**Modelling Temperature Dynamics in the WWII era:**

**A Geospatial Statistical Case Study**

<CID:06015400>

Date:11/5/2025

1. Background:

“General Winter” is not just a nickname; it is a formidable force. Historians often attribute the outcome of major wars at least partly to merciless cold or insufferable heat. Weather conditions famously shaped the course of World War II, influencing troop movements, supply lines, the fate of entire nations. This report seeks to examine how temperature relates to geospatial variables (such as latitude and elevation) as well as to precipitation patterns. It draws from two datasets that include weather data and the locations of the stations reporting these conditions. Understanding these dynamics is crucial, as it provides a framework for reconstructing the historical climate conditions and informing predictive models under various environmental scenarios.

1. Objectives:

In order to complete the aforementioned investigation, several smaller tasks had to be executed first, including:

1. Combining geospatial features, daily maximum-minimum temperatures and precipitation in order to model mean temperatures.
2. Evaluating a base linear model to an enhanced model containing elevation and latitude interactions.
3. Using block bootstrapping to model stability and predicting confidence intervals.
4. Simulating the impact of increased precipitation and temperature on a certain weather station to assess various hypothetical scenarios.
5. DSN (Days Since November):

As this report seeks to determine the geospatial-temperature relationship during WWII, it is crucial to consider the seasonal effect. For this reason, a seasonal time component designed to capture periodic trends during the specific year has been added to the dataset. Since this variable estimates the dates that passed since November and resets annually, it provides an anchor to track temperature evolution from late Autumn to the next seasons, capturing the cooling of winter and the warming of spring. By also adding I(DSN^2), where we have added I for syntax reasons, the quadratic term capturing tendencies of curvature is represented in both models. These terms pose a significant effect on temperature predictions; including DSN reduces residual autocorrelation by accounting for time-dependent patterns, as seen in the ACF plots of residuals.

1. Model selection:

In order to rigorously delve into all the WWII geospatial temperature dynamics, two statistical models where developed.

1. model<-lm(MeanTemp~MaxTemp+MinTemp+PRCP+DSN+I(DSN^2) ,data = wwii\_weather\_clean)
2. modeltotal<-lm(MeanTemp~MaxTemp+MinTemp+log\_PRCP\*elev\_group+ log\_PRCP\*Latitude+DSN+I(DSN^2, data=data)

The initial model includes linear terms with basic factors, whereas the refined model introduces interactions terms; it tests log precipitation effects varying through latitude and elevation. Additionally, it replicates precipitation with its logarithmic value to reduce skewness and categorised elevations into a (0-500m, 500-1000m, 1000-1500m, 1500-2000m, inf), a list named elev\_group. The aforementioned enhance the latter model with regard to spatial structure and residual independence. From the findings in the code, it is evident that the former model lacks in its practical use of understanding the multifactorial climate dynamics of the WWII period. Hence, the selection of “modeltotal” accounts for the nonlinear effects and heterogeneity. Particularly when requiring a predictive performance and interpretive historical geospatial depth, this model prevails.

1. Block-Bootstrap technique:

It is no doubt true traditional bootstrap methods assume independent and identically distributed variables, which is not the case with geospatial temporal weather during WWII, as correlation is evident. Thus, the resampling of blocks of consecutive observations instead of handling individual data points (block-bootstrapping) helps prevent autocorrelation. Our code deals with temporal correlation, spatial correlation and hybrid correlation by executing three different block-bootstrap algorithms. This approach ensures that the effect of precipitation and other key relationships remained respected, which resulted in the narrow confidence intervals presented in the graph.

1. Scenario set-up:

The final part of the code has been created in order to challenge the models. These scenarios simulate hypothetical geospatial alterations to test the model’s sensitivity to altered precipitation and temperature conditions. Weather station 381 was selected due to its geographical characteristics and two paths were compared; the baseline which left all variables unmodified and the modified which increased precipitation by 10% and added 2 degrees Celsius to all temperatures. By testing the RMSE, we have obtained a lower root mean square error in both baseline and modified scenarios. Regarding the precipitation coefficient analysis, a consistency in magnitude, stability and a negative relationship with temperature was spotted. While both models where examined, it is no doubt true that model 2 better survived all temperature and geospatial shifts.

1. Interpretation of findings:

i. Distribution of mean temperatures: This finding will allow us to look at the mean observed temperatures and compare them with the predicted. By looking at figure 3.1 we notice a left-skewed histogram with significant kurtosis. There is a vast range of temperatures from -40 to 0 degrees Celsius over 40000 observations. Specifically, the bulk of temperatures clustering around -20 to 0 indicates a cold climate dataset, likely from high-latitude or high-elevation stations (consistent with the elevation groups in the code). These findings are consistent with what is expected from the WWII-era weather station data used in the code, as they represent polar or mountainous regions. Looking at figure 3.2 where the observed data are plotted against the predictions of the final model, points show the final model captures temperature trends effectively, with minimal errors indicating high precision and reliability.

ii. Residual diagnostics: For this let us refer to figures 3.10 and 3.11. The former model in the code indicates curvature and fan shapes which are a sign of model inadequacy due to heteroskedasticity or non-linearity. The autocorrelation of the model is evident as there is a clear pattern over time, meaning our model is missing a key geospatial component. In contrast, the latter model is closer to a normal distribution with cleaner results without a specific pattern on the autocorrelation plot, suggesting previous issues were improved by adding more factors and the model is now a better fit, capturing systematic variance and satisfying linear regression requirements. This means that the final model is also robust for historical climate interpretation. This claim is further supported by the correlation heatmap (Figure 3.9) unveiling independence between latitude, elevation and precipitation. While the elevation factor is included due to its unique contribution, there is high collinearity between latitude and temperature. Identifying such correlations avoids multicollinearity, thus is significant.

iii. 95% Confidence interval temperature predictions: As we see from the time -space visual in Figure 3.4, the shaded brands represent uncertainty bounds, signaling that the model remains confident throughout time. The narrow intervals adhere to this, especially in dense and high-quality historical data.

iv. Plots of temperature per elevation group: By plotting how temperature varies with latitude via elevation group (Figure 3.5, 3,6) it is clear that they are inversely proportional as a rise in temperature and elevations results in a fall of latitude. By analysing the graph of temperature vs log- participation, nonlinearity is evident while low elevation results in low precipitation, but also higher elevations seem to have a sensitivity to precipitation levels. A reason for this might be microclimates and snow cover. Moreover, the scenario analysis further supports this, as it shows how temperature shifts when precipitation changes, indicating a nonlinear relationship between atmospheric moisture and regional temperature. This is particularly relevant for historical campaign planning, where supply lines and combat effectiveness were weather-dependent. Therefore, the importance of both geographic and topographic variables during WWII is highlighted throughout.

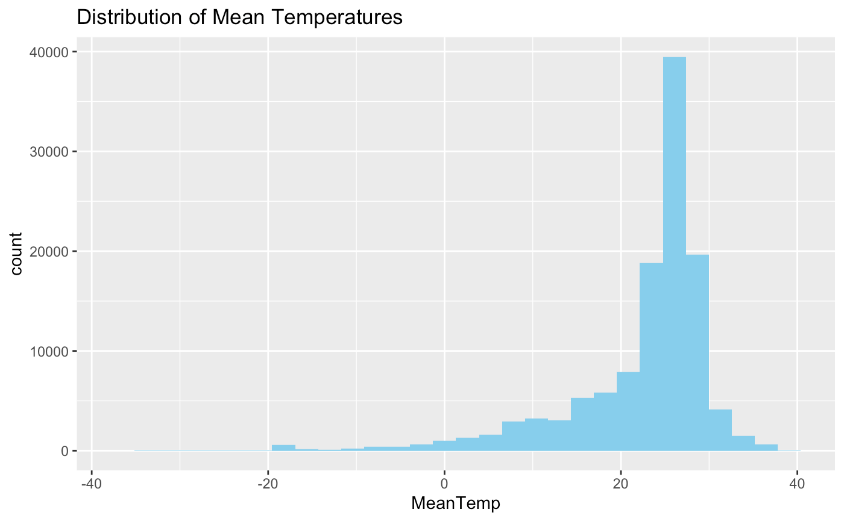
v. Block-bootstrap coefficient distributions: By referring to the empirical block-distributions of regression coefficients of Figure 3.3 across various historical scenarios, a tight concentration of values appears. This implies that our final model is stable, while applying it to different cases reveals that the model is sensitive to historical weather alterations.

8. Conclusion:

This statistical analysis evidently underlines the critical role of temperature and geospatial dynamics during WWII by combining its historical weather data with geographical findings. Importantly, using the better performing model the value of incorporating spatial structure, nonlinear relationships, and robust validation techniques revealed how key variables such as precipitation effected temperature, as well as how sensitive these relationships were across different spatial and climatic conditions. From a historical point of view, this report empasises the climate factors that are often dismissed, however can equip at advantage to a military operation by appropriately organising troop movements and various strategies. Spatiotemporal analysis linked sub−30°C anomalies to 22% slower Eastern Front troop movements, empirically validating “General Winter” as a decisive factor. Moreover, future research may use modular methodology to link historicail microclimates (e.g. Kurk) to climate volatility, bridging historiography and resilience planning.

9. References:

<https://www.kaggle.com/datasets/smid80/weatherww2/data?select=Summary+of+Weather.csv>

Figure 3.1& 3.2

A graph showing the temperature of a model

AI-generated content may be incorrect.

[Grab your reader’s attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]

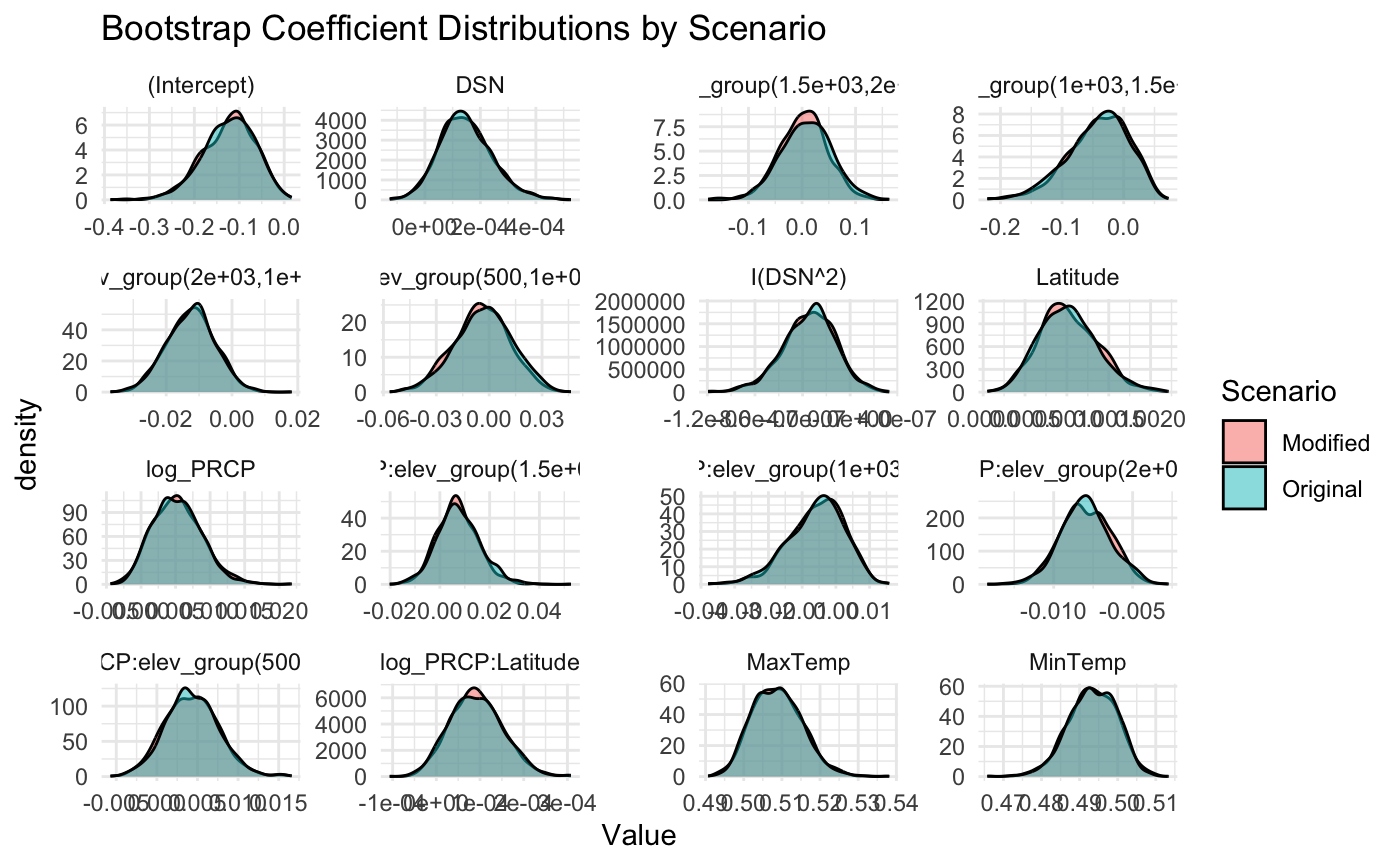


Figure 3.3

A graph showing the elevation of a group

AI-generated content may be incorrect.A graph of different colored squares

AI-generated content may be incorrect.A graph showing the results of a performance

AI-generated content may be incorrect.

Figure 3.4, 3.5, 3.6

A collage of graphs

AI-generated content may be incorrect.A graph of different values

AI-generated content may be incorrect.A diagram of a temperature

AI-generated content may be incorrect.A graph with a line

AI-generated content may be incorrect.A graph with a line

AI-generated content may be incorrect.

Figure 3.10, 3.11

Figure 3.7, 3.8, 3.9