Discussion on the regression results

This analysis aimed to identify the most relevant dataset and significant explanatory variables influencing industrial water withdrawals to develop an accurate final model. The dataset was segmented based on circular water usage into two distinct subsets: low-circular (Kreis with predominantly singular, water usage) and high-circular (Kreis employing extensive water recycling practices). Additionally, a full dataset (IWW) was analyzed to provide a comprehensive baseline comparison.

The low-circular dataset demonstrated the strongest statistical significance and highest explanatory power, evidenced by a substantial R² value (0.557) and highly significant relationships between economic activity, represented by gross value added (log\_gva), and water withdrawals. The intercept and log\_gva coefficients were both statistically significant, strongly suggesting that industries with limited water recycling directly link their water usage to economic productivity. Conversely, the high-circular dataset showed a considerably weaker relationship (R² = 0.046) between economic activity and water withdrawals, and the intercept was not statistically significant. The full dataset (IWW) exhibited moderate explanatory power (R² = 0.175), indicating that aggregating data from industries with varying circular water use practices obscured the strength of the relationships.

Moreover, analyses of climatic variables (summer\_TXK, summer\_TMK, hot\_days25, hot\_days30, mean\_temp) were initially performed to evaluate their potential influence on industrial cooling water withdrawals. Surprisingly, no statistically significant correlations were identified between these climatic variables and cooling water demand across multiple model configurations (Cooling1 to Cooling5). This outcome suggests that, within the studied region and industrial context, cooling water withdrawals may be more driven by production-related economic factors rather than climatic variations. Potential explanations include limited variation in climatic conditions during the analysis period, relatively consistent industrial cooling requirements independent of short-term weather variations, or the resolution and specificity of available climatic data.

Considering these findings, the most scientifically sound and data-supported approach for the final model would involve primarily utilizing the low-circular dataset, with log\_gva as the key explanatory variable. Incorporating climatic variables is not recommended for the main model given their lack of significance and low explanatory power observed in the analyses. Future research could explore additional climatic parameters or a broader spatial-temporal scope to reassess potential climatic influences on industrial water withdrawals.

Methodology steps for the splitting of the dataset

The dataset was classified into two subsets based on the degree of circular water usage at the district (Kreis) level. First, the total amount of single-use water and total water withdrawals were calculated for each district. Subsequently, the ratio of single-use water to total water usage (singular ratio) was determined for each district. Districts in the top quartile (highest 25%) of singular water use ratios were classified as **low circular (LOW\_C)**, indicating lower recycling or recirculation rates and higher reliance on single-use water practices. Conversely, districts below this threshold were classified as **high circular (HIGH\_C)**, reflecting greater adoption of water recycling. This classification was then merged back into the main dataset, resulting in two subsets (**low circular** and **high circular**) for subsequent analyses.

Methodology for the base year = 2004

To analyze changes in gross value added (GVA) within each Kreis over time, a baseline-centered approach was adopted by designating 2004 as the reference year. Specifically, for each Kreis, we extracted the natural logarithm of its GVA in 2004, denoted log(GVAi,2004), and subtracted this value from the logarithm of GVA in all other years. Formally, the standardized log GVA is given by:

log(GVAi,t)std  =  log(GVAi,t)  −  log(GVAi,2004)

where t indexes the time period. This transformation centers each kreis’s GVA trajectory on its own 2004 level, thereby controlling for Kreis-specific initial conditions and ensuring that subsequent analyses focus on relative changes from this baseline. As a result, estimated coefficients in our panel regressions can be interpreted in terms of departures from the 2004 reference point, enhancing the clarity and comparability of the findings.