https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

Open Access Earth System Science Data Discussions

The Application of Machine Learning Algorithms to the Global Forecast of Temperature-Humidity Index with High Temporal Resolution

Pantelis Georgiades[1,2], Theo Economou[2,4], Yiannis Proestos[2], Jose Araya[2], Jos Lelieveld[2,3], and Marco Neira[2]

[1]Computation-based Science and Technology Research Center CaSToRC, The Cyprus Institute, Nicosia, Cyprus
[2]Climate and Athmosphere Research Centre (CARE-C), The Cyprus Institute, Nicosia, Cyprus
[3]Max Planck Institute for Chemistry, Mainz, Germany
[4]Department of Mathematics and Statistics, University of Exeter, Exeter, UK

Correspondence: Pantelis Georgiades (p.georgiades@cyi.ac.cy) and Marco Neira (m.neira@cyi.ac.cy)

Abstract.

Climate change poses a significant threat to agriculture, with potential impacts on food security, economic stability, and human livelihoods. Dairy cattle, a crucial component of the livestock sector, are particularly vulnerable to heat stress, which can adversely affect milk production, immune function, feed intake, and in extreme cases, lead to mortality. The Temperature Humidity Index (THI) is a widely used metric to quantify the combined effects of temperature and humidity on cattle. However, most studies estimate THI using daily-level data, which fails to capture the full extent of daily thermal load and cumulative heat stress, especially during nights when cooling is inadequate. To address this limitation, we developed a machine learning approach to temporally downscale daily climate data to hourly THI values. Utilizing historical ERA5 reanalysis data, we trained an XGBoost model and generated hourly THI datasets for 12 NEX-GDDP-CMIP6 climate models under two emission scenarios (SSP2-4.5 and SSP5-8.5) extending to the end of the century. This high-resolution THI data provides a more accurate assessment of heat stress in dairy cattle, enabling better predictions and management strategies to mitigate the impacts of climate change on this vital agricultural sector.

1 Introduction

Climate change, driven by anthropogenic greenhouse gas emissions, is a multifaceted challenge with profound implications for ecosystems and human societies alike (IPCC, 2023; Malhi et al., 2020). The agricultural sector, which has been the cornerstone of global food security and economic activities for the past centuries, is particularly vulnerable to climate change and variability (Abbass et al., 2022). Within this sector, livestock farming emerges as a critical area of concern due to its susceptibility to environmental stressors, making the assessment and management of climate impacts critical for sustaining agricultural productivity and livelihoods (Cheng et al., 2022; Escarcha et al., 2018).

Dairy farming, an integral component of the livestock industry, is particularly sensitive to climatic conditions. Economic losses due to heat stress in the US alone are estimated at $1.5 - $1.7 billion per year, accounting for approximately 63.9% of

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

the national yearly losses of this economic sector (North et al., 2023; St-Pierre et al., 2003; Cartwright et al., 2023). Predictive models for the US forecast monetary losses as high as $2.2 billion by the end of the century (Mauger et al., 2014) .

The effects of heat stress on cattle are determined by complex interactions between environmental factors (particularly temperature and humidity) and biological parameters. Modern-day breeds of dairy cattle are the result of intensive genetic selection, aimed primarily at increasing milk productivity. However, this increased productivity is genetically linked to physiological traits such as greater metabolic rates and increased feed intake, both of which augment endogenous heat generation in the animals, thereby making high-productivity breeds particularly susceptible to heat stress (Kadzere et al., 2002; Moore et al., 2023).

Dairy cows depend on evaporative heat loss as their main thermoregulatory mechanism (Zhou et al., 2023). Therefore, when exposed to increased temperatures, they rely heavily on their ability to dissipate heat by either sweating or panting in order to regulate their body temperature. Water evaporation rates are negatively correlated with the relative humidity of the surrounding environment, so a cow's ability to regulate its body temperature is progressively diminished with increasing moisture in the air (Bohmanova et al., 2007). As a consequence, even moderate increases in temperature can have severe biological repercussions under high-humidity conditions. Heat stress has been linked to a plethora of deleterious effects in dairy cattle, including reductions in milk yield and quality, decreased reproductive success, decreased feed intake, body-weight loss, reduced immune function, altered behaviour and, in extreme cases, mortality (Burhans et al., 2022; Cartwright et al., 2023; Kadzere et al., 2002; Polsky and von Keyserlingk, 2017).

The Temperature Humidity Index (THI) is a relatively simple, non-invasive metric developed to quantify the levels of thermal stress caused by the combined effects of temperature and humidity on cattle. Its calculation requires meteorological data that is generally easy to access (i.e. air temperature and relative humidity), and its correlation with physiological parameters has been validated by a large body of literature, (Bohmanova et al., 2007; Ravagnolo et al., 2000; Bouraoui et al., 2002; Brügemann et al., 2012; Igono et al., 1992; Bernabucci et al., 2014). This index has served as reference for early warning systems, such as the Livestock Weather Safety Index (North et al., 2023) with specific THI thresholds identified for mild, moderate and severe heat stress (Baca et al., 2019; North et al., 2023).

In most of the available scientific literature, THI values are estimated using daily-level data (e.g. daily averages or daily extremes in temperature and humidity, etc.). The reason for this is twofold: On one hand, working at finer temporal resolutions (e.g. hourly) generally requires the processing and storage of very large datasets, which can pose logistic and computational difficulties. On the other hand, data provided by traditional climate models is also available at daily or coarser temporal resolutions. Unfortunately, daily-level calculations can neither accurately estimate the daily thermal load caused by fluctuating climatic conditions across each day (e.g. diurnal vs nocturnal temperatures), nor capture cumulative effects over consecutive days, particularly during periods when night-time conditions do not allow for efficient heat dissipation (St-Pierre et al., 2003; Hahn, 1997). This situation underscores the need for finer-scale climate projections that can more accurately reflect the environmental stressors impacting dairy cattle, thereby allowing for better forecasts of the expected impacts of climate change on this key economic sector.

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024


Open Access Earth System Science Data Discussions

In the last few decades there have been enormous advancements in computational capabilities and machine learning algorithms have paved the way for significant improvements in the spatial and temporal resolution of climate data (Huntingford et al., 2019). Such innovations allow for the spatial and temporal downscaling of global climate models to generate high-resolution projections that are more aligned with the practical needs of agricultural planning and management. Our study aims to bridge the gap between coarse-resolution climate projections and the required fine-scale environmental data necessary for effective farm management under changing climatic conditions.

2 Methodology

We utilized a well-established machine learning algorithm, specifically the 'Extreme Gradient Boost' (XGBoost) model, to temporally downscale daily climate projections to hourly THI values. We opted for the XGBoost model for its computational efficiency compared to Random Forest and other analogous algorithms, specifically for our use case. Additionally, the implementation of Random Forest in Python does not support incremental learning, which was crucial for this study due to the vast amount of data the model needed to process during training. Furthermore, the model was trained on CPU rather than GPU due to memory limitations of our available GPUs, necessitated by the very large dataset.

Our approach involves the training of the model using the ERA5 reanalysis dataset, which contains historical hourly data (Hersbach et al., 2020). The model was subsequently applied to generate hourly THI projections until the end of the century, based on bias-adjusted climate projections from the NASA NEX-GDDP-CMIP6 datasets (Thrasher et al., 2022). We developed data using twelve climate models and concentrated on two distinct Shared Socioeconomic Pathways (SSPs): SSP2-4.5 and SSP5-8.5, which represent moderate and high greenhouse gas emissions scenarios, respectively, aiming to encompass a broad range of potential climatic outcomes.

In the following sub-sections, we present the details of our methodology, including data sources, preprocessing procedures, feature selection, model training, and evaluation procedures.

2.1 Data

Two distinct sources for climate data were used in this study:

2.1.1 ERA5 reanalysis

The ERA5 reanalysis data, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the Copernicus Climate Change Service, combine historical observations into global estimates using forecasting models (Hersbach et al., 2020). This data set is provided at a spatial resolution of 0.25° and hourly temporal resolution (atmosphere component). For the purposes of this study, we retrieved the variables *t2m* (temperature at 2 m) and *d2m* (dewpoint temperature at 2 m), for a time period spanning from 1980 to 2020, from the "ERA5 hourly data on single levels from 1940 to present" entry available in the Copernicus Data Store (CDS), using the Python API.

We estimated the relative humidity variable using the Magnus formula (WMO, 2021), as follows:

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

Open Access Earth System Science Data Discussions

$$e(T_d) = 6.1078 \cdot \exp\left(\frac{17.1 \cdot T_d}{235 + T_d}\right) [hPa] \tag{1}$$

$$e_s(T) = 6.1078 \cdot \exp\left(\frac{17.1 \cdot T}{235 + T}\right) [hPa]. \tag{2}$$

where T and T_d are the ambient and dew point temperature in degrees Celsius, respectively. $e(T_d)$ is the vapor pressure at temperature T_d and $e_s(T)$ the saturation vapor pressure at temperature T. Finally, the relative humidity can be calculated by taking the ratio of the two, as follows:

$$RH = 100 \cdot \frac{e}{e_s} [\%] \tag{3}$$

2.1.2 NEX-GDDP-CMIP6

The NEX-GDDP-CMIP6 ensemble dataset comprises global downscaled climate change scenarios. These were derived from the General Circulation Model (GCM) runs conducted under the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Thrasher et al., 2022). It includes global downscaled and bias adjusted projections from ScenarioMIP model runs and feature a 0.25° spatial resolution and daily temporal resolution. The data for twelve climate models and two greenhouse gas emissions scenarios (SSP2-4.5 and SSP5-8.5) were retrieved in netCDF format from the NCCS THREDDS data service. From these datasets, we utilized the daily average, minimum and maximum temperatures, as well as the mean relative humidity variables. Table 1 presents the full list of NEX-GDDP-CMIP6 models used in this study to generate hourly THI projections until the end of the century.

Table 1. List of NEX-GDDP-CMIP6 models used in this study to generate hourly THI predictions.

No.	Model Name	No.	Model Name
1	ACCESS-ESM1-5	7	GFDL-ESM4
2	CMCC-CM2-SR5	8	INM-CM4-8
3	EC-Earth3	9	INM-CM5-0
4	EC-Earth3-Veg-LR	10	MIROC6
5	FGOALS-g3	11	MRI-ESM2-0
6	GFDL-CM4	12	NorESM2-MM

2.2 Feature selection

The computation of hourly THI values from the ERA5 dataset was performed using the following formula:

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

Open Access Earth System Science Data Discussions

$$THI = (1.8 \times T + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times T - 26) \quad (4)$$

where T denotes the temperature in degrees Celsius ($^{\circ}C$) and RH represents the relative humidity in percentage ($\%$) (Yeck, 1971).

To ensure compatibility with the variables available in the NEX-GDDP-CMIP6 datasets, we generated features from the hourly ERA5 dataset as follows:

- Daily minimum, maximum and average temperature.
- Daily average THI, calculated using the daily average temperature and average relative humidity.
- Daily average relative humidity

Lastly, we included the hour of the day and day of the year features to account for diurnal and seasonal variations of THI and the land-sea mask, ranging from 0 (sea) to 1 (land), to differentiate between terrestrial and maritime environments. Table 2 shows a complete list of the input variables used to generate the training and inference data from ERA5 and NEX-GDDP-CMIP6 data, respectively.

Table 2. List of the input variables used in this study.

Feature	Long Name	Feature	Long Name
THI_	Daily average THI	dayLength	Length of day (hours from sunrise to sunset)
t2m	Daily average temperature	rhmean	Daily average relative humidity
t2mmax	Daily maximum temperature	dayOfYear	Day of the year
t2mmin	Daily minimum temperature	hourOfDay	Hour of the day
lsm	Land Sea Mask		

2.3 Model Training

An XGBoost regressor model was employed to perform the temporal downscaling from daily to hourly resolution. Three models of increasing complexity were trained to explore the trade-off between model performance and computational efficiency. The parameters of each of these models is presented in Table 3.

Table 3. The parameters of the three XGBoost models trained to temporally downscale daily climate data to hourly THI values.

Model no.	Lambda regularisation	Max depth	Number of parallel trees	Learning rate
1	1	5	10	0.1
2	1	5	20	0.1
3	5	6	30	0.01

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

The selection of the predictive model was partially influenced by the necessity for an approach capable of incremental learning. This requirement was dictated by the sheer volume of the training dataset, which precluded the possibility of training on the entire dataset simultaneously, due to technical limitations. The *xgboost* library, implemented in Python, was chosen for its ability to accommodate this need as well as the well-established accuracy and speed compared to other ensemble learning models (Chen and Guestrin, 2016; Sheik et al., 2024). The framework facilitated the training of the model in monthly increments, commencing from the year 1980 and concluding in 2017. To ensure the continuity and assess the model's performance over time, checkpoints were stored at the end of each training increment (monthly). The first month of 2018 was used as a test set throughout the training procedures.

2.4 Model Evaluation

The performance of the trained models was assessed using ground truth data derived from the ERA5 dataset for the period spanning from February 2018 until December 2020, which wasn't seen by the model during training. This evaluation phase aimed to establish the models' predictive accuracy and their ability to generalize to unseen data. Model performance was quantitatively evaluated using standard statistical metrics, including the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2).

2.4.1 Implementation Details

The data manipulation and feature engineering were performed using Python 3.11 and the *xarray*, *numpy* and *pandas* libraries. The input variables were scaled to the 0-1 range using the *MinMaxScaler* method of the *scikit-learn* library. Lastly, the *xgboost* library was used to implement the training and inference procedures for temporally downscaling daily climatic variables to hourly THI values.

3 Results

3.1 Model Training and Evaluation

Consistent with the methodologies put forth in the Methods section, this study included the training of three XGBoost regression models, each varying in complexity, to establish the optimal parameterization for the prediction of hourly Temperature Humidity Index (THI) values from daily climatic inputs. Figure 1 illustrates the progression of two key performance indicators—Mean Absolute Error (MAE) and Mean Squared Error (MSE)—throughout the training phases. These metrics were computed upon the conclusion of each training epoch, corresponding to monthly intervals spanning from January 1980 to December 2017. The evaluation was conducted using a test dataset, which included global data from January 2018, to validate the models' predictive accuracy and generalization capability.

Furthermore, the right panel of Figure 1 presents a ground truth versus prediction plot for Model 1's inference on the validation set (February 2018 - December 2020). We employed a density plot instead of a scatter plot to facilitate visualization

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

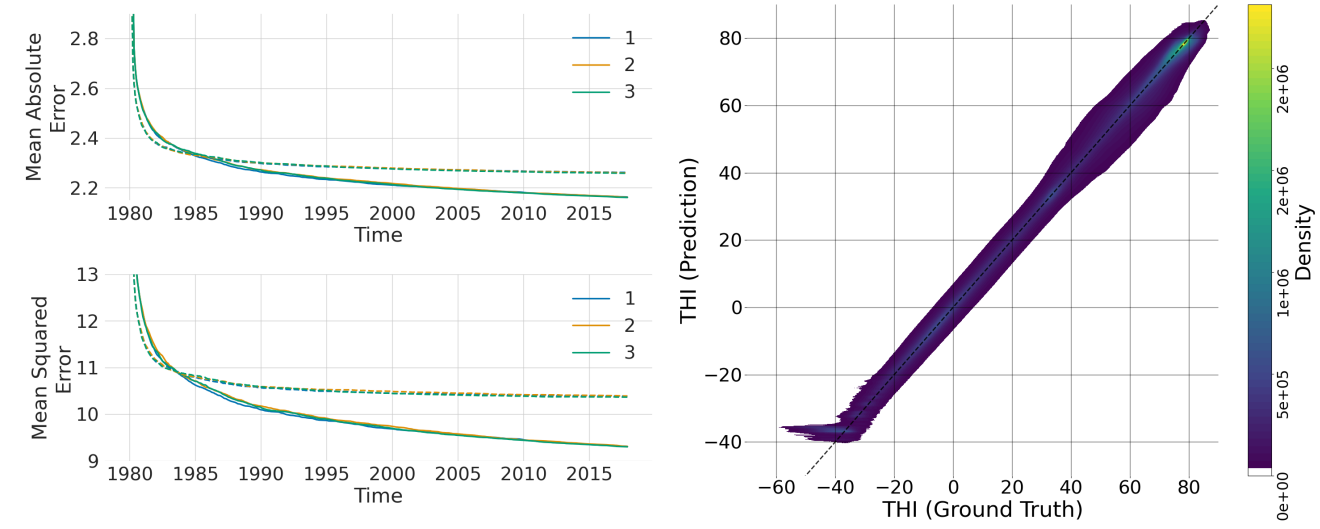

Figure 1. Evolution of Mean Absolute Error (MAE) and Mean Squared Error (MSE) throughout the training process for three distinct XGBoost models, each represented by a different color. The left panel shows the MAE (top) and MSE (bottom) metrics over training epochs, conducted on a monthly basis from January 1980 to December 2017. Solid lines depict the metrics evaluated on the training set in each epoch, while dashed lines represent the MAE and MSE evaluated on the test set at each epoch. The right panel displays a density plot of the predictions from Model 1 versus the ground truth for the validation data set (2018-2020), where the diagonal line indicates the optimal prediction performance.

of the clustering behavior within the large dataset (approximately 10 billion data points). As evident, the model demonstrates good performance, with the majority of points concentrated near the diagonal, representing optimal prediction. However, a potential limitation is observed at the lower end of the Thermal Humidity Index (THI) range. Here, the model appears to exhibit a prediction floor around ∼-40 THI. It is important to note that these low THI values are of minimal interest for heat stress studies, as they primarily correspond to regions like Antarctica without human and cattle populations.

The models were trained on a single compute node, which was equipped with two AMD EPYC/Milan 64-core CPUs and 256 GB of RAM. During both the training and inference phases, each model was configured to utilize 128 parallel processes, optimizing computational efficiency. Consistent with the approach outlined in the Methods section, the training process for the models was executed incrementally, with the dataset being segmented into monthly intervals. This approach facilitated the storage of checkpoints at the conclusion of each epoch, allowing for a systematic evaluation and resumption of the training process without loss of progress. It total, the models were trained on approximately 130 billion examples; areas comprised entirely of sea or ocean were omitted.

The performance metrics of the three models were found to be closely comparable across the evaluation criteria. Both the Mean Absolute Error (MAE) and Mean Squared Error (MSE) demonstrated a continuous decrease throughout the training epochs, albeit at a diminished rate of reduction as the training progressed. It is noteworthy that these metrics were also assessed

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

Open Access Earth System Science Data Discussions

Table 4. MAE, MSE and R^2 performance metrics evaluated on the last epoch during the training process, the test set for the three models (data from Jan 2018) and the evaluation set (Feb 2018 - Dec 2020). Furthermore, the total training time and the time needed by the models to perform inference on a single year are presented.

	Training set			Test set			Validation set			Total training	Time to evaluate
Model no.	**MAE**	**MSE**	R^2	**MAE**	**MSE**	R^2	**MAE**	**MSE**	R^2	**time (hrs)**	**1 year (min)**
1	2.163	9.306	0.942	2.262	10.386	0.940	3.432	19.014	0.943	**564**	$\sim$ **205**
2	2.160	9.294	0.940	2.260	10.392	0.941	**3.402**	**18.624**	**0.944**	792	$\sim$ 375
3	**2.159**	**9.279**	**0.939**	**2.259**	**10.366**	**0.944**	3.403	18.754	**0.944**	1156	$\sim$ 480

using a test set that was not seen by the model during the training phase, ensuring the evaluation of the model's predictive capability on unseen data. As the complexity of the models increased, notably, Model 3 required a substantially longer duration for training compared to its counterparts. Furthermore, an observable convergence between the curves representing MSE and MAE was observed, in both the training and test sets, indicating a stabilization in the models' performance over time. Table 4 shows the performance metrics of the three models across the training and test sets, as well as the total training time and time needed to perform inference on a single year and scenario. Lastly, the metrics obtained from the validation set were closely comparable across all three models; MAE was found to be $\sim$3.4, MSE $\sim$19 and R^2 $\sim$0.94.

To further assess the comparative performance of the three trained models, we performed inference using data from 2018 to 2020 (ERA5 reanalysis) to evaluate the precision of the THI predictions relative to the ground truth. Figure 2 displays THI predictions at four randomly selected grid points and time intervals, representing diverse climatic conditions: a permanent frost region (top left), a moderate climate (top right), and two hot climate regions (bottom panels). Across these examples, the outputs from all three models closely followed the real THI fluctuations during the 10-day periods shown. Additionally, the predictions generated by the three models were nearly identical, as shown in Figure 2.

The outputs from the three models on ERA5 data align well with the ground truth THI. To evaluate the similarity of their performance on CMIP6 future projection data, we conducted inference using a single year of data from the ACCESS-ESM1-5 model (year 2020 under scenario SSP2-4.5) for all three models. Figure 3 presents the THI outputs at four randomly selected geographical locations and time points over a 10-day period. The outputs from all models closely match each other, corroborating their consistency. Combined with the previously obtained performance metrics, this indicates that the three models exhibit similar performance on both ERA5 and CMIP6 data. Consequently, we opted for the simplest model (Model 1) due to its significant computational cost savings compared to the other models. The marginal improvements in performance metrics did not justify the additional tens of thousands of CPU-hours required for the more complex models, given the close similarity in their outputs.

To further assess model performance and identify potential systematic errors, we employed global maps at randomly chosen time points from the validation set. These maps visualized the Temperature Humidity Index (THI) using both ground truth data and the chosen model's predictions, along with their difference. Representative examples are shown in Figure 4. Deviations from the ground truth are evident in various regions across the globe at these specific hourly time points.

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

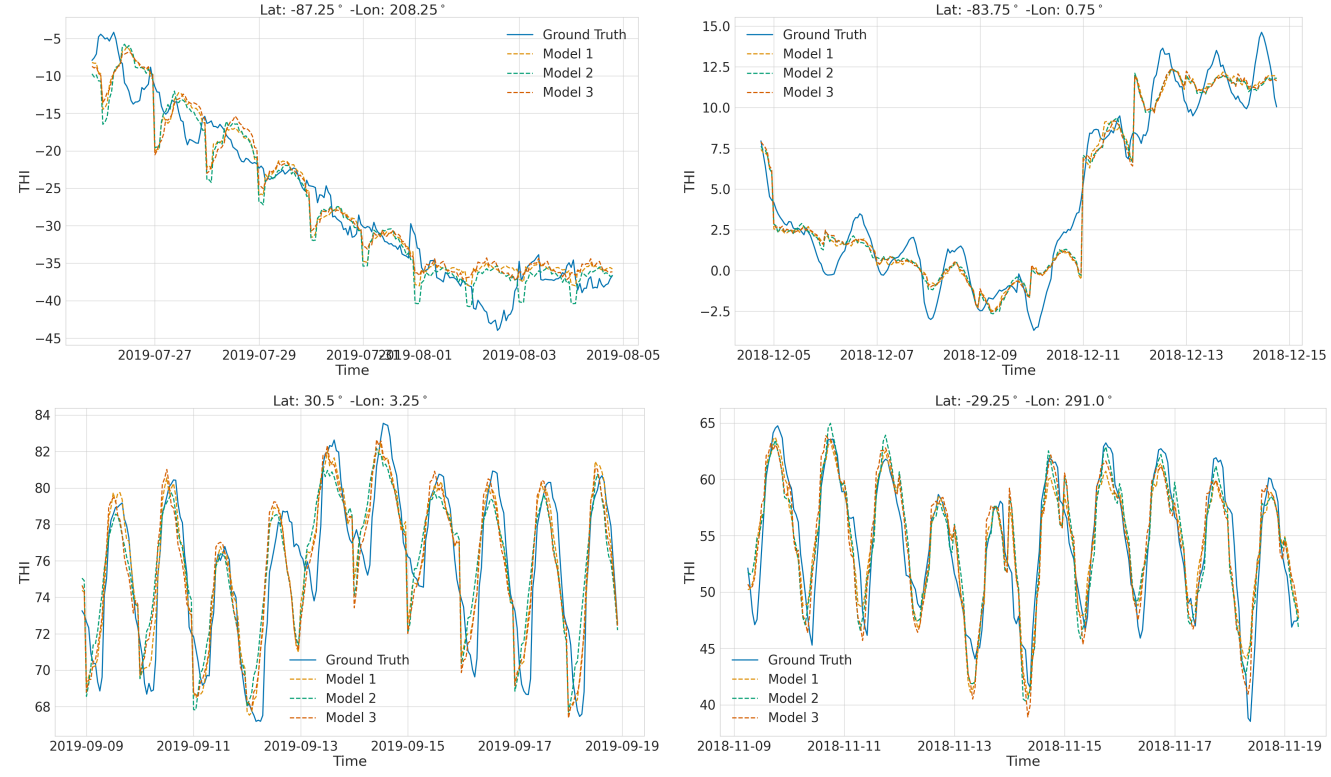

Figure 2. Comparative visualization of hourly THI calculations using ERA5 reanalysis data (ground truth) and the outputs from the three XGBoost models, each differentiated by unique colors. The four-panel display (2x2 arrangement) showcases THI profiles across varying climatic conditions: the top left panel presents an example from a cold climate region, the top right panel illustrates a moderate climate scenario, and the bottom panels depict examples from warm climate regions. This arrangement provides a comprehensive overview of the models' performance and accuracy in replicating ground truth THI values across a spectrum of environmental conditions. This comparison explains further the density q-q plot.

To investigate the presence of systematic errors, such as consistent over/underestimation in specific areas, we averaged the differences across the entire validation set. This resulted in a new set of global maps presented in Figure 5. The mean difference between ground truth and prediction reveals minimal deviations from zero across most of the world, suggesting an absence of systematic errors. However, a slight underestimation (approximately less than 1 THI unit) is observed near the borders of India and Pakistan, and in the eastern part of the Arabian Peninsula. Conversely, a notable overestimation of THI is present in a large portion of Antarctica. This overestimation is of limited concern for this study's scope, as Antarctica is an uninhabited region with no risk of heat stress. Due to the low relevance of heat stress in this uninhabited region and to optimize computational resources, Antarctica was excluded from further inference using Model 1.

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

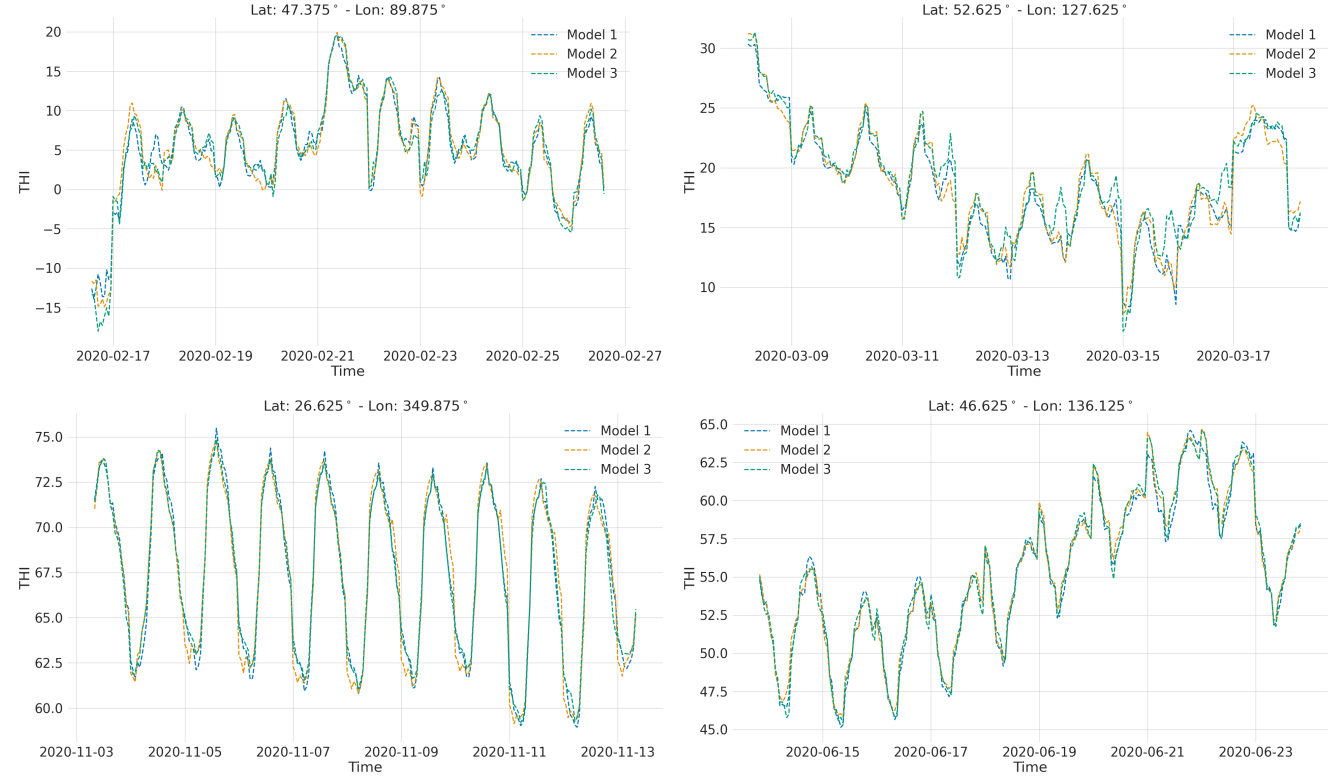

Figure 3. Comparative visualisation of the THI profiles from the three models' predictions from the ACCESS-ESM1-5 model (SSP2-45 scenario) at the year 2020. The four panels show four randomly selected grid points and the prediction from each model is colour-coded.

Building upon these findings, we employed Model 1 for inference using 12 GDDP NASA-NEX CMIP6 models under two distinct climate scenarios: SSP2-4.5 (representing a moderate stabilization emission scenario) and SSP5-8.5 (representing a business-as-usual scenario with rising emissions until the end of the century). Utilizing all combinations of climate models and scenarios, we generated datasets spanning the period 2020 to 2100. These datasets are publicly available at https://doi.org/10.26050/WDCC/THI for further investigation and use in climate change impact studies (Georgiades, 2024).

4 Conclusions

Climate change, driven by anthropogenic emissions, entails a significant risk to ecosystems and societies worldwide. One of the anticipated consequences is rising global temperatures. The agricultural sector, vital for global food security and economies, is particularly vulnerable. Dairy farming, a crucial sub-sector, faces significant economic challenges due to heat stress impacting dairy cattle. Heat stress in dairy cows is commonly quantified using the Thermal Humidity Index (THI), a simple metric

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

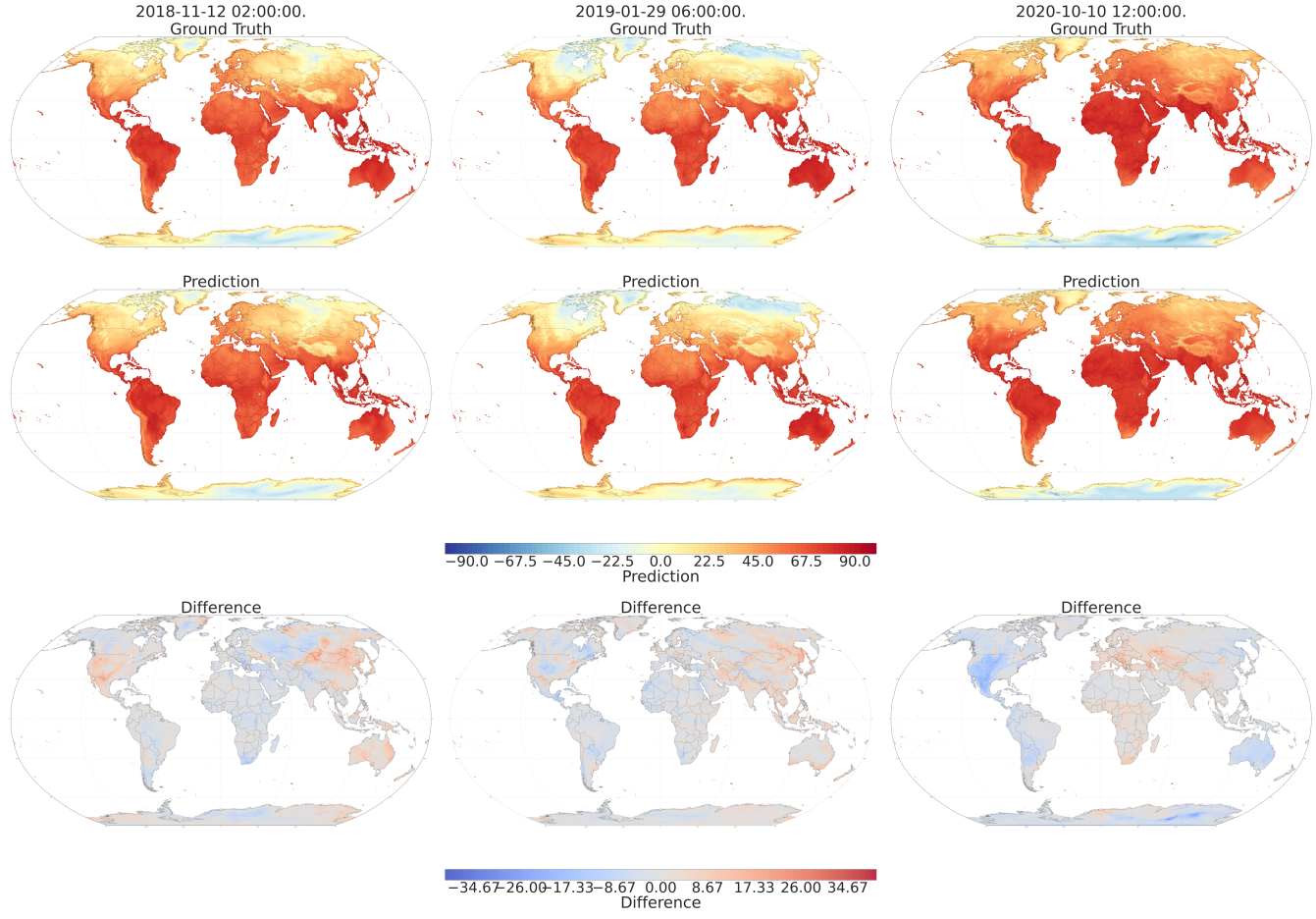

Figure 4. Comparison of Ground Truth THI and Model Predictions. The first row displays the Ground Truth THI values from ERA5 data for three randomly selected time points within the evaluation period. The second row shows the corresponding THI predictions from our model. The third row illustrates the differences between the Ground Truth and the model predictions (Ground Truth - Prediction).

requiring only temperature and humidity data. However, most existing literature utilizes daily THI values, lacking the necessary granularity to capture the crucial intraday climatic variability for accurate heat stress estimation.

To address this limitation, we trained a machine learning model (XGBoost regressor) on global hourly historical reanalysis data (ERA5) to effectively downscale daily climate variables to hourly THI values. Our models demonstrably performed well against ground truth data from an independent validation period. The implicit assumption in this approach is that the diel cycle of the THI does not alter significantly under climate change scenarios.

Leveraging the good performance and agreement between the three models, we employed the most computationally efficient model to generate global hourly THI projections until the end of the century. This involved utilizing 0.25° GDDP NASA-NEX CMIP6 data with 12 climate models and two emission scenarios (SSP2-4.5 and SSP5-8.5).

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

Open Access Earth System Science Data Discussions

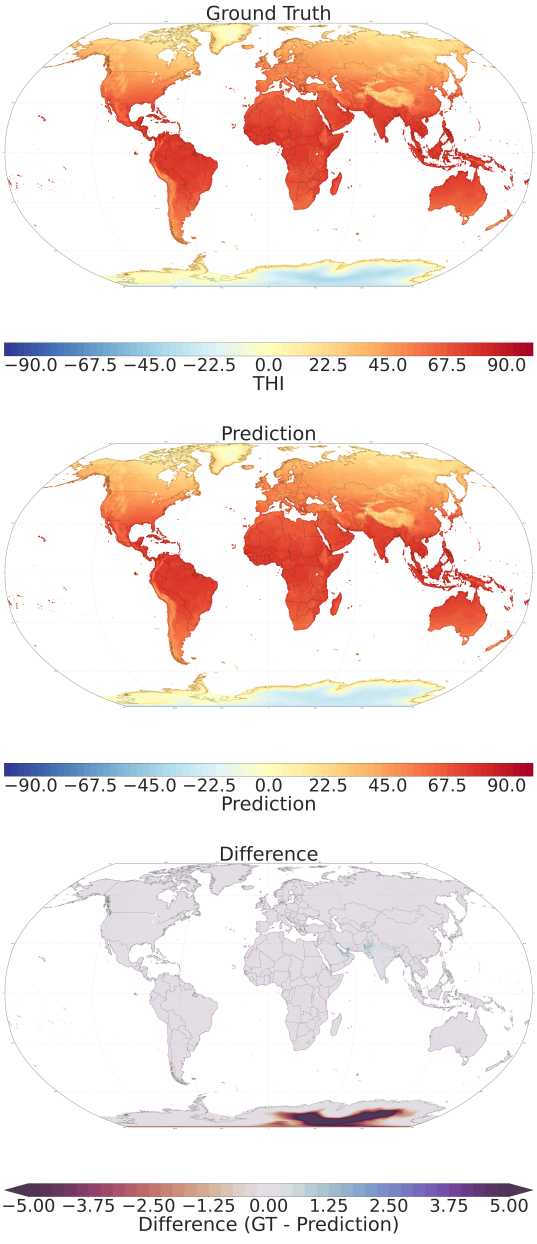

Figure 5. Comparison of Ground Truth THI and model predictions averaged over the validation set (2018-2020). In the top panel, the mean THI over the three year period is presented, whereas in the middle panel, the corresponding THI predictions from our model is shown. The third panel presents the difference between the two (Ground Truth - Prediction).

The generated hourly THI datasets hold significant potential to contribute towards the optimization of heat stress management in the dairy industry. These datasets can empower stakeholders with the ability to create highly accurate and geographi-

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

cally specific heat stress risk assessments for individual farms. This information can then be used to develop targeted mitigation strategies, allowing farmers, agricultural communities and organizations to proactively manage heat stress and optimize animal well-being and production efficiency. Furthermore, incorporating these datasets into climate change adaptation plans allows policymakers and the dairy cattle sector to develop long-term strategies for ensuring the sustainability of the dairy industry in the face of a changing climate. Ultimately, this research paves the way for a more resilient and sustainable future for dairy farming.

5 Code and data availability

Code to reproduce the results presented in this paper is available in the public github repository github.com/pantelisgeor/Temperature-Humidity-Index-ML. The code provided is written in Python and the workload can be executed by running a series of bash scripts, as documented in the repository description. The code is provided under an MIT lisence, which allows for users to freely use and modify the code.

The data produced in this study are available at https://doi.org/10.26050/WDCC/THI (Georgiades, 2024) in NetCDF format, with an hourly temporal resolution and 0.25° spatial resolution. The datasets are published under a CC BY 4.0 license.

Author contributions. PG: Conceptualization, Data curation, Format Analysis, Methodology, Software, Validation, Original draft preparation, TE: Methodology, Review & editing, YP: Data curation, Software, Resources, Review & editing, JA: Data curation, Software, Review & editing, JL: Review & Editing, MN: Conceptualization, Methodology, Validation, Original draft preparation

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank the High Performance Computing Facility of the Cyprus Institute for their support in the computational and storage needs for this study.

This research was supported by the EMME-CARE project that has received funding from the European Union's Horizon 2020 Research and Innovation Program, under Grant Agreement No. 856612, as well as matching co-funding by the Government of Cyprus. This research was also supported by the PREVENT project that has received funding from the European Union's Horizon Europe Research and Innovation Program under Grant Agreement No. 101081276 and the European High Performance Computing Joint Undertaking (JU) and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia under grant agreement No 101101903.

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

References

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., and Younis, I.: A review of the global climate change impacts, adaptation, and sustainable mitigation measures, Environmental Science and Pollution Research, 29, 42 539–42 559, https://doi.org/10.1007/s11356-022-19718-6, 2022.

Baca, A., Wiener, S., Treasure, E., and Myers, J. M.: United States Department of Agriculture, https://www.climatehubs.usda.gov/content/southeast-hub-cattle-heat-stress-fact-sheet, accessed: 28/03/2024, 2019.

Bernabucci, U., Biffani, S., Buggiotti, L., Vitali, A., Lacetera, N., and Nardone, A.: The effects of heat stress in Italian Holstein dairy cattle, Journal of Dairy Science, 97, 471–486, https://doi.org/10.3168/jds.2013-6611, 2014.

Bohmanova, J., Misztal, I., and Cole, J.: Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress, Journal of Dairy Science, 90, 1947–1956, https://doi.org/10.3168/jds.2006-513, 2007.

Bouraoui, R., Lahmar, M., Majdoub, A., Djemali, M., and Belyea, R.: The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate, Animal Research, 51, 479–491, https://doi.org/10.1051/animres:2002036, 2002.

Brügemann, K., Gernand, E., König von Borstel, U., and König, S.: Defining and evaluating heat stress thresholds in different dairy cow production systems, Archives Animal Breeding, 55, 13–24, https://doi.org/10.5194/aab-55-13-2012, 2012.

Burhans, W., Rossiter Burhans, C., and Baumgard, L.: Invited review: Lethal heat stress: The putative pathophysiology of a deadly disorder in dairy cattle, Journal of Dairy Science, 105, 3716–3735, https://doi.org/10.3168/jds.2021-21080, 2022.

Cartwright, S. L., Schmied, J., Karrow, N., and Mallard, B. A.: Impact of heat stress on dairy cattle and selection strategies for thermotolerance: a review, Frontiers in Veterinary Science, 10, https://doi.org/10.3389/fvets.2023.1198697, 2023.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, https://doi.org/10.48550/ARXIV.1603.02754, 2016.

Cheng, M., McCarl, B., and Fei, C.: Climate Change and Livestock Production: A Literature Review, Atmosphere, 13, 140, https://doi.org/10.3390/atmos13010140, 2022.

Escarcha, J., Lassa, J., and Zander, K.: Livestock Under Climate Change: A Systematic Review of Impacts and Adaptation, Climate, 6, 54, https://doi.org/10.3390/cli6030054, 2018.

Georgiades, P.: Temperature Humidity Index GDDP-NEX-CMIP6 ML projections, https://doi.org/10.26050/WDCC/THI, 2024.

Hahn, G. L.: Dynamic responses of cattle to thermal heat loads, Journal of Animal Science, 77, 10, https://doi.org/10.2527/1997.77suppl_210x, 1997.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., and Yang, H.: Machine learning and artificial intelligence to aid climate change research and preparedness, Environmental Research Letters, 14, 124 007, https://doi.org/10.1088/1748-9326/ab4e55, 2019.

Igono, M. O., Bjotvedt, G., and Sanford-Crane, H. T.: Environmental profile and critical temperature effects on milk production of Holstein cows in desert climate, International Journal of Biometeorology, 36, 77–87, https://doi.org/10.1007/bf01208917, 1992.

https://doi.org/10.5194/essd-2024-344
Preprint. Discussion started: 22 August 2024

IPCC: Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, https://doi.org/10.1017/9781009325844, 2023.

Kadzere, C., Murphy, M., Silanikove, N., and Maltz, E.: Heat stress in lactating dairy cows: a review, Livestock Production Science, 77, 59–91, https://doi.org/10.1016/s0301-6226(01)00330-x, 2002.

Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., and Knowlton, N.: Climate change and ecosystems: threats, opportunities and solutions, Philosophical Transactions of the Royal Society B: Biological Sciences, 375, 20190 104, https://doi.org/10.1098/rstb.2019.0104, 2020.

Mauger, G., Bauman, Y., Nennich, T., and Salathé, E.: Impacts of Climate Change on Milk Production in the United States, The Professional Geographer, 67, 121–131, https://doi.org/10.1080/00330124.2014.921017, 2014.

Moore, S. S., Costa, A., Penasa, M., Callegaro, S., and De Marchi, M.: How heat stress conditions affect milk yield, composition, and price in Italian Holstein herds, Journal of Dairy Science, 106, 4042–4058, https://doi.org/10.3168/jds.2022-22640, 2023.

North, M. A., Franke, J. A., Ouweneel, B., and Trisos, C. H.: Global risk of heat stress to cattle from climate change, Environmental Research Letters, 18, 094 027, https://doi.org/10.1088/1748-9326/aceb79, 2023.

Polsky, L. and von Keyserlingk, M. A.: Invited review: Effects of heat stress on dairy cattle welfare, Journal of Dairy Science, 100, 8645–8657, https://doi.org/10.3168/jds.2017-12651, 2017.

Ravagnolo, O., Misztal, I., and Hoogenboom, G.: Genetic Component of Heat Stress in Dairy Cattle, Development of Heat Index Function, Journal of Dairy Science, 83, 2120–2125, https://doi.org/10.3168/jds.s0022-0302(00)75094-6, 2000.

Sheik, S., Mohammed, R., Teeparthi, K., and Raghuvamsi, Y.: Machine Learning-Based Prediction of Intergranular Corrosion Resistance in Austenitic Stainless Steels Exposed to Various Heat Treatments, Journal of The Institution of Engineers (India): Series D, https://doi.org/10.1007/s40033-024-00675-y, 2024.

St-Pierre, N., Cobanov, B., and Schnitkey, G.: Economic Losses from Heat Stress by US Livestock Industries, Journal of Dairy Science, 86, E52–E77, https://doi.org/10.3168/jds.s0022-0302(03)74040-5, 2003.

Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., and Nemani, R.: NASA Global Daily Downscaled Projections, CMIP6, Scientific Data, 9, https://doi.org/10.1038/s41597-022-01393-4, 2022.

WMO: Guide to instruments and methods of observation. Wmo.int, World Meteorological Organization (WMO), https://library.wmo.int/idurl/4/68695, 2021.

Yeck, R. G.: Guide to environmental research on animals, National Academy of Sciences, 1971.

Zhou, M., Groot Koerkamp, P., Huynh, T., and Aarnink, A.: Evaporative water loss from dairy cows in climate-controlled respiration chambers, Journal of Dairy Science, 106, 2035–2043, https://doi.org/10.3168/jds.2022-22489, 2023.