Referee comments are shown in black. Our responses in bold and yellow.

RC1: ‘Comment on hess-2023-255’, Anonymous Referee #1, 05-12-2023

General comment:

This work can represent a very significant contribution to improve our knowledge about the effect of forest structure and climate on water resources (or blue and green waters) in a very heterogenous context of climates and forest typologies such as Spain. Although it is true that it has been proposed at country level, in my opinion, the results presented are those expected in terms of the differences observed between the biomes and the types of forests present in the study area, which gives credibility to the results from modelling apart from other about the validation of results based on field data (transpiration, water in the soil, runoff, etc.). Its subsequent use in studies at higher spatial resolution (at the basin scale, for example) could help managers to better define forest management with a strong hydrological foundation (to improve green water, blue water, or both) and therefore promoting ecohydrology-based management to improve water resource availability at regional scales.

I value very positively the great effort made in terms of obtaining information for the characterization of the forest inventory plots, as well as other information necessary to apply the model on a scale as ambitious as Iberian Spain, with such diverse forest typologies and climates and extension.

Thank you very much for your positive comments.

Q1.1. However, I believe that the work still needs to be greatly improved to be finally published in a journal like HESS. In this sense, I consider very important to justify the model performance to produce realistic estimates for all the forest typologies addressed in the work. Although it is true that a recent work is indicated (De Cáceres et al., 2023) where the validation for some species with experimental field data is presented, this work should at least indicate how the latter has addressed the calibration and validation of the model, and not only relegating this information to the conclusions section. As I indicated, I understand this is not straightforward given you have not experimental plots in all the forest typologies, but I understand more efforts should be carried out to justify the use of the model for the entire territory of Spain and the validity of their results. In this sense, maybe you can consider available databases to discuss about the model performance (such as that for sapflow from https://essd.copernicus.org/articles/13/2607/2021/) when comparing your estimates to some hydrological component such as transpiration or runoff.

R1.1. We agree that a validation in more tree species or biomes can be necessary since De Cáceres et al. (2015) and De Cáceres et al. (2021) didn’t cover the whole country and all the forest typologies and the evaluation of De Cáceres et al. (2023) focused on forest dynamics. Nevertheless, there are a lack of field work data at country scale. Our current efforts include evaluating the model in other (temperate) forest types, using datasets like PROFOUND. However, we felt that for this paper the model evaluation should be made at larger scales, in accordance

with the scale of application. Thus, we have included a comparison with two flow water products, for green water and blue water respectively, at regional scale that can be more adequate for the scale of this study. We answered with more details in the R1.13.

On the other hand, although the fundamentals of the model are understood quite clearly, I believe that the work should not simplify relevant information to understand how the calculations have been carried out in the study, either in the materials and methods section or as additional information. For example, it is not explained how the LAI is calculated or how the data obtained for the IF3 plots are regionalized or how the modeled fluxes are extrapolated to the entire Spanish forest territory (except for the Canary Islands). In my opinion, this aspect must be clearly addressed, since in previous works related to your model performance, I am not able to understand how this is addressed. I believe this work is the correct place to give a proper explanation about how spatial dimension is considered within your model framework.

R1.2. We answered the question about the calculation of the LAI in the R.1.23. The question about the regionalization was answered in R1.15.

I believe English flows well and the manuscript is easy to be followed due to its clarity and well structure. I have, however, several suggestions and comments which have been added as comments within the manuscript. In this sense, apart from considerations about the modelling approach, I would recommend reviewing the introduction section for improving it (see please my comments on it).

Q1.3. L22: I recommend including values for this LAI plateau depending on forest type/climate. Take into account you did not include any quantitate information through the entire abstract, thus I would add more "numbers" to improve it in order to be less qualitative

R1.3. Thanks for the suggestion. We have included the values in the abstract: (around 2.5-3). The plateau can be seen for each functional group in Figure 6. Additionally, we have added some more quantitative results in the abstract.

Q1.4. L25: Too general conclusion, Please be more specific based on your results

R1.4. We changed the conclusions to: "This study highlights how the green water is decoupled of the blue water across the forests typologies of whole Spain and it how the species functional traits (deciduous vs evergreen) can influence the blue water production."

Q1.3. L34: Remove "On the other hand"

R1.3. Done.

Q1.4. L36: please include reference to green water

R1.4. Done: (Llorens et al., 2011)

Q1.5. L43: please define LAI; also, it is not the only variable used for characterizing stand structure. You need further justify why LAI can be the most optimum variable to describe stand structure rather than density or cover, among others

R1.5. We have included the definition of LAI in the introduction: "Stand forest structure is also key in rainfall partitioning through its relationship with stand LAI (total area of leaves of the canopy per unit horizontal ground area) which is a key variable determining transpiration (Granier et al., 2000) and/or the interception of vegetation." In the Methods section (Statistical analysis) we have explained why LAI (and also BA) are the variables used to describe stand structure: "Basal area serves as an indicator of canopy cover, as it is calculated by summing the diameters at breast height (dbh) of all trees per hectare, and higher basal area values can result in increased interception. LAI determine the transpiration of the trees through of the leaf surface."

Q1.6. L50: I miss more recent works in Mediterranean forests, and specially those carried out in Spain

R1.6. We included one field work carried out in a Mediterranean forest in Spain and another one in a Mediterranean forest in Israel: (Campos et al., 2016; Qubaja et al., 2020)

Q1.7. L67: I don´t understand the logic followed here; you first focused on two really contrasted scales and then you said that the effects of stand structure and traits are not well examined in the literature.

R1.7. We removed the sentences about the stand structure and traits to avoid the confusion.

Q1.8. L74: In my opinion, this is the key message of your introduction, However, I consider you should further review more recent articles in order to better justify your objectives but also improving the statements of your introduction. I find the introduction two general and not really supported by recent literature.

R1.8. Thanks for your suggestion. We added more information in the text: "The stand structure variables have been studied at local scale (Benyon et al., 2017; Simonin et al., 2007), but the landscape ecohydrological simulations have been carried out without incorporating the role of the stand structure variables or the differences in species and functional composition (Hoek van Dijke et al., 2022; Mastrotheodoros et al., 2020)." We hope the statement of our objectives is more clear now.

Q1.9. L95: You should also include information regarding geology, soils, geomorfology, Variation in climatic aspects is important but also other aspects should be added to give a correct general description of your study area.

R1.9. Thank you for your comment. But we did not include information about geology or soils because the basic descriptions of geology/soils in Spain mainly

focus on the difference between acidic and basic pH, which is not relevant for this work.

Q1.10. L101: please include references to these climatic values

R1.10. The climatic values were derived of the average climatic extracted for the SFI3 plots. But we removed the values to avoid confusion.

Q1.11. L111: LAI is in the end the most important stand parameter you considered. In this sense, you should include a detailed explanation about how it is estimated based on SFI data or whatever you considered.

R1.11. Thanks for pointing at this shortcoming of the original text. The variables and calculations that are based on SFI data are described in detail in a new section:

"2.5. Parameter estimation

Data from forest inventory plots included tree height (*H*) and tree diameter at breast height, which was used to obtain estimates of foliar biomass (hence leaf area after multiplying by *SLA*) and crown ratio (*CR*) via species-specific allometries (see Table S1-3 of De Cáceres et al., 2023 for more details). In the model, *SLA* (Specific Leaf Area (sq *mm/mg*)), *the ratio of leaf area to leaf dry mass,* is constant for every species (De Cáceres et al., 2023). Leaf area index (LAI in $m^2 \cdot m^{-2}$) was calculated from the foliar biomass (in $kg \cdot m^{-2}$) by using a *specific leaf area* coefficient (*SLA*, in $m^2 \cdot kg^{-1}$) that is species-specific (LAI = foliar biomass * *SLA*). Taxon-specific parameter details are shown in supplemental material of De Cáceres et al., 2023. Soil data of each forest inventory plot was extracted from the SoilGrids database (Hengl et al., 2017). For all plots four soil layers down to a total depth of 4 m were initially considered, but the deepest layers were merged into a rocky layer (95 % of rocks) following the depth of the R horizon. A monotonous increase in rock fragment content across soil layers from the surface to the rocky layer was defined based on surface stoniness classes determined in SFI3 plot surveys."

Q1.12. L144: nothing is said about blue and green water calculations. I understand runoff and soil percolation are considered as blue water but further information is required to understand how this is addressed on your study, In addition, what about soil evaporation as key component of green water? I would suggest to include a final paragraph explaining this

R1.12. We have included an explanation of the calculation of blue and green water in a new section:

2.6. MEDFATE simulations:

MEDFATE was run on each selected SFI3 plot using daily weather data (temperature and precipitation, PET, radiation and relative humidity) corresponding to a 10-yr period centered on the year of the SFI3 sampling (1997-2008). We calculated the blue water as the sum of the runoff and deep water and

the green water as the evapotranspiration (that included the sum of the transpiration, interception and soil evaporation). The percentage of interception and soil evaporation about green water now are showed in the Supplemental material (Figure S9-S14):

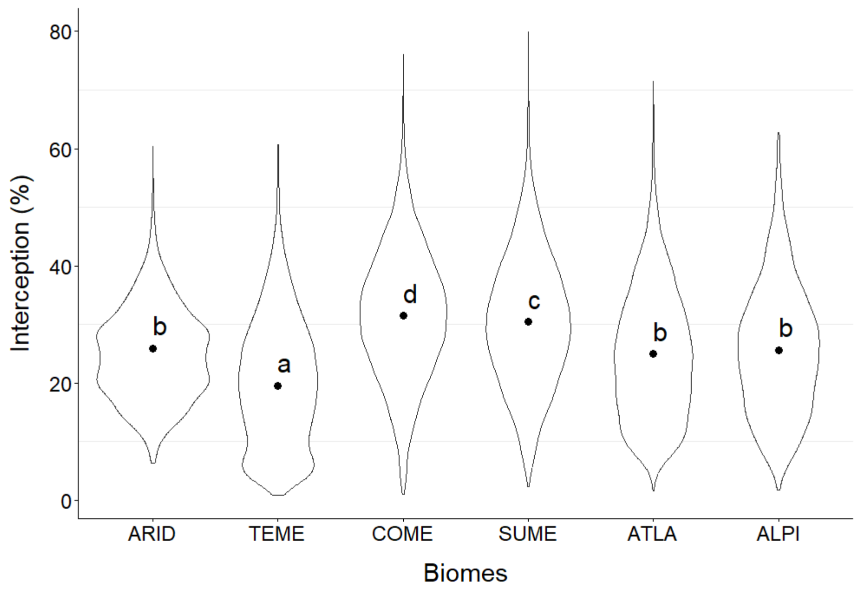

Figure S9. Percentage of interception in the different biomes. Different letters denote significant differences among biomes (*p* < 0.01) after Tukey's test. Abbreviations of the biomes: ARID, Arid; TEME, Temperate Mediterranean; COME, Continental Mediterranean; SUME, Submediterranean; ATLA, Atlantic; ALPI, Alpine.

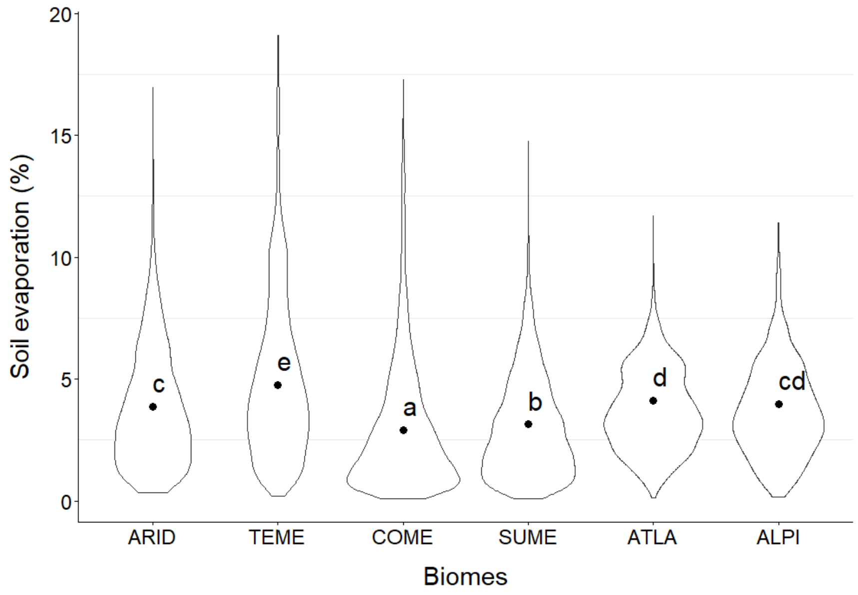

Figure S10. Percentage of soil evaporation in the different biomes. Different letters denote significant differences among biomes (*p* < 0.01) after Tukey's test. Abbreviations of the biomes: ARID, Arid; TEME, Temperate Mediterranean; COME, Continental Mediterranean; SUME, Submediterranean; ATLA, Atlantic; ALPI, Alpine.

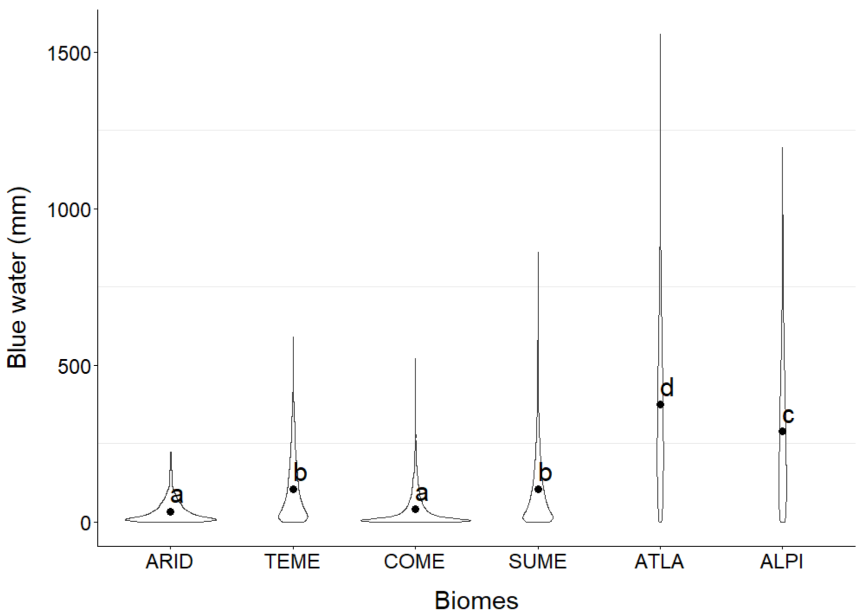

Figure S11. Blue water amount in the different biomes. Different letters denote significant differences among biomes (*p* < 0.01) after Tukey's test. Abbreviations of the biomes: ARID, Arid; TEME, Temperate Mediterranean; COME, Continental Mediterranean; SUME, Submediterranean; ATLA, Atlantic; ALPI, Alpine.

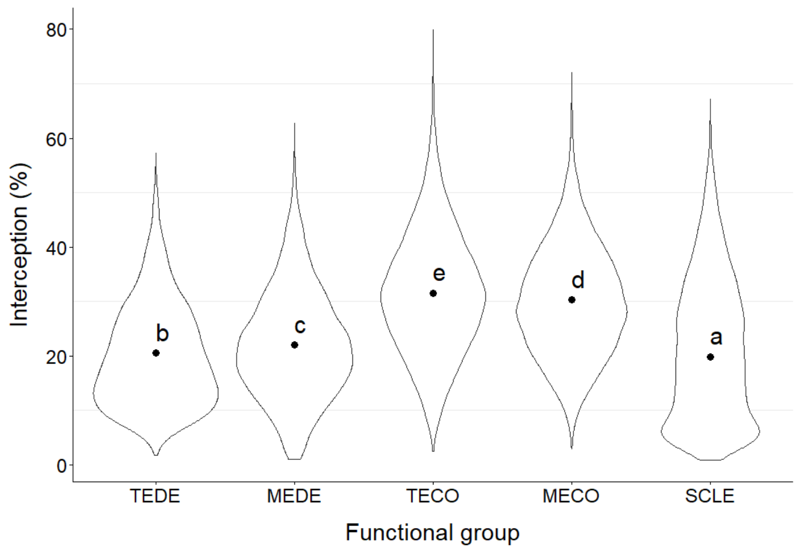

Figure S12. Percentage of interception of the different functional groups of species. Different letters denote significant differences among forest types ($p < 0.01$) after Tukey's test. Abbreviations of the functional groups: TEDE, Temperate deciduous; MEDE, Mediterranean deciduous; TECO, Temperate coniferous; MECO, Mediterranean coniferous; SCLE, sclerophylls.

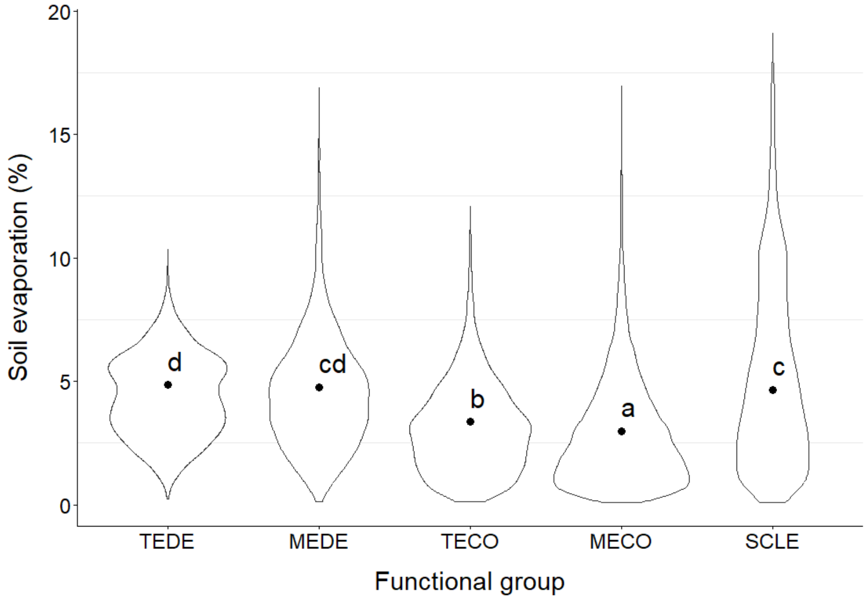

Figure S13. Percentage of soil evaporation of the different functional groups of species. Different letters denote significant differences among forest types ($p < 0.01$) after Tukey's test. Abbreviations of the functional groups: TEDE, Temperate deciduous; MEDE, Mediterranean deciduous; TECO, Temperate coniferous; MECO, Mediterranean coniferous; SCLE, sclerophylls.

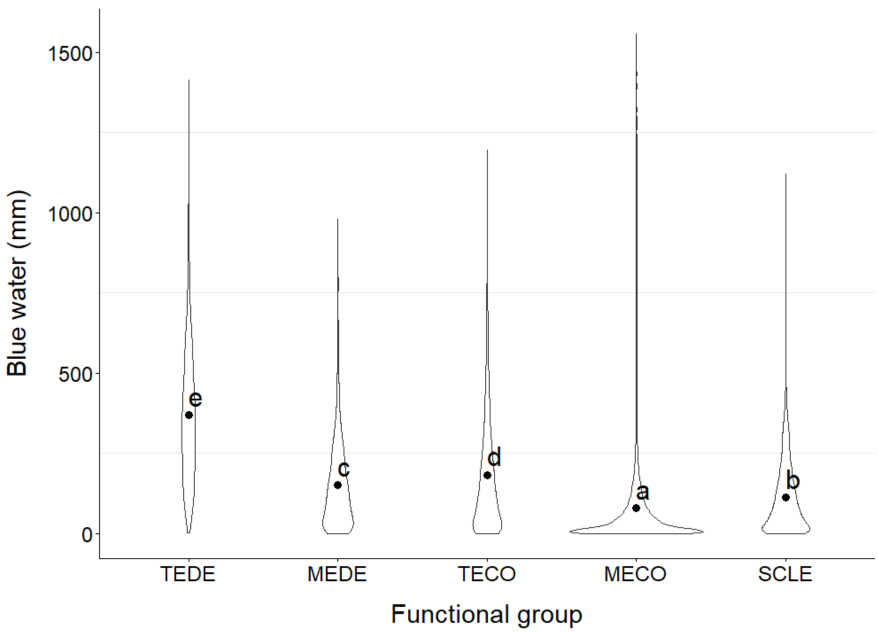

Figure S14. Blue water amount of the different functional groups of species. Different letters denote significant differences among forest types ($p < 0.01$) after Tukey's test. Abbreviations of the functional groups: TEDE, Temperate deciduous; MEDE, Mediterranean deciduous; TECO, Temperate coniferous; MECO, Mediterranean coniferous; SCLE, sclerophylls.

Q1.13. L154: I think further work should be presented here in order to justify the use of this model to simulate all the forest typologies presented in Spain, since nothing is said about calibration and validation ot the model in the manuscript. I understand the model has been properly calibrated and validation for some of the forests but not all the studied ones within the manuscript. For those, you should clearly justify how your model estimates can be considered close to reality

R1.13. We can not validate in all Spain due to the lack of field data for them. But we included a paragraph in a new section (2.7) where we did a comparative between the results of MEDFATE with the blue water from the ministry and the evapotranspiration from GLEAM. The map comparatives and statistical correlations are showed in the Supplemental material (Figure S7 and Figure S8). Theses maps showed the same regional patterns for these sources with MEDFATE. Although we know that a validation with field data it is better at country scale it is very harsh.

"2.7 Model evaluation

MEDFATE predictions have already been evaluated at the forest stand scale in terms of soil moisture dynamics, plant transpiration and water status in Mediterranean forest (De Cáceres et al., 2021, 2015). Given the focus of the present work, we evaluated regional-scale patterns of green and blue water predicted by MEDFATE against those produced by alternative methods. First, we did a comparison between the results for average blue water of MEDFATE with the average blue water from the Precipitation Runoff Integrated Model (SIMPAL acronym in Spanish) of the Spanish government. SIMPAL model is calibrated with stations that measure stream flows across Spain and it is interpolate a 1 km^2 resolution for the whole country (Estrela et al., 2012). Second, we compared green water patterns predicted by MEDFATE against those of GLEAM, which is derived from satellite data and covers the world with a resolution of 0.25 degrees (Martens et al., 2017). We used the GLEAM v3.7a that defines the evapotranspiration as the sum of transpiration (from short and tall vegetation), interception (from tall vegetation), soil and open water evaporation and snow sublimation. The comparative indicated that the blue water of the SIMPAL model and the green water of GLEAM followed the same regional patterns that MEDFATE (see Figure S7 and S8). SIMPAL reaches higher values of blue water in the Atlantic biome and lower values in west Temperate Mediterranean than MEDFATE. In opposition GLEAM evapotranspiration is higher in the west Temperate Mediterranean and lower in the Atlantic forest than MEDFATE. MEDFATE models at stand scale whereas SIMPAL and GLEAM models are at regional scale. At stand scale the evapotranspiration is high in the Atlantic forest and then the blue water is lower.

West Temperate Mediterranean is characterized by forest of low basal area and LAI (Table S2, Figure S4). Therefore the evapotranspiration is lower and blue water higher than surrounding forest. In other words, the differences observed seem to arise from the fact that MEDFATE takes into account the stand structural characteristics, which are difficult to represent in models based on interpolation (SIMPAL) or remote sensing (GLEAM).”

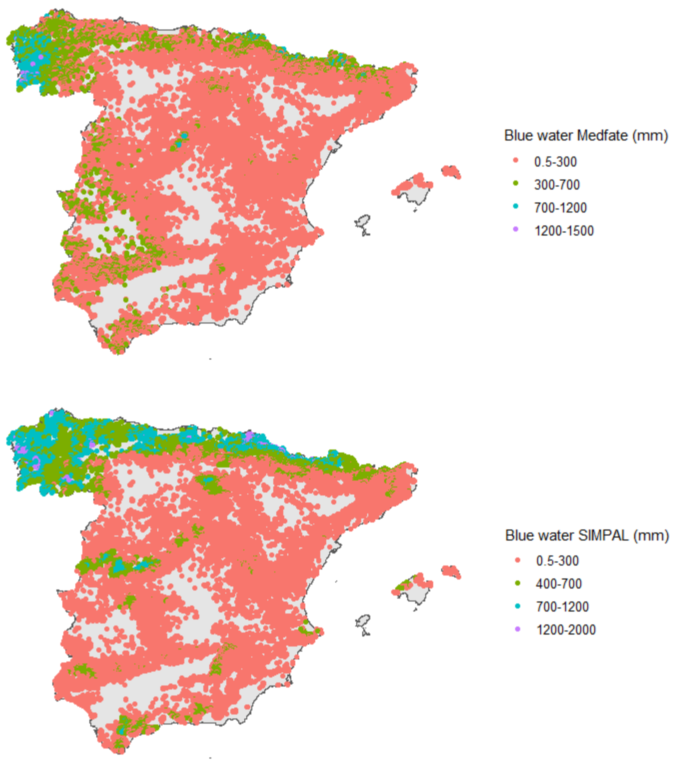

Figure S7. Average blue water maps for SFI3 plots according to MEDFATE and SIMPAL (Spearman correlation=0.70, R^2=0.354).

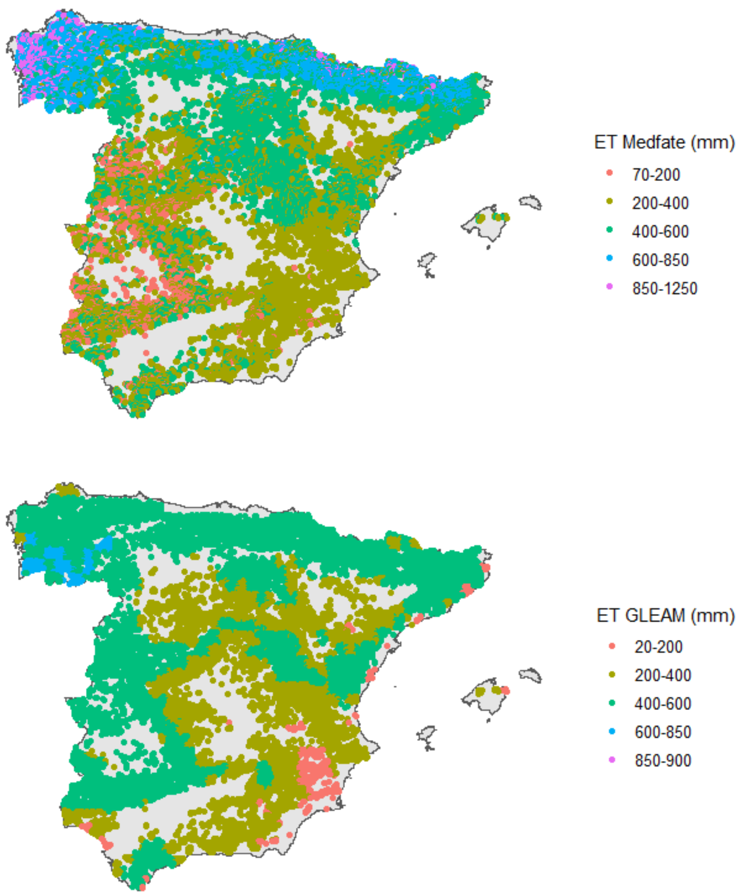

Figure S8. Average evapotranspiration (ET) maps for SFI3 plots according to MEDFATE and GLEAM (Spearman correlation=0.62, R^2=0.218)

Q1.14. L186: how did you incorporate these physiological parameters within your study? How did you calculate stand transpiration based on parameters which are depending on tree species?

R1.14. We included more information about how stand transpiration was estimated: "The estimation of maximum transpiration for the entire stand (Emax,stand), excluding considerations for soil water deficit, relies on the daily Penman's potential evapotranspiration (PET) and an empirical relationship established by Granier et al. (1999). But MEDFATE modified the Granier equation with aTmax and bTmax: If it is assumed that the entire leaf area of a stand corresponds to a single cohort i, the equation is:

$$\frac{E_{\text{max,stand}(i)}}{\text{PET}} = a_{\text{Tmax}} \cdot \text{LAI} + b_{\text{Tmax}} \cdot \text{LAI}^2,$$

where aTmax and bTmax represent species-specific parameters for cohort i. Assuming reliable species-specific estimates are accessible for aTmax and bTmax,

the equation can be applied to calculate Emax,stand(i), which denotes the maximum stand transpiration when dominated by the species of cohort i. Once Emax,stand(i) is determined for each species in the stand, the portion of SWR absorbed by a particular cohort i is employed to estimate its maximum transpiration (Emax(i)) from Emax,stand(i) (Korol et al., 1995).”

Q1.15. L188: I believe you should explain how your model is extrapolated to the whole Spain mainland based on SFI3 data. see in the figure 1 you finally considered a square pixel as your spatial domain, but noting is said about this in your M and M section. I understand you run your model for each IF polygon and then you calculate each water cycle component at a daily scale. However, from this water balance estimate to whole territory, there are several steps you should accomplished and they are not clearly explained in your study. Please consider to improve the explanation when moving stand scale into landscape scale for obtaining blue water variables.

R1.15. We ran the MEDFATE model for every SFI3 plot. The results are at daily scale and we averaged the values (365 days x 10 years) for the analyses. In the maps we showed theses averages values after we rasterized the SFI3 plots. In the rasterization we transformed a vectorial point database (the SFI3 plots) in a raster pixel of 1 km^2. We did not an extrapolation since every pixel have the same value of the SFI3 plot where it was transformed. The SFI is distributed with a systematic survey along the forested areas with a density of ~1 plot/km^2. Then we don’t need extrapolate the data since it cover the main forest areas of Spain. The pixel without a SFI3 plot don’t have values (it is the reason because the map do not cover the whole country since many pixels are “NA” where there are not forest areas). The rasterization is only a visual way of show the same values of the vectorial SFI3 plot database. In the Figure S7 (referenced in R1.13) we show the results of the blue water comparison between MEDFATE with SIMPAL model. In this case we did not transform the vectorial data in a raster pixel. But the data are the same that in the map of Figure 1A. Therefore, the rasterization it was only an aesthetic decision. However we included the information about the rasterization in the Figure 1: “Maps were realized with a rasterization of the SFI3 plots results at 1 km^2.”

Q1.16. L208: include topography as well

R1.16. Done.

Q1.17. L234: I would consider to modify this title to be clearly different than the next one. Something like this "Spatial patterns of blue water of Spanish biomes" could be applicable

R1.17. Thanks for your suggestion. We have changed the title to: “Spatial patterns of blue water in the Spanish biomes”

Q1.18. L238: Why did your consider this value as your threshold? It is based on a reference, on your data distribution? Please clarify this

R1.18. It was a descriptive value according to the maps. We changed the sentence to: "had values between 300-500 mm"

Q1.19. L241: I would suggest to interpret your blue water estimates based on mm and % of precipitation in two separated parts. These metrics are indicating different aspects, and thus I think they should be individually treated

R1.19. We included more explanations about the blue water amount (mm) but the patterns are very similar with blue water percentages.

Q1.20. L251: what about absolute values? are they indicating the same or not? I think that including another table for water depth is very pertinent.

R1.20. We included the Tukey test for the absolute values in the Supplemental material (Figure S11 and Figure S14 showed in R1.12). The results are very similar to the relative blue water.

Q1.21. Table 1. further explanation about differences on estimates values are required. For example, which is the meaning of a value of -1.9 instead of 0.22? I would include more information about beta regression coefficients in the statistical analyses section

R1.21. We included the explanation in the Table 1 and Table 2: "A positive value indicates that an increment occurred in the biome with respect to the observed values in the intercept".

Q1.22. Table A2. you should include information regarding the three last columns

R1.22. The information about the indices was explained in the section 2.3. The indices do not have units.

Q1.23. Figure 2A: how was LAI simulated? This aspect should further explained, and especially, given the high importance of this parameter when explaining blue water differences among forest types

R1.23. We have included this explanation: "Leaf area index (LAI in $m^2 \cdot m^{-2}$) was calculated from the foliar biomass (in $kg \cdot m^{-2}$) by using a *specific leaf area* coefficient (SLA, in $m^2 \cdot kg^{-1}$) that is species-specific (LAI = foliar biomass * SLA)."

References:

Benyon, R. G., Nolan, R. H., Hawthorn, S. N. D., and Lane, P. N. J.: Stand-level variation in evapotranspiration in non-water-limited eucalypt forests, Journal of Hydrology, 551, 233–244, https://doi.org/10.1016/j.jhydrol.2017.06.002, 2017.

Campos, I., González-Piqueras, J., Carrara, A., Villodre, J., and Calera, A.: Estimation of total available water in the soil layer by integrating actual evapotranspiration data in a remote sensing-driven soil water balance, Journal of Hydrology, 534, 427–439, https://doi.org/10.1016/j.jhydrol.2016.01.023, 2016.

De Cáceres, M., Molowny-Horas, R., Cabon, A., Martínez-Vilalta, J., Mencuccini, M., García-Valdés, R., Nadal-Sala, D., Sabaté, S., Martin-StPaul, N., Morin, X., D'Adamo, F., Batllori, E., and Améztegui, A.: MEDFATE 2.9.3: a trait-enabled model to simulate

Mediterranean forest function and dynamics at regional scales, Geoscientific Model Development, 16, 3165–3201, https://doi.org/10.5194/gmd-16-3165-2023, 2023.

Granier, A., Biron, P., and Lemoine, D.: Water balance, transpiration and canopy conductance in two beech stands, Agricultural and Forest Meteorology, 100, 291–308, https://doi.org/10.1016/S0168-1923(99)00151-3, 2000.

Hoek van Dijke, A. J., Herold, M., Mallick, K., Benedict, I., Machwitz, M., Schlerf, M., Pranindita, A., Theeuwen, J. J. E., Bastin, J.-F., and Teuling, A. J.: Shifts in regional water availability due to global tree restoration, Nat. Geosci., 15, 363–368, https://doi.org/10.1038/s41561-022-00935-0, 2022.

Llorens, P., Latron, J., Álvarez-Cobelas, M., Martínez-Vilalta, J., and Moreno, G.: Hydrology and Biogeochemistry of Mediterranean Forests, in: Forest Hydrology and Biogeochemistry: Synthesis of Past Research and Future Directions, edited by: Levia, D. F., Carlyle-Moses, D., and Tanaka, T., Springer Netherlands, Dordrecht, 301–319, https://doi.org/10.1007/978-94-007-1363-5_14, 2011.

Mastrotheodoros, T., Pappas, C., Molnar, P., Burlando, P., Manoli, G., Parajka, J., Rigon, R., Szeles, B., Bottazzi, M., Hadjidoukas, P., and Fatichi, S.: More green and less blue water in the Alps during warmer summers, Nat. Clim. Chang., 10, 155–161, https://doi.org/10.1038/s41558-019-0676-5, 2020.

Qubaja, R., Amer, M., Tatarinov, F., Rotenberg, E., Preisler, Y., Sprintsin, M., and Yakir, D.: Partitioning evapotranspiration and its long-term evolution in a dry pine forest using measurement-based estimates of soil evaporation, Agricultural and Forest Meteorology, 281, 107831, https://doi.org/10.1016/j.agrformet.2019.107831, 2020.

Simonin, K., Kolb, T. E., Montes-Helu, M., and Koch, G. W.: The influence of thinning on components of stand water balance in a ponderosa pine forest stand during and after extreme drought, Agricultural and Forest Meteorology, 143, 266–276, https://doi.org/10.1016/j.agrformet.2007.01.003, 2007.