We would like to thank the reviewer for their thorough response to our manuscript. Their comments will be very helpful to improve our manuscript. We are glad to take the opportunity of this discussion format to address the points they raise.

My first comment is more a question out of curiosity, as the overestimation of LAI by the model intrigues me. It will strongly depend on the formulation of leaf area dynamics in Eq. 1, so how confident are you that this equation (or more the parameterization of this equation) is correct? How many trees were for example used to derive the allometric relations? Besides, the species specific parameters (like SLA, and a1,a2) are not reported, also not in the Supplement, so can you add these? So in general, could your leaf area formulation be the reason for the observed over-estimation?

The equation for leaf area and its parameterization are taken directly from the original TreeMig. They have originally been parameterized by Bugmann (1994, 1996) for the gap model FORCLIM, which shares many process formulations with TreeMig. The basis for the parameterization is the dataset collected by Burger (1945 - 1953), consisting of measurements of tree height, diameter and leaf area on 583 trees of five species or species groups. Tree species not represented in the dataset are assigned to one of the represented species, still following Bugmann (1994). The number of trees for each species, and specific parameters are given in the appendix of Bugmann (1994). Since this document is not widely available, we will repeat this information in a new version of the Supplement.

I am also a bit confused by equation 3. The fractional cover in LSM's or remote sensing products is often related related to LAI by the Lambert-Beer relation: FC = 1 – exp(-K * LAI), where LAI is the total leaf area index (or crown index), FC is fractional cover, K is an extinction coefficient, (e.g. Bréda, 2003; Choudhury, 1987; Monsi, 2004). The extinction coefficient is a function of leaf inclination and often set to 0.5. Here, this seems to be set to 1 for all species, which seems a bit high, is that correct? In addition, why are the exponents summed? Shouldn't you just add up the different final fractional covers of the species when the area stays the same? This is also what you describe on page 17 (if I am not mistaken), where you take the cumulative sums of the classes.

The calculation of fractional cover in FORHYCS is independent from the calculation of LAI. Instead, it is based on crown area, which is calculated from tree height using species-specific empirical relationships (the formula and its species-specific coefficients are reported in the supplementary material of Zurbriggen et al. (2014), Section B5). This way, it is not necessary to estimate an extinction coefficient.
The procedure used here and in Zurbriggen et al. (2014) was originally developed by Crookston and Stage (1999). It is based on the assumption that trees are randomly distributed in space (which is consistent with the way light penetration is calculated in TreeMig) and accounts for overlap between crowns. On page 17, the same procedure is applied. For example, applying Eq. 3 to the upper 3 height classes will return the fractional cover for the trees belonging to these classes, accounting for overlap between them. This assumes that shading of lower parts of the crowns by smaller trees can be neglected.
In a revised version of the manuscript, these two assumptions (random distribution and no shading by shorter trees) will be explicitly stated in Section 2.2.3 (currently p. 17). In

addition, to clarify that the same procedure is being used, a modified version of Eq. 3 will be introduced in the same section:

$$f_{c,i} = \left[1 - exp\left(-1 \times \sum_{sp=1}^{nspc} \sum_{hc=i}^{nhcl} (n_{sp,hc}/833) \times CA_{sp,hc}\right)\right],$$

where $f_{c,i}$ is the fractional cover of the i upper height classes.

In addition, LAI and fractional cover are compared by two newly developed error measures, which only compare on one specific moment in time. However, getting the seasonality right in these models is quite important, and one of the minimum things the model should be able to represent is the seasonal signal. Did you compare the timeseries of simulated and observed LAI? It's rather simple to do, and, in my view, provides much more information then the error measures as presented by the authors. So how well is the seasonality captured by the model?

Indeed, these metrics focus on the canopy structure at full foliage cover. This is intentional, as their purpose is to evaluate the forest structure and improvements over stand-alone TreeMig (hence the comparison with TreeMig in Fig. 7 and 8).
An evaluation of the intra-annual variations in leaf area was not carried out for the following reasons:
  - As seen on Fig. 6, the observed and simulated distribution of species do not match well (for the reasons discussed in Section 4.2, p. 34 l 4-9). Therefore, a good fit to phenological observations is not to be expected.
  - In another study (Speich et al., 2018a), the sensitivity of a water-balance model (corresponding to the surface water balance part of FORHYCS) to vegetation properties was assessed. It was found that long-term water partitioning was not very sensitive to growing season length (defined as the number of days with full foliage), as compared to LAI at full foliage. As long-term annual streamflow is the main hydrological output of interest in this case study, we chose not to evaluate simulated phenology in detail.
  - While they do not constitute a validation, the error metrics obtained during the calibration of the phenology submodel (reported in Tables S3 and S4 in the Supplement) give an indication of its strength. As discussed in the Supplement, spring phenology can be reproduced reasonably well in most cases, whereas autumn phenology is more problematic. This is consistent with other studies where empirical phenological models were applied.

I am also a bit confused on how the effect of elevated CO2 is studied. How can you evaluate the effect of elevated CO2 if you switch off the stomatal response to high CO2 (P22.L3)? Do you mean you keep the conductance the same? It would be much more interesting to keep the feedbacks in place, so why you do this? However, later on in the manuscript, the stomatal conductance is discussed, so can you clarify what you do exactly?

Our formulation here may indeed be unclear. In most simulation runs, the effect of elevated CO_2 on stomatal resistance (Eq. 12) is active, i.e. stomatal resistance is impacted by atmospheric CO_2 concentration. As this is a new addition (there is no CO_2 effect either in PREVAH or in the water balance model described in Speich et al. (2018a)), the purpose of the NCS runs is to test the strength of this effect. Therefore, in the NCS runs, the stomatal effect of CO_2 is switched *off*, i.e. Eq. 12 is set to 1. Comparing

the NCS runs with the standard runs will give an indication of how strong the CO_2 effect in the model is.

I am not too familiar with PREVAH, unfortunately, but the authors state that the model structure is similar to HBV. That would mean there are also several parameters that do not relate to vegetation (such as recession parameters, routing, snow parameters), so how are these determined? It can also be seen in Figure 5 that snow melt and recessions are quite off compared to the observations, which is probably just due to the remaining parameters. It may also affect the conclusions based on the climate change scenarios, as the snow melt is highly affected by the temperature changes.

Indeed, there are various parameters related to non-vegetation aspects. Some of these parameters are constant for the whole study area; others are spatially variable and have a different value for each grid cell. The spatially variable parameter values were determined in a previous study (Zappa and Bernhard 2012) based on the regionalization method of Viviroli et al. (2009). The spatially constant parameters were also taken from previous studies. As a different dataset for soil water holding capacity was used in this study (see Section 2.2.2), some parameters were manually adjusted to improve the optical fit of the streamflow lines. Due to the proof-of-concept nature of this study, a full calibration was not undertaken.
In a revised version, we will provide the reference for the spatially variable parameter values, as well as the values used for the spatially constant parameters, in the Supplement.

My last, but most important comment is on several key-findings which do not seem to be (entirely) supported by data. For example, one of the key findings presented in the manuscript concerns the effect of the climate change scenarios on streamflow. However, the result of only one specific catchment is shown in Figure 10, how do the results for the other catchments look like? Similarly, an analysis on elevated CO2-levels is described, but no results are shown in any of the graphs. Please add some graphs and evidence to support the statements you make here.

We agree that these figures are necessary to give the full picture. At the end of this document, we include the future streamflow projections and differences between model configurations (equivalent to Fig. 10) for the other four catchments, as well as the differences in streamflow for model runs with and without CO_2 effect on stomatal resistance. In a revised version, we will include these figures in the Supplement and refer to them in the Discussion.

Minor comments

I would like to suggest to change names of the modelling scenarios into more meaningful names, or add clarifications in the text when discussing a certain scenario. Names like Succ_TM_BEK, or T6_P10, are not very informative and make it hard to understand what happens without looking at the table all the time.

We welcome this comment, as this may indeed be a factor that makes it difficult to follow the text. We will take this into consideration when submitting a new version, and modify the text and figures accordingly.

P8.L8-9. In this way, the equation does not seem consistent in units. What is the unit of Pd,sp?

The phenological status $p_{d,sp}$ is dimensionless and ranges from 0 to 1. Its purpose is to scale leaf area when the foliage is not fully developed (in autumn, winter and spring). We forgot to specify the unit [-] for this variable in the text and will correct this in a revised version.

P8.L10. This seems a rather arbitrary number to me, why 833 m2 ?

This number has its origin in the gap model FORCLIM (Bugmann 1994, 1996), from which many process formulations of TreeMig were taken. In FORCLIM, 833 m^2 is the reference area of a simulated forest plot (roughly equivalent to the crown area of a large, dominant tree).

P8.L18-20. How does crown area relate to leaf area?

As noted in our response to the second comment, leaf area and crown area are calculated independently from each other, both using empirical relationships with tree size.

P9.L25-30. So the used transpiration values are model outputs, correct?

Yes, both actual and potential transpiration are simulated in the surface water balance part of the model.

P10.L9. Is fDS not a single yearly value, as DI is a single year value too? What do you use to calculate the geometric mean in that case?

Indeed, there is some information missing here to properly follow. Modeled tree growth depends on an environment-dependent function ranging from 0 (maximum stress) to 1 (unstressed conditions). This function is the geometric mean of three functions:

- The drought stress function fDS (Eq. 6)
- The effect of degree-day sum (Eq. S8 in the supplement)
- A stress function describing the effect of nitrogen supply

The last function is not mentioned in the manuscript, as the nitrogen supply is kept constant over the whole study area and period. Nevertheless, as this part of the model cannot be described without this function, it will be included into the supplement.

P10.L18. "i.e. with a decrease...is at kDT", I am not sure I follow, can you please clarify?

There is a mistake in this sentence – this is not about LAI, but tree height. The correct version is (also modified for additional clarity): "The former is parameterized following

Rasche et al. (2012), i.e. species-specific maximum tree height may be reduced as a function of the bioclimatic indices DI and DDEGS. The parameter k_{redmax}, which is also species-specific, indicates the fraction of maximum height that can be attained by trees if one of the environmental vitality functions is at its minimum. The more severe of the two reductions (drought or degree-days) is applied."

P11.L1. he –> the

Thank you – this will be corrected.

P11.L26. Why did you use these numbers? Seems a bit arbitrary.

First, we noticed that there was a mistake in the way Eq. 12 is reported. The parameter j_c is equivalent to (1-a) in Medlyn et al. (2001) (their Eq. 5). Therefore, the correct version of Eq. 12 is:

$$f_5 = \left(1 - j_c \left(\frac{\min(C_a, 700)}{350}\right) - 1\right)^{-1}.$$

The values for j_c were set based on the results reported by Medlyn et al. (2001): coniferous species had a value of $(1 - a)$ between 0 and 0.2, whereas broadleaves had values up to 0.4. Therefore, for conifers, a value of 0.1 was selected. For broadleaves, as there seemed to be some acclimation for trees growing in elevated CO_2, a more conservative (than 0.4) value of 0.25 was chosen. The value for mixed forests corresponds to the arithmetic mean of the two.

P13. So PPo includes the carbon costs? What are these values based on?

This variable is described in Speich et al. (2018b), and combines the plant-specific characteristics of Eq. 13 as follows:

$$PP_o = \frac{\gamma_{r,20} D_r}{L_r w_{ph}},$$

where $\gamma_{r,20}$ is the root respiration rate at 20 °C. The actual root respiration rate is dependent on annually averaged temperature via a Q_{10} function, as described in Speich et al. (2018b). These details are indeed important, and will be included in the new version.

P16.L1-5. If a large amount is diverted by pipelines, can you compare modelled and observed discharge? Which sub-catchments are affected by this?

The streamflow data used in this study was obtained from the company operating the power plants and includes the amount of water diverted through the different pipelines. From this, time series of natural streamflow were reconstructed, which were used as observations.

P16.L27. "As the sampling plots... a larger area." This sentence is a bit unclear to me, what do you mean?

Here, we explain why it does not make sense to compare simulations and observations at the scale of single inventory sampling plots (which is sometimes still being done). New formulation to clarify: „As the sampling plots of the NFI are distributed on a regular grid, each plot is randomly selected from all forest plots in that region, and may not be considered representative for a larger area. *It is therefore not sensible to compare simulated and observed biomass at the scale of single inventory plots. Instead*, the 245 NFI plots in the study area were aggregated to seven classes based on aspect and elevation, with four elevation bands for North-facing plots and three for South-facing plots. This way, each class has a sample size of at least 30 plots, which ensures that the averages are representative.“

P17.L19. Please correct reference

Thank you for catching this.

P22.L31 The plot only shows Acer spp., so how can I see this?

This is indeed not visible on Fig. 6. However, it seems that differentiating between the three Acer species will add little value to the figure (especially as the Acer spp. band is quite thin in all plots), while making the plots harder to read. We will rephrase the text to make it clear that this is not visible on the plots.

P24.L21. This sounds a bit counter-intuitive, wouldn't you expect a higher biomass when there is no LAI-reduction? What is the reason for this?

An important point to keep in mind (which should be made clearer in the manuscript) is that TreeMig/FORHYCS does not explicitly simulate carbon cycling/allocation (see also response to the next point). Hence, there is no direct effect of LAI reduction on biomass. Such effects are implicitly simulated through the environmental stress functions such as Eq. 6.
Two main effects happen in the model as a result of LAI reduction: the drought index (Eq. 5) is lower than it would be without LAI reduction; and the light distribution is modified, i.e. lower height classes get more light than they would get without LAI reduction. These two effects both promote tree growth, which explains why the model simulates higher biomass. This higher growth also eventually leads to greater mortality, after the number and size of trees have grown fast for some years. This explains the more dynamic pattern when LAI reduction is activated (Fig. S9).
These effects indeed need to be discussed in the text, and will be included in Section 4.2 (Effect of coupling on forest simulations) in a revised version.

P29.L10-11. Is this not counter-intuitive too? You would expect (also because of equations 7 and 8, where additional carbon is allocated for roots under stress) that the roots will go deeper in case of drier scenarios, and that LAI would go down, correct?

As noted in the response to the previous comment, there is no simulation of carbon uptake and allocation in the model. Eqs. 7 and 8 have no (direct) influence on the development of roots – they only influence the LAI reduction (which is switched off on the simulation described here). These equations were included because they represent a sound and plausible way to parameterize reduction of leaf area due to environmental

stress. This comes at the price of inconsistent formulations between Eqs. 7-11 (where carbon allocation fraction to roots is calculated as an auxiliary variable to determine the degree of leaf area reduction) and Eq. 13 (which determines rooting depth). As the allocation fractions of Eqs. 7-11 are not used anywhere else, we argue that this inconsistency can be tolerated. However, in a new version of this manuscript, we will need to make it clearer that Eqs. 7-11 only affect leaf area.

The rooting depth scheme used in FORHYCS assumes that plants dimension their rooting systems to optimize for net carbon gain. In this scheme, rooting depth does not necessarily increase with a drier climate. In some cases (like the low-elevation regions under drying scenarios, as discussed here), it may not be worth it for the plants to invest more carbon into roots. This is a point that should be included in the discussion (where a link can be made to Speich et al. (2018b), where the behavior of this rooting depth scheme was examined under various environmental conditions).

P31.L4-5. You described earlier that Eq. 12 was set to 1, correct?

Only in the NCS runs – in all other runs, the effect of CO2 concentration on stomatal resistance is activated (see our response to the corresponding point above).

P32.L4-14. How different are the landuses eventually at the end of the runs? Are the differences mainly due to different forest covers under the different scenarios?

There is no other mechanism for land cover change in the model than forest growth or retreat. In the runs with land-cover change enabled, the forest biomass in the cells initially belonging to the „potentially forested" land cover classes reaches up to 150 t/ha at the end of the simulation (these values vary with climate scenario and elevation band). This information is indeed important to follow the presentation and discussion of these results, and will be included in the new version.

P33.L14. Based on the data as shown, you cannot claim that the sensitivity of streamflow to vegetation properties varies spatially. This would also mean you need to have the same vegetation in different places and observed different changes in streamflow, but I believe that is not the case.

This sentence was indeed formulated in an ambiguous/misleading way, and this is not the point that we wanted to make here (the sensitivity of water balance to absolute values of vegetation properties was the subject of another of our studies (Speich et al. 2018a)).

Rather, the point here is that the effect of model coupling (i.e. of dynamically varying the values of vegetation properties) varies spatially. This statement (which will be corrected/clarified in the new version) is directly based on the results discussed immediately before.

P37.L7-8. "the greatest effects occurred at low elevations, and in regions currently above the treeline", where do you show this? Please back this up with some evidence in the main manuscript, especially when it is a key-finding.

This is directly linked to the statement discussed in the previous point, with the greatest changes in streamflow occurring in the Chippis subcatchment (Fig. 10 and its

equivalents for the other subcatchments, to be included in the supplement) and in the currently unforested, high-elevation subcatchments if forest is allowed to grow there (Fig. 11).

Eq3. Please define all variables and subscripts

Table1. There two Succ_noHmax-scenarios, please correct.

Table 2. Please describe the abbreviations in the caption or replace them with a description.

Fig1a. Please define SFC also in the figure.

OK to all

Fig5. There are a couple of things that seem a bit odd to me in this plot. In Fig 5a, Prevah seems to be much closer to the observations then the other two model set-ups, but in Table 2 the KGE-values are lower. Is that correct? In addition, Forhycs00 and Forhycs11 are on top of each other in Figure 5a, whereas Figure 5b suggests a difference of up to 0.3 m3/s.

What may create some confusion in this figure is that the lines in Fig. 5a are presented as 30-day rolling averages, while the line in Fig. 5b is not. For example, at the time of the greatest difference (Fig. 5b) in early 2005, the lines on Fig. 5a depart slightly but visibly from each other. If the lines on Fig. 5a were not presented as rolling averages, this difference would be more pronounced, but the plot would be more difficult to read.

Code availability: I would suggest to share your code on github or gitlab, instead of the supplement. Please also add links to the actual datasets used in the study.

It is feasible to create a public repository for the model code. However, due to the different levels and structure of documentation in the code of the original models, it may be rather difficult to study the model code.
Due to restrictions from the provider of the meteorological data, it is unfortunately impossible to give access to the data used to drive the model.

Appendix A: Why is there an appendix in the main manuscript and also a supplement? Should Table A1 not just be part of the Supplement then?

This can be done, it is certainly a good idea to keep the main document as lean as possible.

References

Bugmann, H. (1994): On the Ecology of Mountainous Forests in a Changing Climate: A Simulation Study, Ph.D. thesis, ETH Zurich, Zurich.

Bugmann, HKM (1996): A Simplified Forest Model to Study Species Composition Along Climate Gradients. Ecology 77, 7, 2055-2074. https://doi.org/10.2307/2265700

Burger, H. 1945, 1947, 1948, 1950, 1951, 1952, 1953. Holz, Blattmenge und Zuwachs. Mitteilungen der Schweizerischen Anstalt für das forstliche Versuchswesen 24:7-103, 25:211-279, 25:435-493, 26:419-468, 26:583-634, 27: 247-286, 28:109-156, 29:38-130.

Crookston, NL, Stage, AR (1999): Percent canopy cover and stand structure statistics from the Forest Vegetation Simulator. Technical report, US Department of Agriculture. https://www.fs.usda.gov/treesearch/pubs/6261

Medlyn, B. E., Barton, C. V. M., Broadmeadow, M. S. J., Ceulemans, R., De Angelis, P., Forstreuter, M., Freeman, M., Jackson, S. B., Kellomaki, S., Laitat, E., Rey, A., Roberntz, P., Sigurdsson, B. D., Strassemeyer, J., Wang, K., Curtis, P. S., and Jarvis, P. G. (2001): Stomatal conductance of forest species after long-term exposure to elevated CO_2 concentration: a synthesis, New Phytologist, 149, 247–264, https://doi.org/10.1046/j.1469-8137.2001.00028.x

Speich, M. J., Zappa, M., and Lischke, H. (2018a): Sensitivity of forest water balance and physiological drought predictions to soil and vegetation parameters - A model-based study, Environmental Modelling & Software, 102, 213–232, https://doi.org/10.1016/j.envsoft.2018.01.016

Speich, M. J. R., Lischke, H., and Zappa, M. (2018b): Testing an optimality-based model of rooting zone water storage capacity in temperate forests, Hydrology and Earth System Sciences, 22, 4097–4124, https://doi.org/10.5194/hess-22-4097-2018

Viviroli, D, Mittelbach, H, Gurtz, J, Weingartner, R (2009): Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results. Journal of Hydrology 377, 1-2, 208-225, https://doi.org/10.1016/j.jhydrol.2009.08.022

Zappa, M, and Bernhard, L (2012): Klimaänderung und natürlicher Wasserhaushalt der Grosseinzugsgebiete der Schweiz. Technical report. Birmensdorf, Eidg. Forschungsanstalt WSL.

Zurbriggen, N., Nabel, J., Teich, M., Bebi, P., and Lischke, H. (2014): Explicit avalanche-forest feedback simulations improve the performance of a 5 coupled avalanche-forest model, Ecol. Complexity, 17, 56–66, https://doi.org/10.1016/j.ecocom.2013.09.002
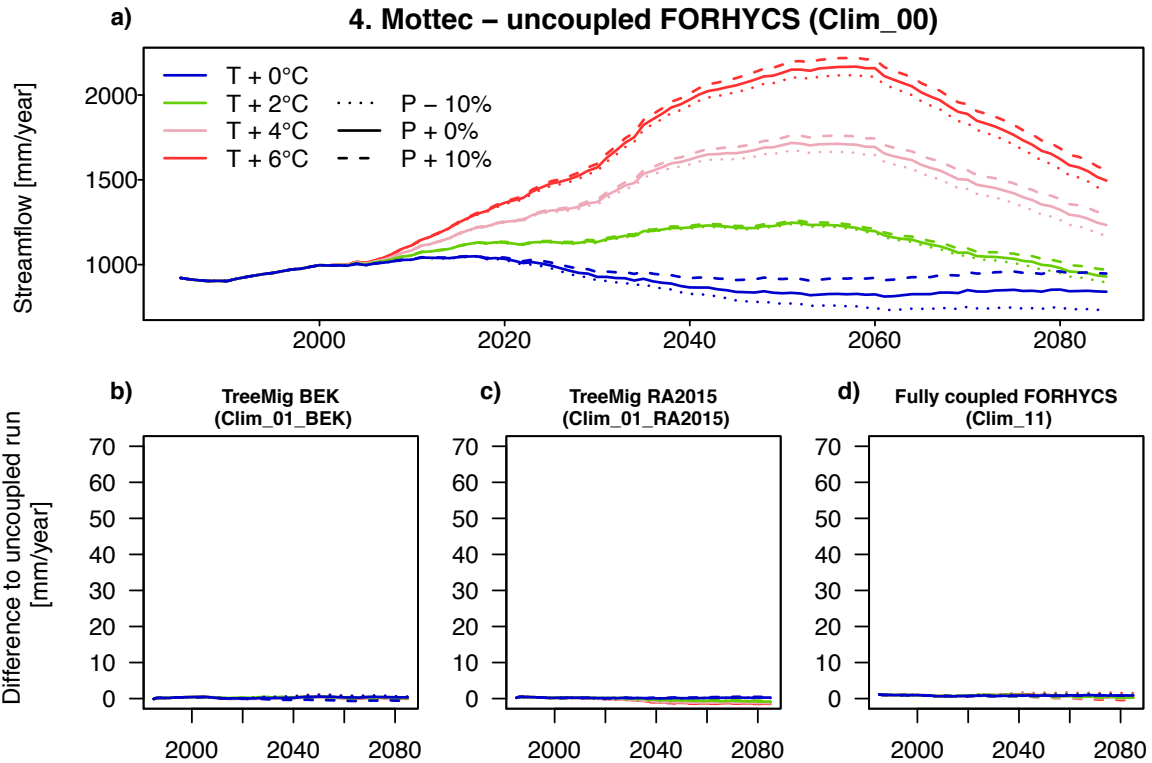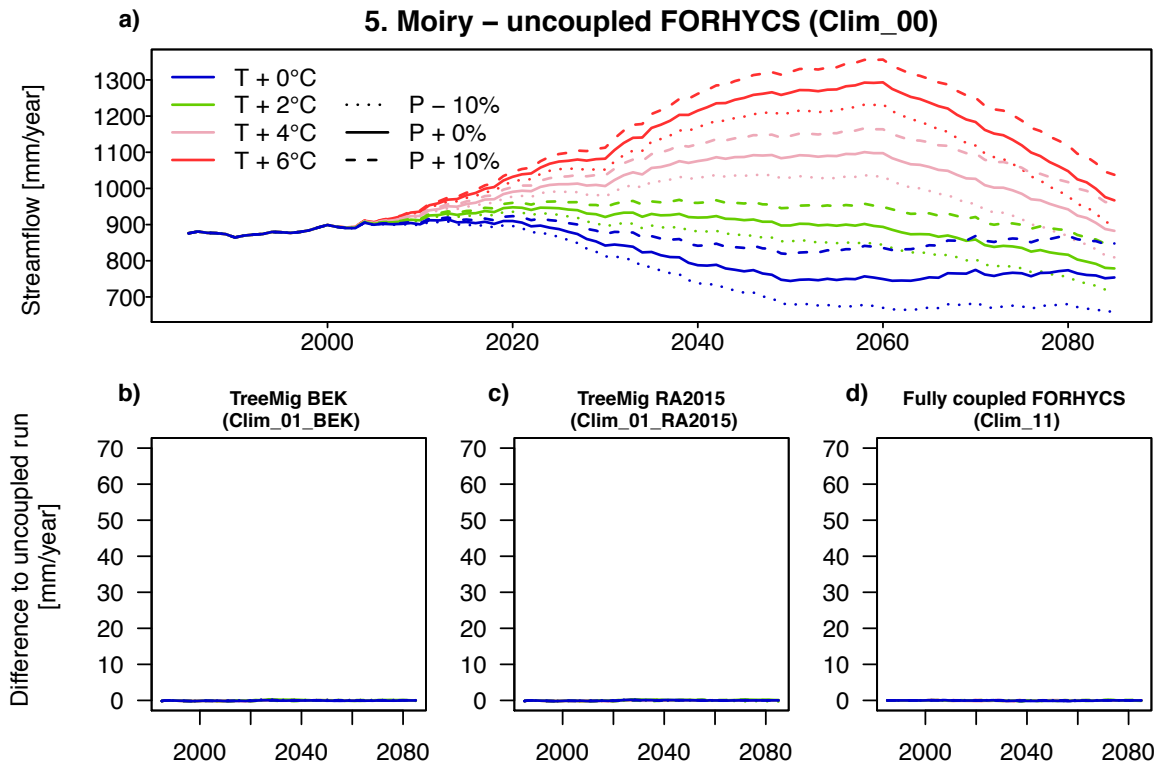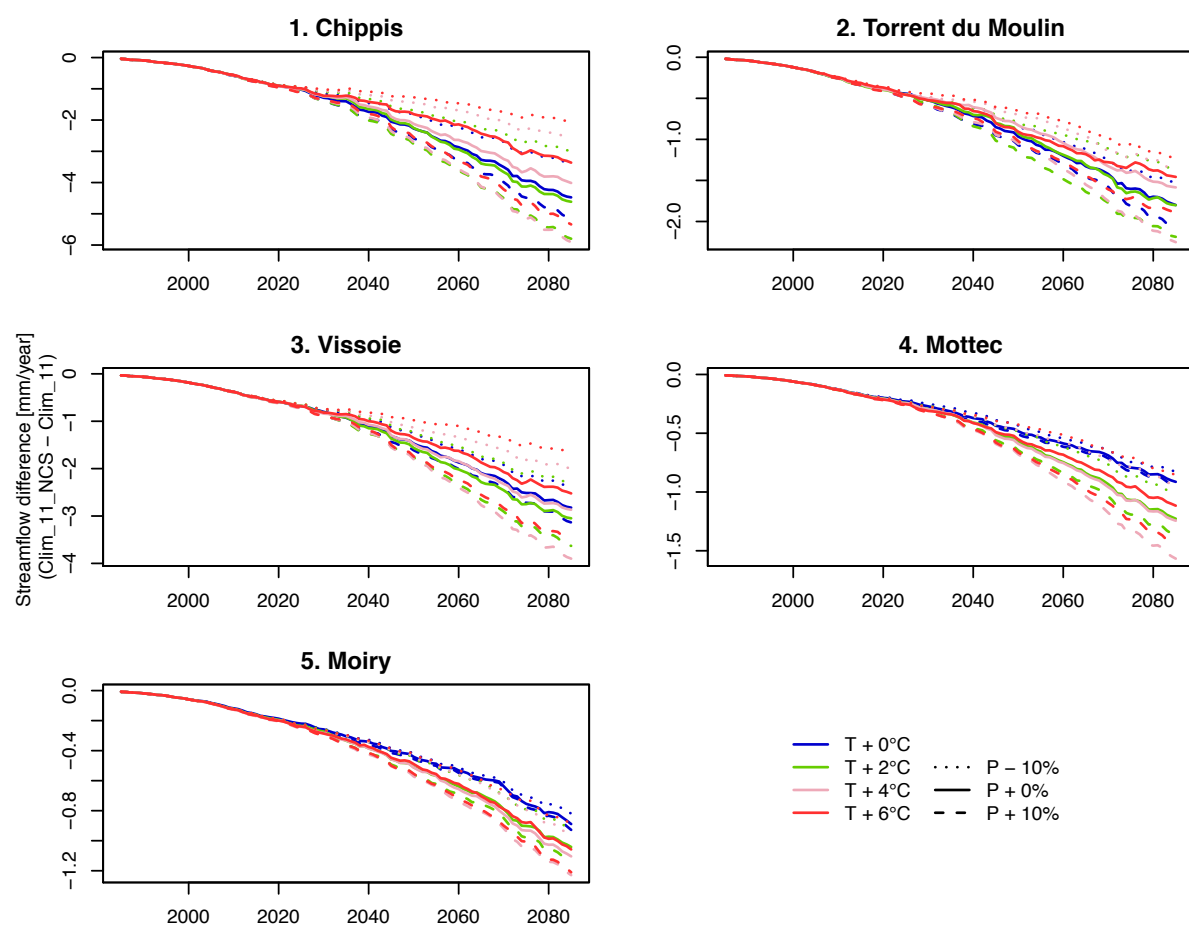
New Figures
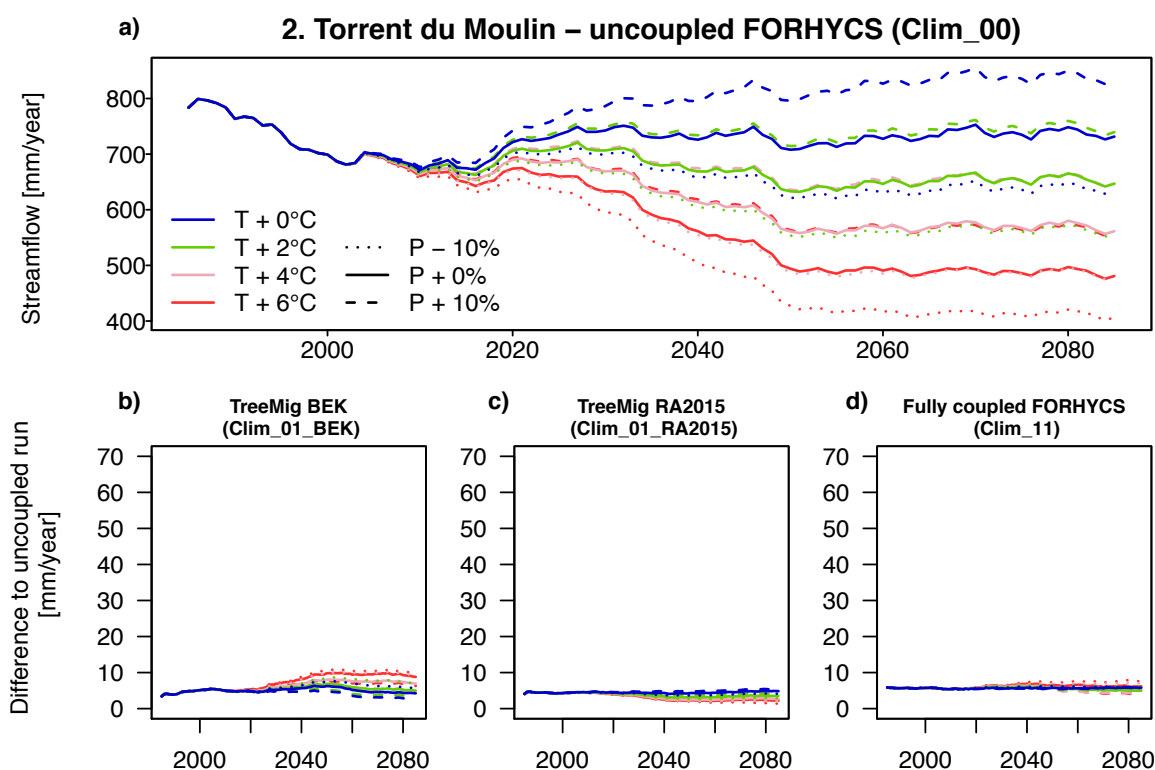
Equivalent to Fig. 10 for the Moulin subcatchment – to be added to the Supplement.
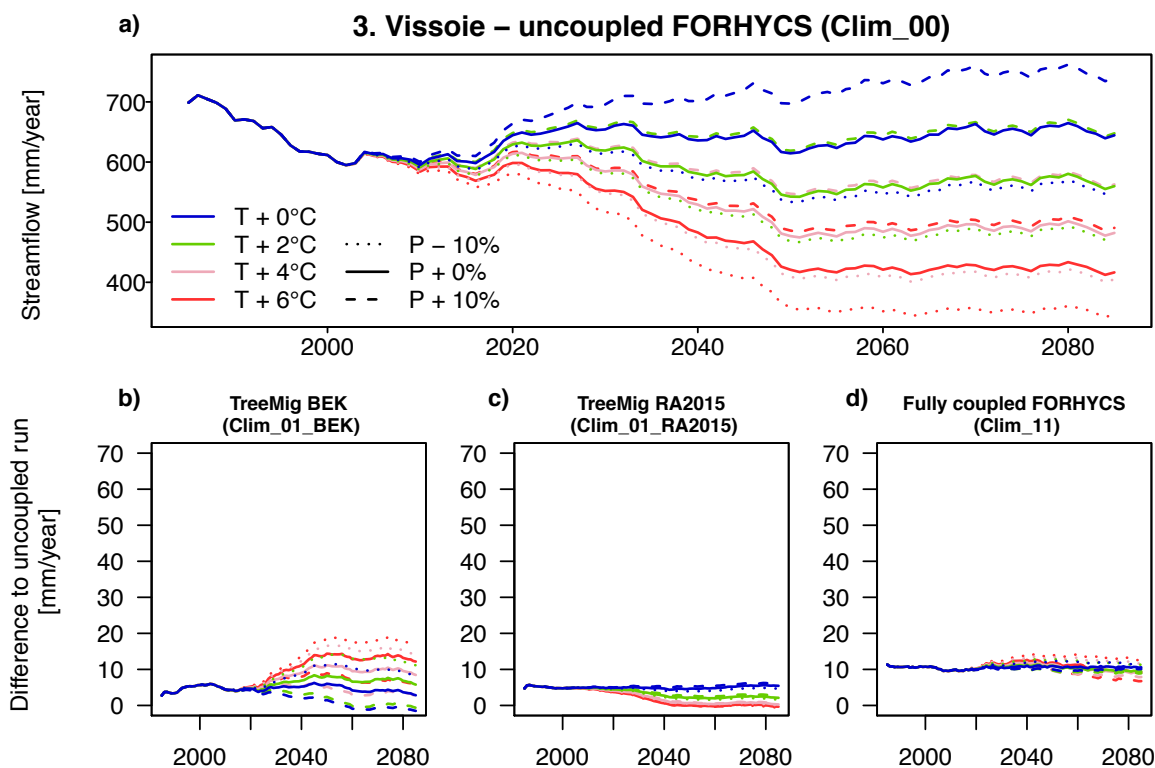
Equivalent to Fig. 10 for the Vissoie subcatchment – to be added to the Supplement.

a)

4. Mottec – uncoupled FORHYCS (Clim_00)

Legend:
- T + 0°C
- T + 2°C
- T + 4°C
- T + 6°C
- ······ P − 10%
- —— P + 0%
- – – – P + 10%

b) TreeMig BEK (Clim_01_BEK)

c) TreeMig RA2015 (Clim_01_RA2015)

d) Fully coupled FORHYCS (Clim_11)

Equivalent to Fig. 10 for the Mottec subcatchment – to be added to the Supplement.

a)

5. Moiry – uncoupled FORHYCS (Clim_00)

Legend:
- T + 0°C
- T + 2°C
- T + 4°C
- T + 6°C
- ······ P − 10%
- —— P + 0%
- – – – P + 10%

b) TreeMig BEK (Clim_01_BEK)

c) TreeMig RA2015 (Clim_01_RA2015)

d) Fully coupled FORHYCS (Clim_11)

Equivalent to Fig. 10 for the Moiry subcatchment – to be added to the Supplement.

Modification of simulated mean annual streamflow if the effect of CO₂ concentration on stomatal resistance is enabled. To be added to the supplement.