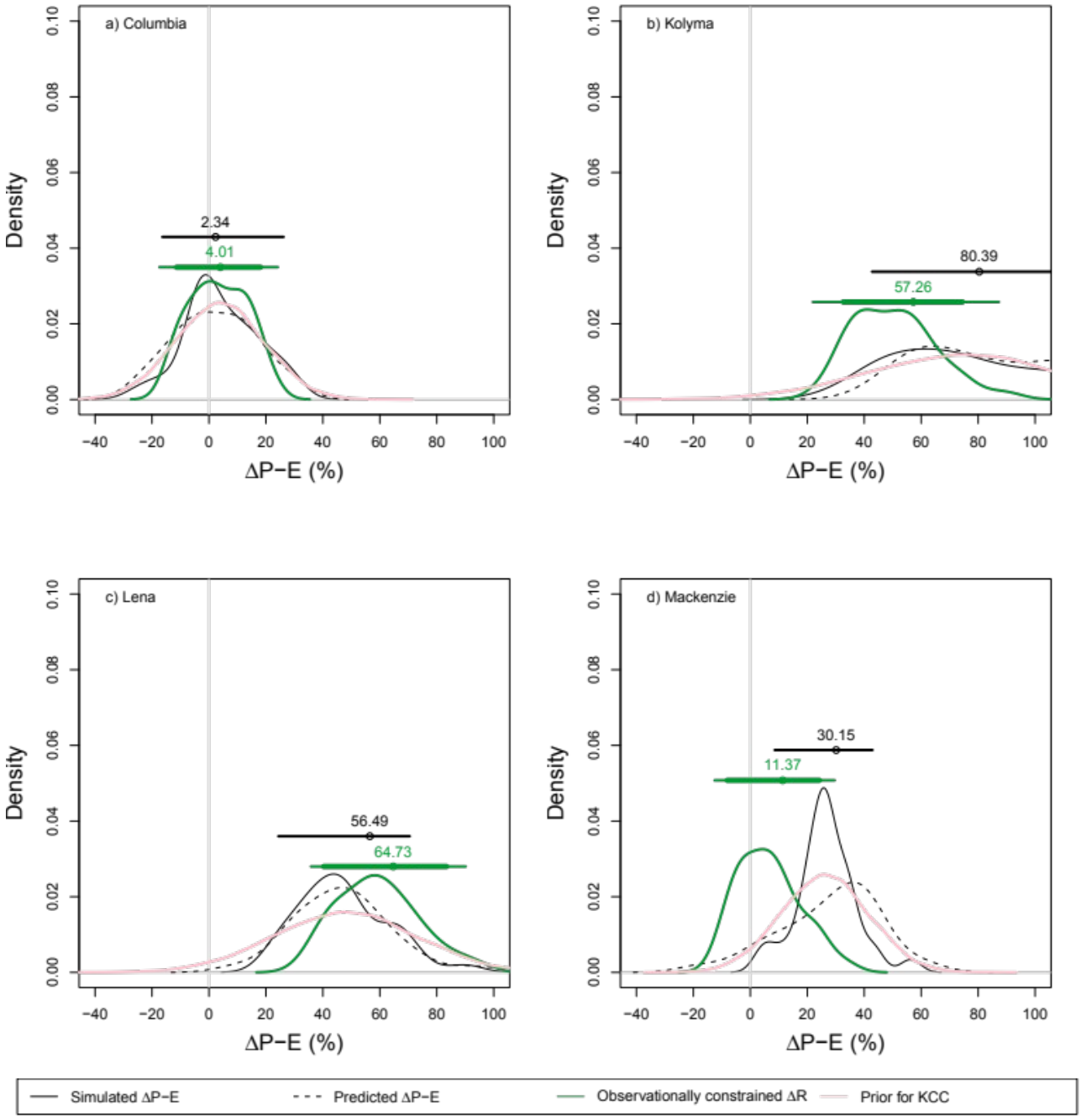

Figure S3: Constrained versus unconstrained distributions of water-year mean P-E relative anomalies (%) from the CMIP6 model ensemble under the SSP5-8.5 high-emission scenario: a) Columbia, b) Kolyma, c) Lena, and d) Mackenzie. Similar to Fig.1 but using P-E as a surrogate of R (although the regression coefficients are still constrained with GRUN runoff reconstructions).

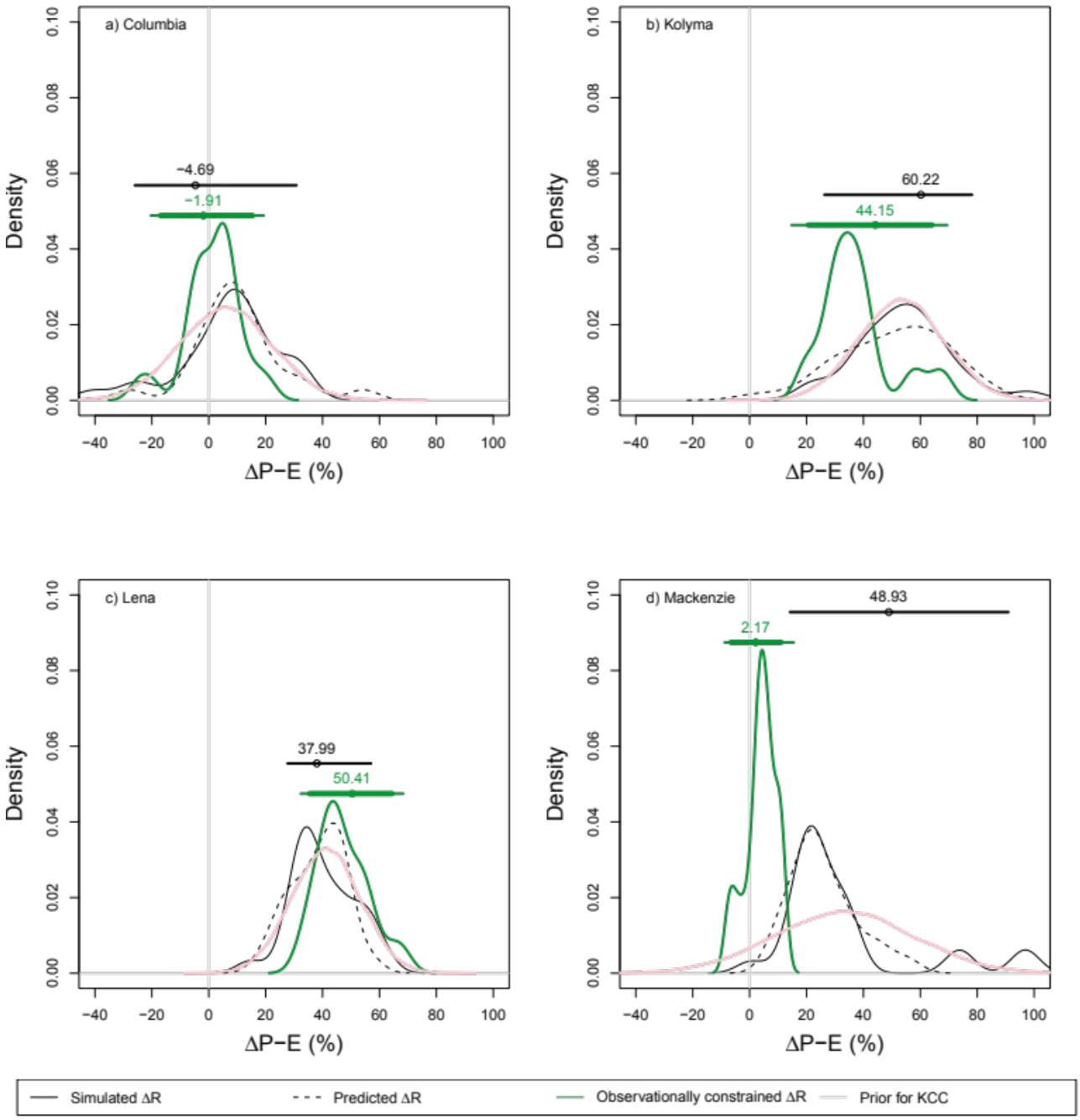

Figure S4: Constrained versus unconstrained P-E relative anomalies (%) from the CMIP5 model ensemble under the RCP8.5 high-emission scenario: a) Columbia, b) Kolyma, c) Lena, and d) Mackenzie. Similar to Fig.1 but using CMIP5 instead of CMIP6 models. The results are sensitive to the choice of the CMIP ensemble, although they are qualitatively consistent regarding the effect of the constraint on the ensemble mean.

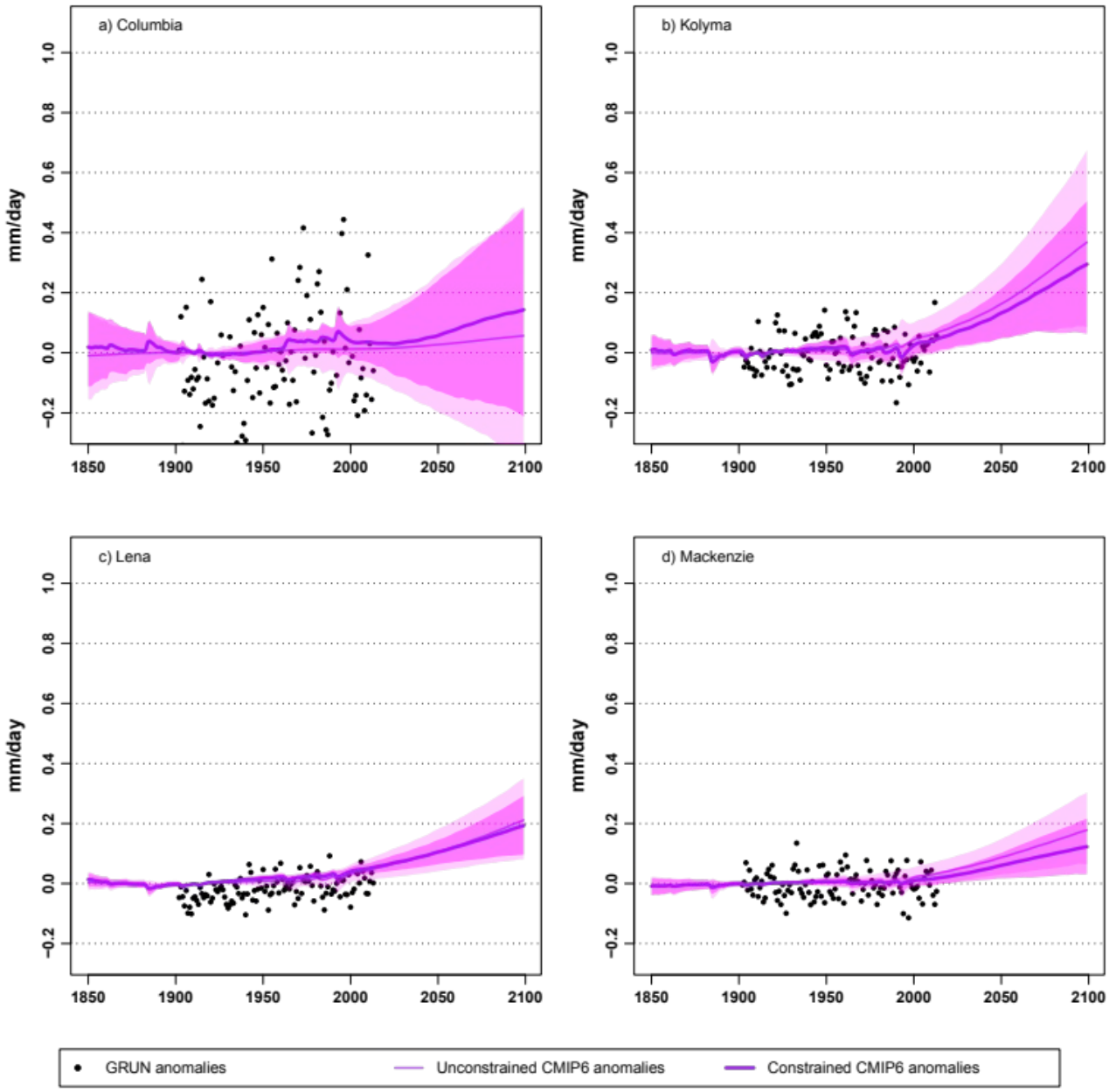

Figure S5: Constrained versus unconstrained water-year runoff anomalies (mm/day) using the rcp81 ensemble (a single realization of each CMIP5 model under the RCP8.5 scenario). Similar to Fig.2 but using the CMIP5 models. Depending on the CMIP model ensemble, the *prior* distribution is not the same but the effect of the KCC constraint on the *posterior* distribution is qualitatively consistent.

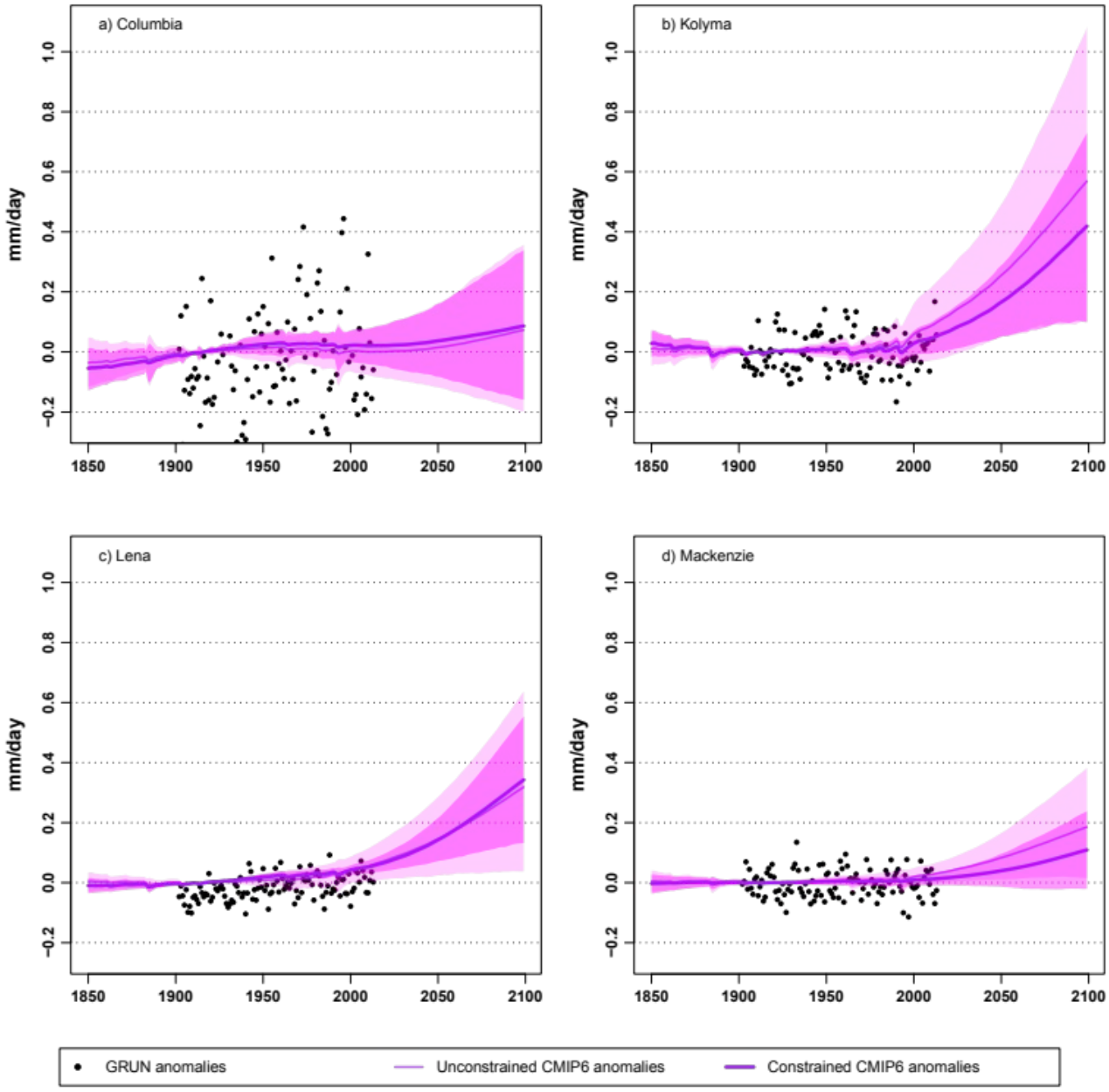

Figure S6: Constrained versus unconstrained water-year runoff anomalies (mm/day) using the ssp50 ensemble (multiple realizations of each CMIP6 model under the RCP8.5 scenario). Similar to Fig.2 but using multiple realizations for a lower number of CMIP6 models. Not surprisingly, differences between constraints using ssp50 and ssp51 scenarios are light, a smoothing of the distribution can be observed when more realizations are used (ssp50), less internal variability appears. There are no clear differences concerning the reduction of the spread.

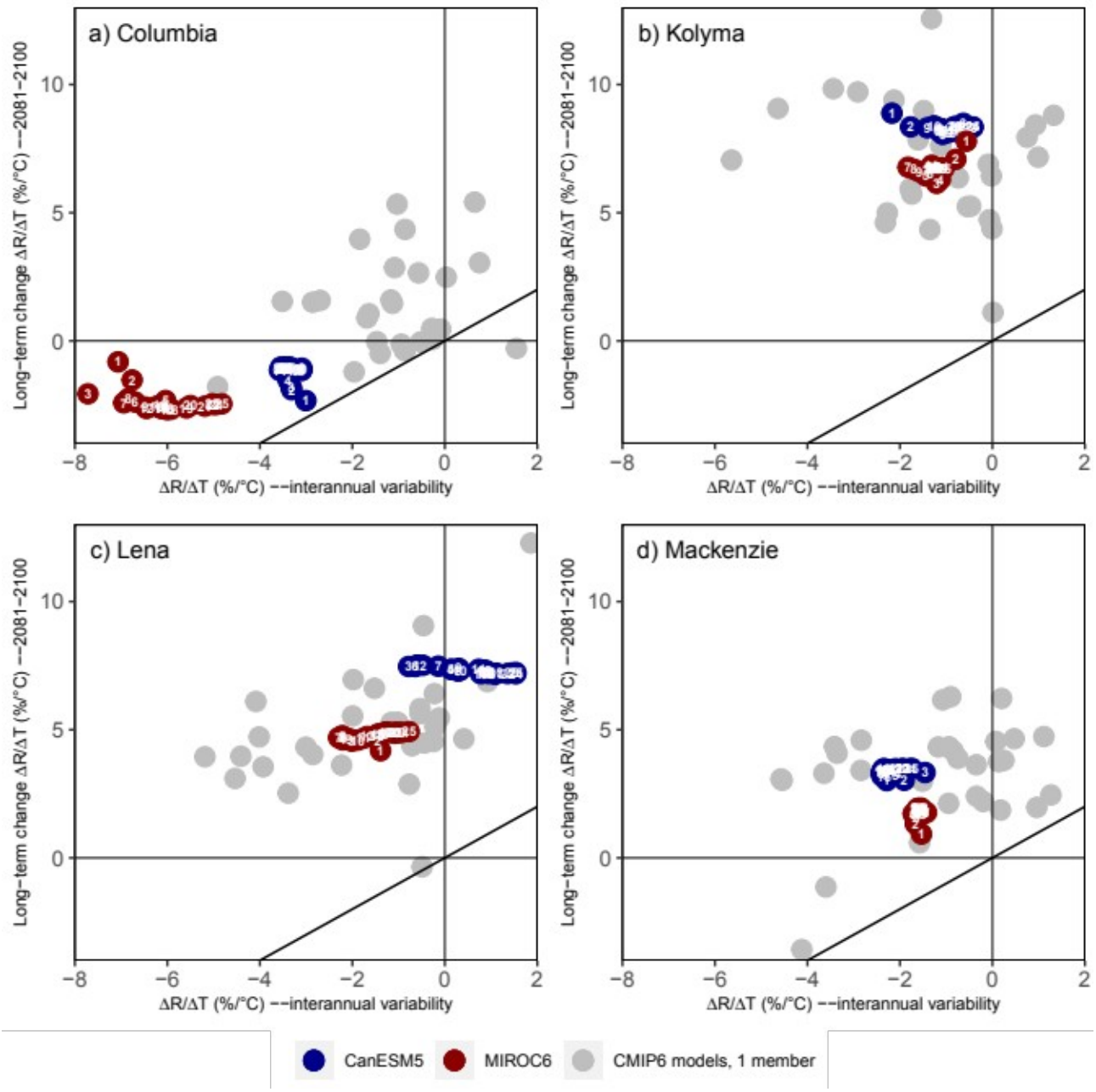

Figure S7: Runoff sensitivity to temperature at short (x-axis) versus long (y-axis) timescales. For each model (grey dots), the variation in runoff relative to the variation in temperature (%/°C) at the basin-scale is computed at two timescales. The long-term runoff sensitivity is estimated as the ratio of the averaged simulated anomalies over the 2081-2100 period compared to the 1902-1930 baseline. The short-term runoff sensitivity is estimated as the interannual variability over the 1902-2013 period (also used to estimate the regression coefficients in L19). The same computations are done for the CanESM5 and MIROC6 models (with 25 available realizations) after averaging an increasing number of realizations (for instance, dot 20 corresponds to average of the first 20 members). There is no obvious link between the runoff sensitivity to temperature at long versus short timescales. Therefore, the L19 hypothesis regarding the timescale independency of the runoff sensitivity to temperature does not seem valid.

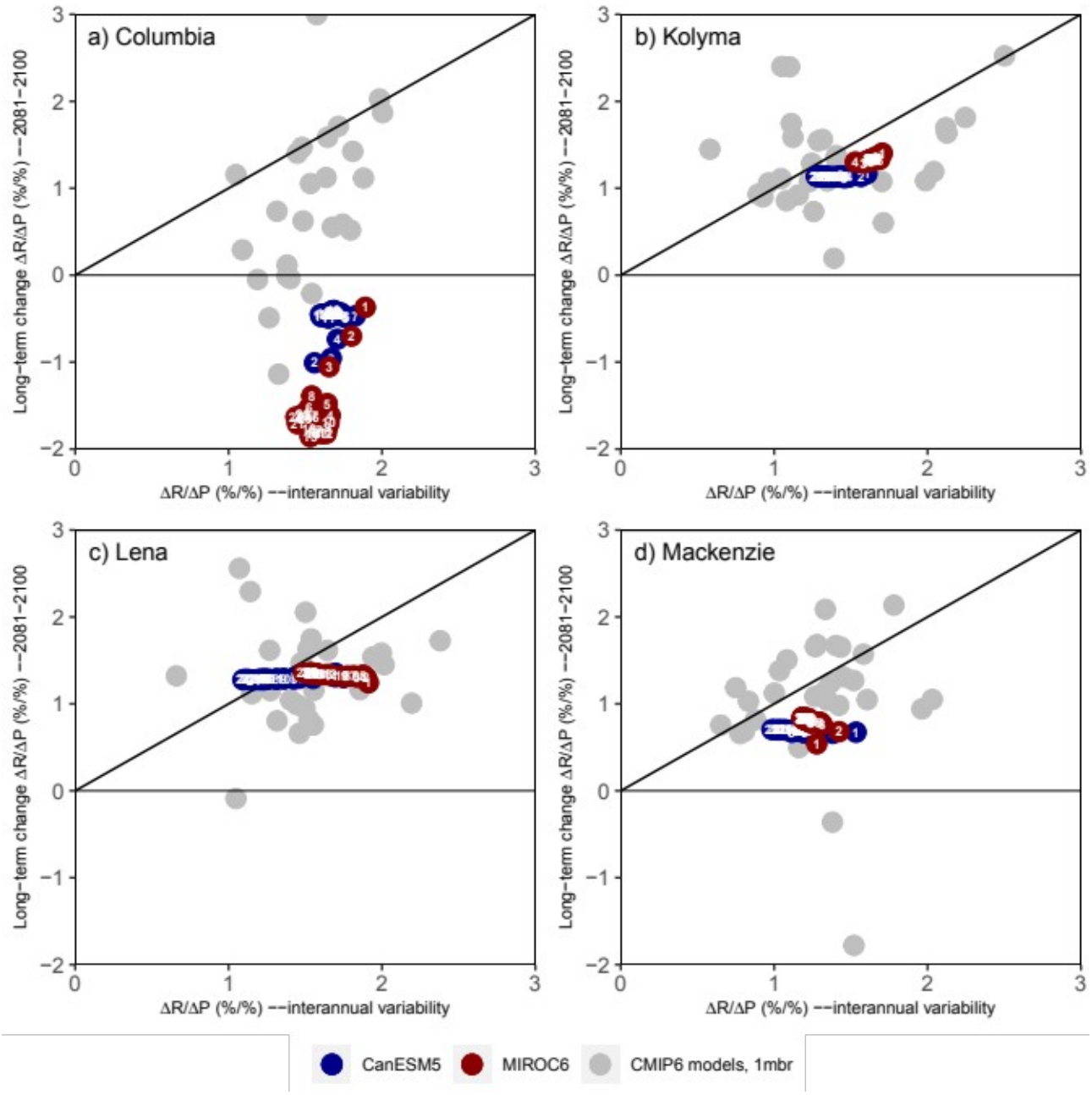

Figure S8: Runoff sensitivity to precipitation at short (x-axis) versus long (y-axis) timescales. Similar to Fig.S7 but comparing long and short-term runoff sensitivity to precipitation instead of temperature. Once again, the interannual runoff sensitivity is not a good surrogate for its long-term sensitivity to precipitation.