S1 Additional figures

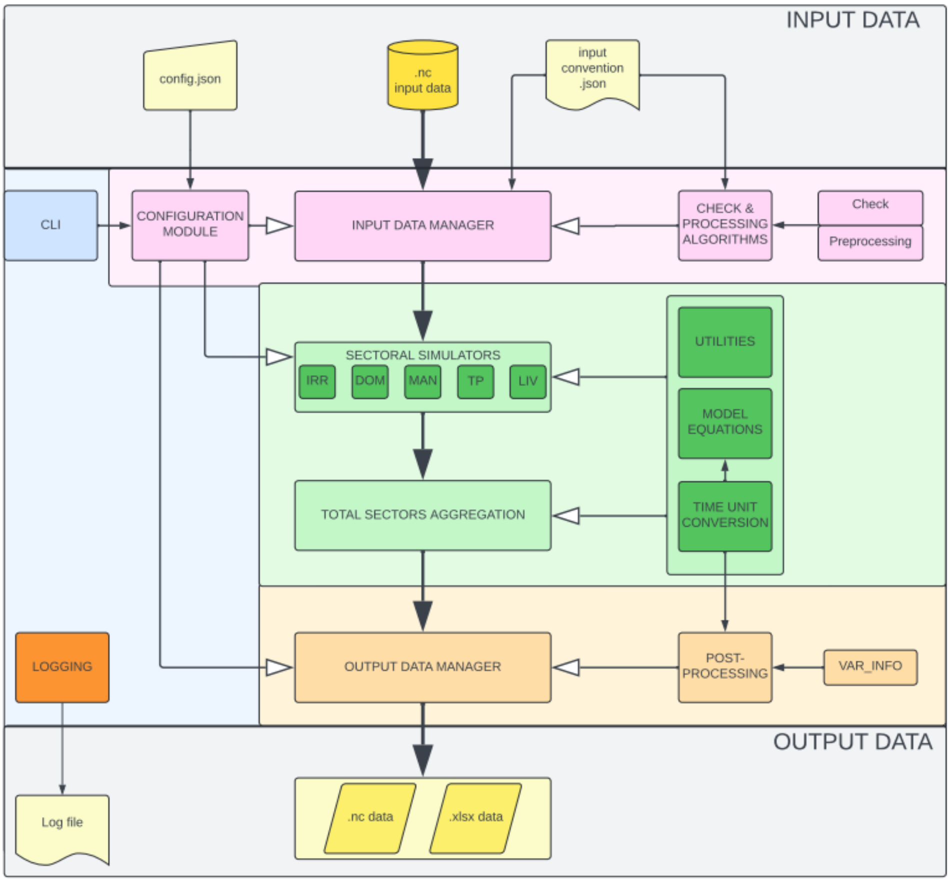


Figure S1: Model-View–Controller architectural pattern of the reprogrammed GWSWUSE software. The Controller package (pink) manages the configuration and input data (e.g., sectoral water use data), the Model package (green) contains core hydrological processes, and the View package (light orange) handles the saving and presentation of model outputs in NetCDF format.

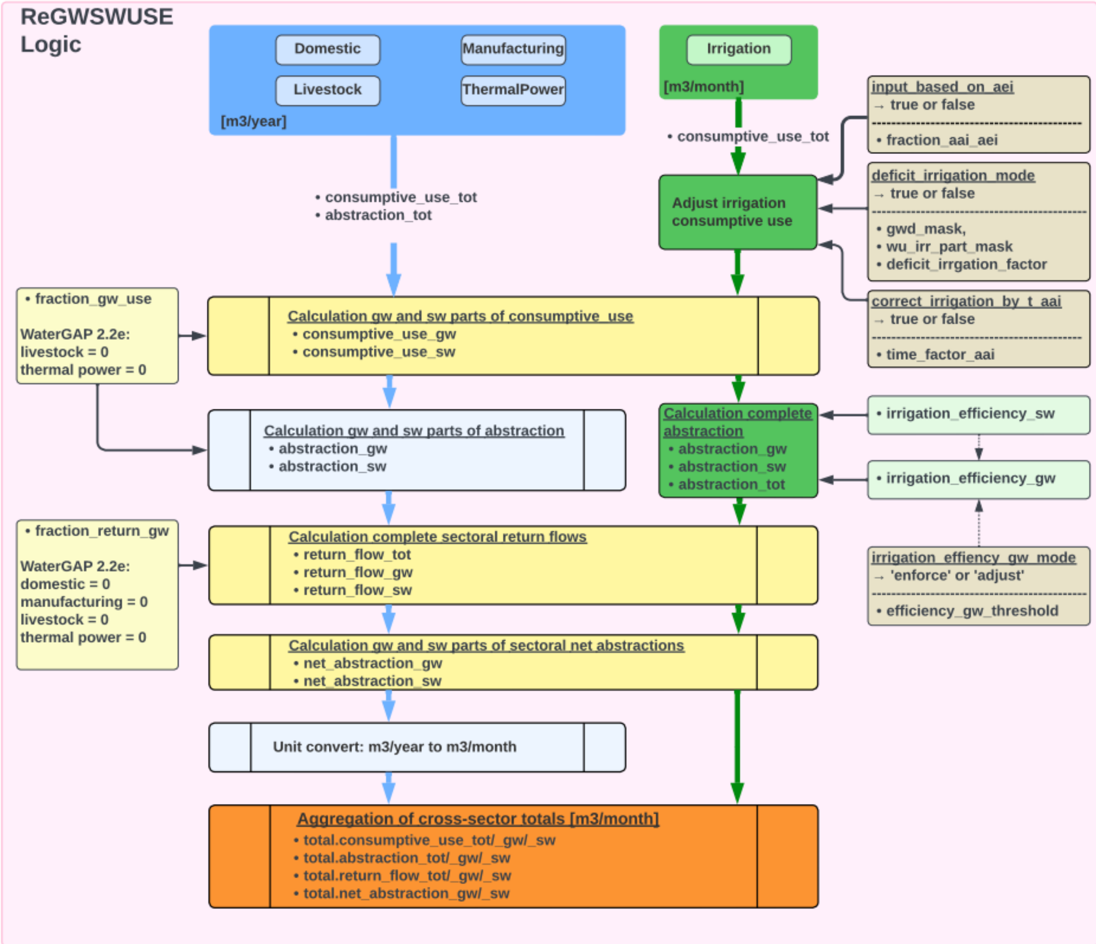


Figure S2: The reprogrammed GWSWUSE framework starting from sectoral water use inputs to aggregated potential net abstractions. Please refer to the external documentation (https://hydrologyfrankfurt.github.io/ReWaterGAP/model_processes/gwswuse/index.html#gwswuse) for a detailed explanation of this framework.

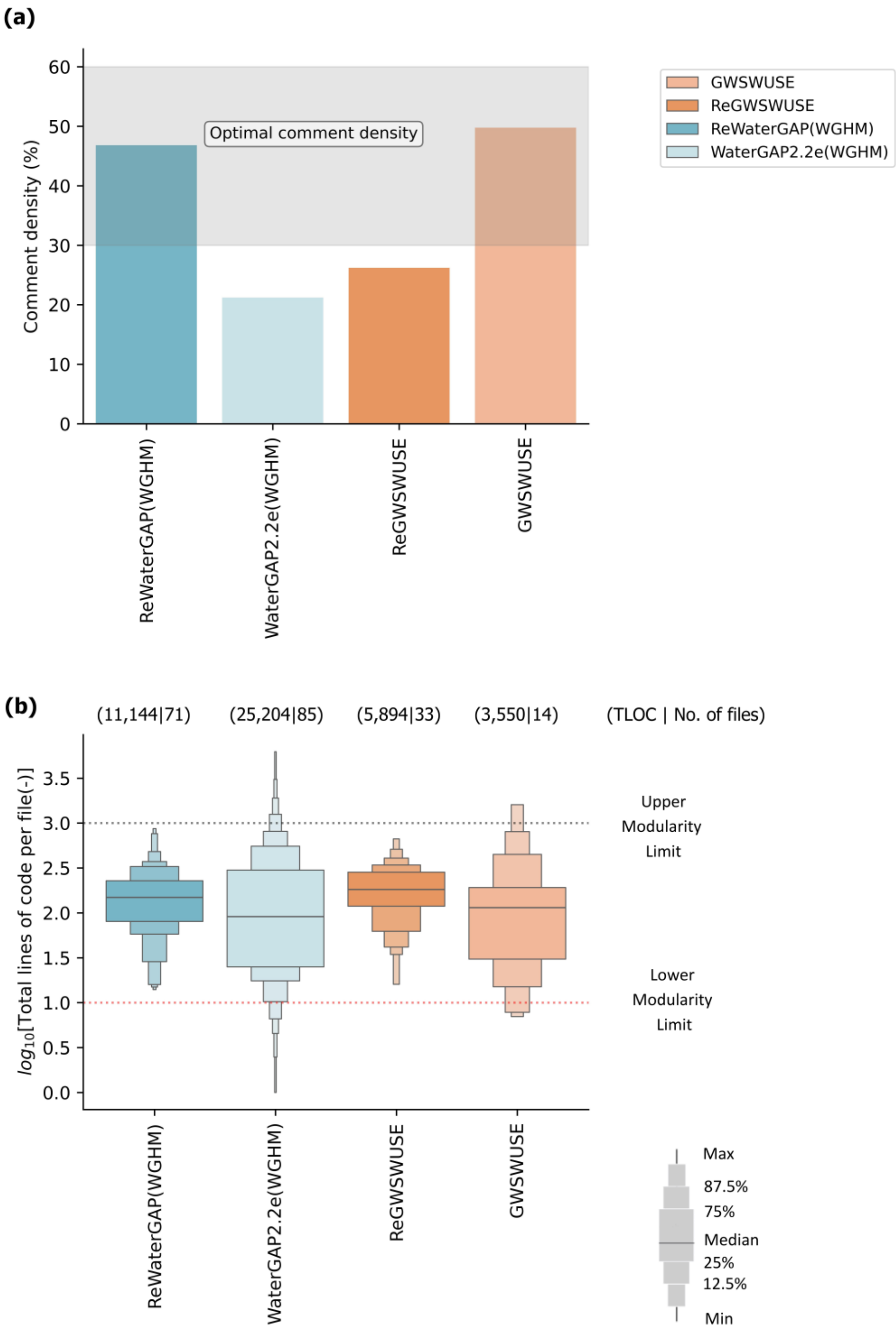


Figure S3: Modularity and commenting practice of two legacy and reprogrammed software. (a) Comment density per model. The grey zone in Fig. S3a denotes the optimal comment density, (b) Letter value plot of the total lines of code per file (logarithmic scale) of each model. The dotted black (red) line shows the upper (lower) modularity limit defined as the maximum of 1,000 (minimum of 10) total lines of code per file. The values (x | y) shown in the upper section of Fig. S3b correspond to the TLOC and number of files per model.

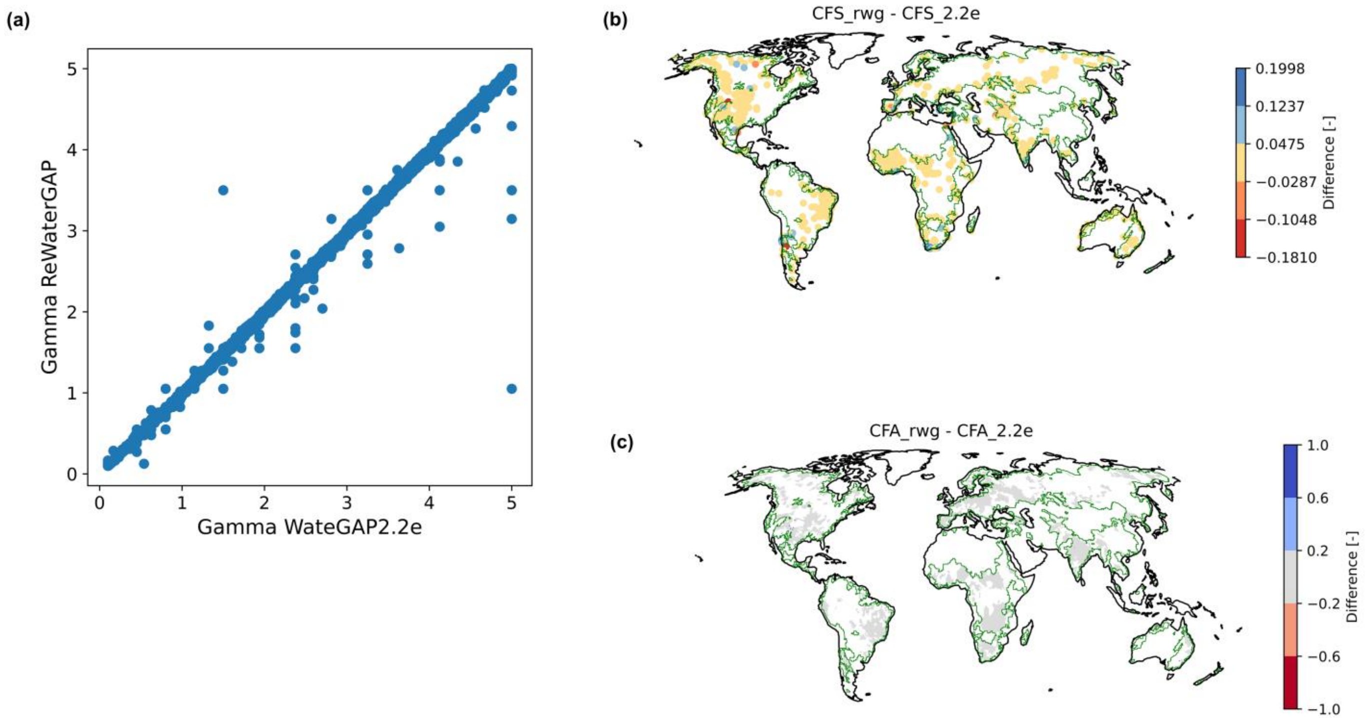


Figure S4: Difference in calibration parameters between the reprogrammed and legacy WGHM (Müller Schmied et al., 2024). (a) Gamma (runoff coefficient), (b) absolute change in areal correction factor (CFA) and (c) absolute change in station correction factor (CFS). The CFA (with range 0.5-1.5) adjusts runoff and actual evapotranspiration at the grid-cell level to maintain mass balance, while the CFS (unconstrained) corrects streamflow at gauging stations to prevent error propagation downstream (Müller Schmied et al., 2024). (b)-(c) Green outlines indicate the boundaries of the calibration basins. Outside these boundaries, Gamma is regionalized (Müller Schmied et al., 2021). (b)-(c) White represents cases where the difference is 0, except for Greenland.

S2 Users' perceptions of the reprogrammed WaterGAP software.

We conducted an online survey to determine perspectives on the reprogrammed WGHM software, focusing on research software sustainability (see questionnaire and associated response in supplement). The survey evaluated the readability, comprehensibility, modifiability, and documentation quality of a code snippet that implements the Priestley-Taylor potential evapotranspiration algorithm in the reprogrammed software. The survey was conducted within approximately one month, receiving 217 clicks, with 64 participants completing it.

The 64 survey participants who completed the survey represented a diverse group, with the majority being PhD students (24 out of 64), scientific staff (13 out of 64), or Postdocs (9 out of 64) (Fig. S5a). On average, participants had approximately 14 years of programming experience, with individual experience ranging from 1 to 50 years (Fig. S5b).

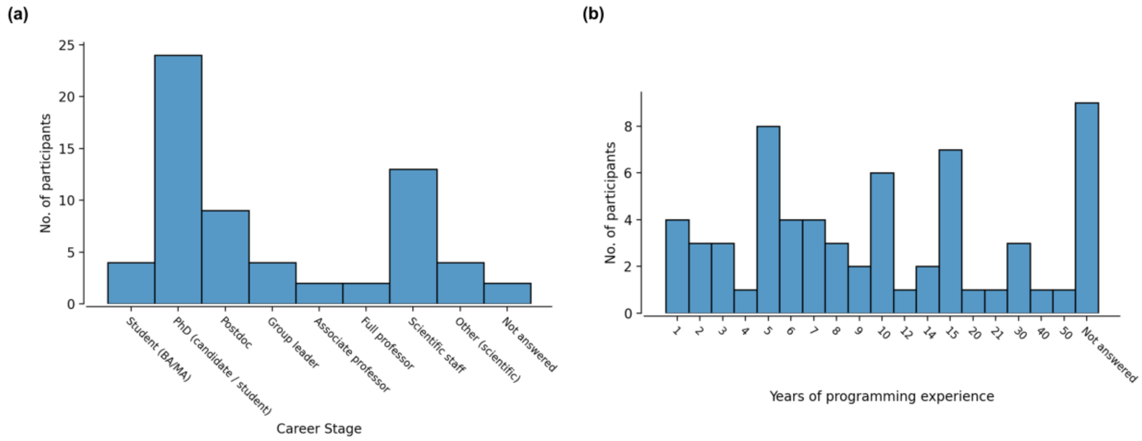
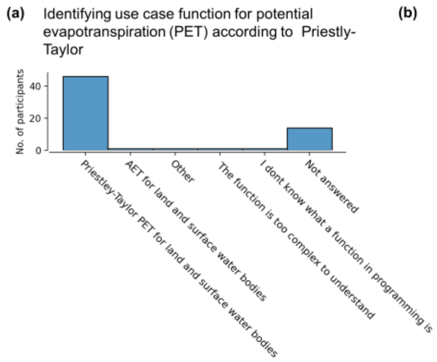


Figure S5: Career stage (a) and years of programming experience (b) for the 64 survey participants.

The survey results demonstrate a high level of code readability, with 46 out of 64 participants correctly identifying the Priestley-Taylor algorithm (Fig. S6a). The survey also examined the ease of code modification by testing participants' confidence in implementing a change to the algorithm. When asked about modifying the code with a new atmospheric constant, 40 out of 64 participants expressed some level of confidence (ranging from slightly to very confident) in their ability to do so (Fig. S6b). In contrast, 17 out of 64 participants did not respond, 3 out of 64 participants indicated having no confidence, and 4 out of 64 participants stated they had no idea how to proceed (Fig. S6b). Additionally, 42 out of 64 participants correctly identified the specific line of code that would require modification (Fig. S6c). Meanwhile, 17 out of 64 participants did not respond, and 5 out of 64 participants selected the wrong line of code.



(b) How confident are you to change the atmospheric constant in the use case code from 101.3kPa to 101.325kpa ?

Category	No. of participants
Not confident	3
Slightly confident	1
Moderately confident	2
Confident	6
Very confident	31
No Idea	4
Not answered	17

(c) Where in the code would you changed the atmospheric constant ?

Category	No. of participants
Correctly Identified the line number in the code	42
Wrongly Identified the line number in the code	5
Did not answer	17

Figure S6: Survey results for the 64 participants on code readability, comprehension, and ease of modification of the Priestley-Taylor potential evapotranspiration (PET) code snippet.

The survey results also showed a high level of documentation readability and comprehension. 46 out of 64 participants agreed (with 32 out of 64 strongly agreeing) that the provided external documentation clearly explained the code (Fig. S7). Furthermore, 37 out of 64 participants understood the algorithm's purpose after reading the external documentation, and 41 out of 64 confirmed that the documentation was not difficult to comprehend (Fig. S7).

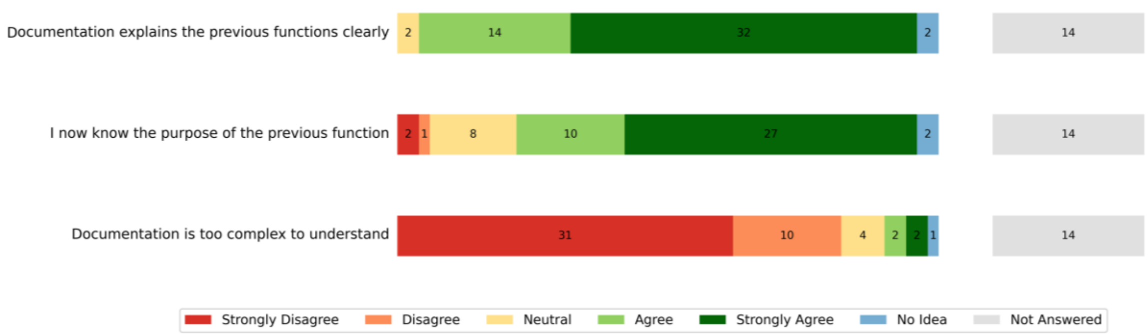


Figure S7: Survey results from 64 participants regarding external documentation readability and comprehension of the Priestley-Taylor potential evapotranspiration (PET) code snippet. The values shown in Figure S7 are the number of participants.

Our user survey has several limitations and potential biases. The survey was distributed to participants at the European Geosciences Union (EGU) 2024 conference, which introduced self-selection bias. Respondents were likely more interested in software sustainability topics, potentially skewing the results toward a more engaged subset of researchers. More importantly, we assessed understanding and perceptions based on a code snippet rather than the full source code, which does not provide a comprehensive evaluation of the software. Due to time constraints, we did not conduct a practical evaluation of reproducibility, guiding participants through executing the reprogrammed software with a tutorial.

References

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geoscientific Model Development, 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., Kynast, E., Peiris, T. A., Schiebener, L., Schumacher, M., and Döll, P.: The global water resources and use model WaterGAP v2.2e: description and evaluation of modifications and new features, Geoscientific Model Development, 17, 8817–8852, https://doi.org/10.5194/gmd-17-8817-2024, 2024.