The manuscript presents the results of modeling bioclimatic conditions in Freiburg. The authors compare the results obtained using numerical models with the results of models using machine learning techniques. In order to determine the spatiotemporal variability of the UTCI index, four ML models emulating appropriate numerical models are used. It was shown that properly trained ML models give results comparable to those of numerical models at much lower computational costs. The approach of using several independent numerical models with different resolutions to obtain the variables necessary to determine UTCI and then emulating them with ML models, so that calculations of long-term variability can be performed in a reasonable time, may constitute a contribution to the development of bioclimatic research methodology in urban areas. From a purely bioclimatic point of view, an important result are high-resolution maps of exceeding specific UTCI thresholds determined on the basis of modeling results with a 1-hour time step for a four-year observation period. The manuscript is fully in line with the scope of the GMD journal, its layout is typical of scientific articles, the argument is logical and contains all the necessary information. As I am not a native speaker, I do not evaluate the correctness of the language, but I did not notice any incorrectness. For the above reasons, I believe that it can be published as is, with only very minor corrections.

Specific comments

Ln. 61 The authors state that "*Nevertheless, the emulated ML model can never exceed the accuracy of the numerical model because it is trained based on the model's results*", but in the results and discussion it turns out that the proposed ML models often give better results than numerical models. I propose to explain this contradiction.

Ln. 72 … *four cardinal wind directions* … – I am used to analyzing the wind field as three-dimensional. Did I misunderstand something?

Ln. 213 "*The error distributions of SUEWS and the MLP across the different stations are similar (Fig. 3a)…*" – I think that Fig. 3a shows the error for all stations rather than the error distributions across stations.

Ln. 227 "*The T_{mrt} U-Net has a slightly lower accuracy than SOLWEIG (RMSE of 6.18 K to 5.86 K; R2 of 0.84 to 0.86)*" (and also in ln. 239) – the acceptable level of accuracy is usually a subjective choice. However, for T_{mrt} in the standard ISO7726 (ISO, 1998), ISO recommends that the error in T_{mrt} estimates should be within ±5°C. Could you please address/discuss this.

Fig. 4 The RMSE of SUEWS predictions (orange lines) are almost invisible - could they be bolded? Please change "Dez" to "Dec".

Fig. 6 Some of the sharp drops in T_{mrt} and UTCI in the SOLWEIG charts (e.g. afternoon 2022-02-11 at the station marked "e") are likely the results of shading, which are directly calculated by SOLWEIG at ground level, while the reference data is from 3.5 m. Similarly at early morning or afternoon at other stations. Am I right? Anyway, could you comment on

these rapid, sawtooth changes in the SOLWEIG charts and their effect on the accuracy statistics.