Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

Context selection for stylistically controlled weather forecast text generation

Tessa van Dijk [1], Dirk Wolters [2], Jasper S. Wijnands [1,*] *Corresponding author: jasper.wijnands@knmi.nl

This study investigates the use of large language models to generate Dutch weather forecasts from NWP data while adhering to KNMI's stylistic and domain-specific requirements. Instead of producing text directly from raw data, we link current NWP inputs to similar historical NWP–text pairs. Dimensionality reduction helps cluster meteorologically and stylistically aligned examples. Early results show that meaningful matches can be found, enabling controlled, example-guided generation and highlighting the promise of context-specific data-to-text forecasting.

Introduction

The Royal Netherlands Meteorological Institute (KNMI) maintains a long-standing practice of issuing public weather forecasts multiple times a day. These are composed by meteorologists in Dutch based on numerical weather prediction (NWP) data. Given the linguistic precision and accessibility requirements of KNMI forecasts, the generated texts must conform closely to an established style guide.

This project explores how LLMs can assist KNMI meteorologists by generating draft weather forecasts that align with their established linguistic and structural conventions. We investigate a context selection-based approach: using NWP data to identify historical forecasts that are semantically and stylistically similar and using these to guide or control model generation (e.g. [2]). The full model pipeline can be observed in Figure 1.

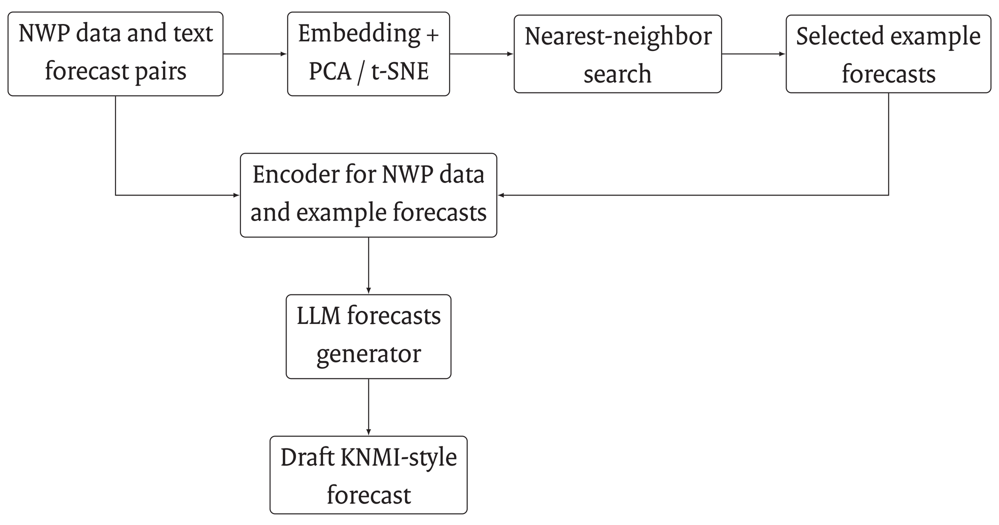

Figure 1: *Schematic overview of the context selection pipeline. Current NWP data are embedded and organised via PCA/t-SNE. Nearest historical NWP–text forecast pairs are selected to provide stylistic and meteorological guidance to an LLM, which generates a draft KNMI-style weather forecast for meteorological review.*

Data description

Two matched datasets spanning 2021–2023 were constructed: (i) NWP forecasts from KNMI's operational HARMONIE-AROME model, including 13 weather variables at nine representative Dutch locations (as seen in Figure 2), and (ii) textual forecasts authored by KNMI meteorologists, with revised versions removed to ensure consistency with the original forecast intent and corresponding NWP data. Each forecast text follows a structured format, with paragraphs corresponding to specific forecast periods, and was aligned with matrices of NWP data for the same time blocks.

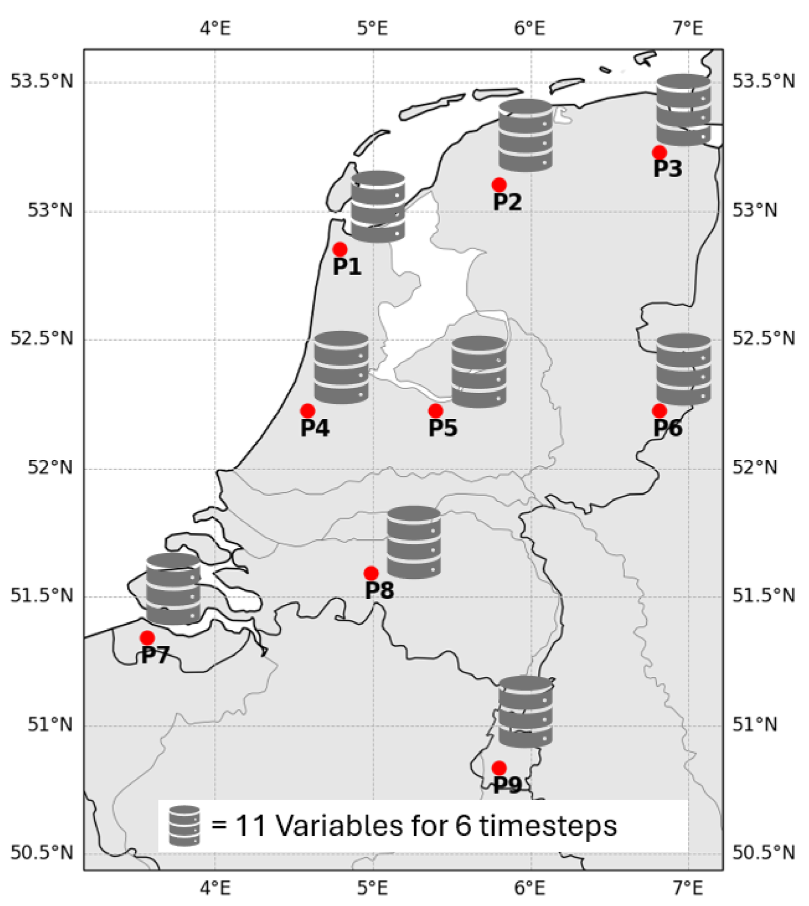

Figure 2: *The nine coordinates for which the numerical weather prediction data is extracted.*

Preliminary results

To guide the model toward stylistically and semantically accurate text generation, we first apply preprocessing and dimensionality reduction techniques to identify semantically and meteorologically similar historical forecasts.

Principal component analysis

Principal component analysis (PCA) was applied to project the NWP dataset into two dimensions to assess whether major meteorological patterns are preserved. As shown in Figure 3, the projections reveal broad gradients for both 'Total Cloud cover' and 'Visibility', each varying along different axes, and several outliers (corresponding to extreme weather conditions). However, most points cluster near the center of the plot, indicating that substantial variability is compressed in the reduced space and not fully captured by PCA. Across the dataset, atypical events such as snowfall or storms form loose groupings, though substantial overlap persists.

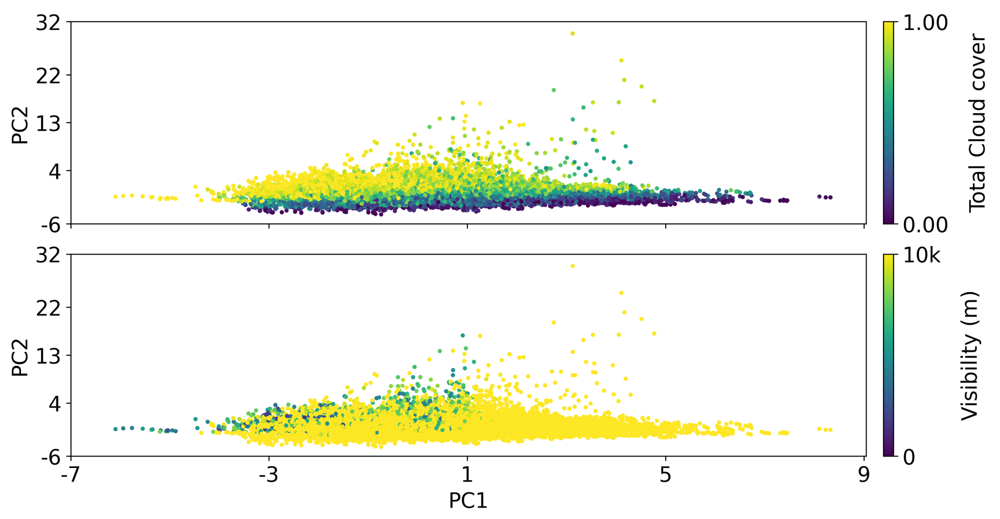

Figure 3: *PCA embedding with overlay of the fraction of total cloud cover (top) and visibility capped at 10 km (bottom).*

t-SNE Analysis

Using t-SNE, the dataset is projected into two dimensions to visualize the latent structure of the NWP data. While no clear clusters form, localized orderings appear when points are colored by specific variables. Overlays for total cloud cover and visibility reveal coherent gradients—cloud cover increasing left to right and visibility varying top to bottom—more separated than in the PCA plot.

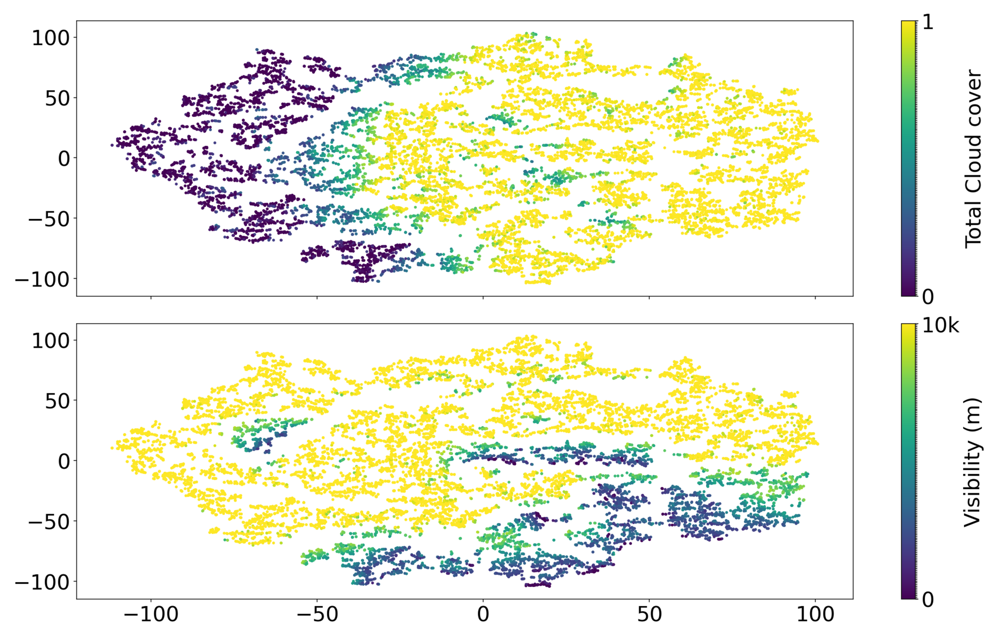

Figure 4: *t-SNE embedding with overlay of the fraction of total cloud cover (top) and visibility capped at 10 km (bottom).*

Table 1 shows a forecast sample and its neighbors. All describe calm, mostly dry conditions with light to moderate wind and similar temperatures. Neighbor 1 is most similar, matching the nighttime setting and northerly flow. Neighbor 2 mainly differs in timing but retains similar dry, cool conditions. Neighbor 3 is the most distinct, noting possible early rain and a westerly to northwesterly wind, though it remains broadly consistent with the stable weather pattern.

Table 1: *Example data point with neighbors from t-SNE, translated to English.*

Original	Neighbor #1	Neighbor #2	Neighbor #3
Tonight there will be few clouds and it will remain dry. Temperatures will drop to 10°C locally in the northeast, reaching 14°C at the coast. The wind will come from a northerly to northeasterly direction and will be light to moderate, at sea and on the IJsselmeer initially moderate to fairly strong.	Tonight will be dry. At the end of the night, there is a chance of fog in the north, which will quickly dissipate. The minimum temperature will be between 8°C in the north and 15°C in the southwest. There will be a slight breeze, with a moderate northerly breeze along the coast and on the IJsselmeer.	This afternoon there will be plenty of sunshine. Cloudy skies will develop in the north. It will remain dry everywhere, with temperatures ranging from 15°C in the northern coastal area to 18°C in the south. There will be a light to moderate northwesterly wind.	Tonight, it may rain in the southeast at first, but elsewhere it will be dry with clear spells. The wind will come from a westerly to northwesterly direction. Wind speed will be light to moderate over land, occasionally strong along the coast.

Embedding-based similarity results

To evaluate how well PCA and t-SNE preserve the embedding structure, we perform a 1-nearest-neighbor (1-NN) retrieval using Euclidean distance, comparing each sample's 1-NN in the projections to its 1-NN in the original embedding space [1, 5]. We then compute the percentage of overlaps (Table 2). Across all models, t-SNE clearly outperforms PCA, achieving roughly 20% overlap compared to PCA's 1–2%. t-SNE also yields higher average similarity scores (79.1%, 98.7%, 90.8%) than PCA (63.3%, 98.1%, 83.0%), indicating that it better preserves semantic neighbors.

Table 2: *Overlap between 1-NNs obtained from PCA/t-SNE projections and the 1-NNs obtained from the embedding space*

Model	PCA	t-SNE
distiluse-base-multilingual-cased-v1	1.8%	20.0%
paraphrase-multilingual-mpnet-base-v2	1.6%	19.0%
RobBERT	1.6%	19.6%

Model training

NWP data was connected to a pre-trained LLM using a separate encoder. By freezing the LLM weights and only tuning the encoder parameters, LLM output can be steered towards the training texts. Various LLMs have been tested, including gpt2-small-dutch (0.1B parameters) [6], EuroLLM (1.7B parameters) [3] and gpt-oss (21B parameters) [4]. As expected, validation performance improves substantially when upgrading the base LLM model, but only for Dutch-supporting LLMs (see Fig. 5).

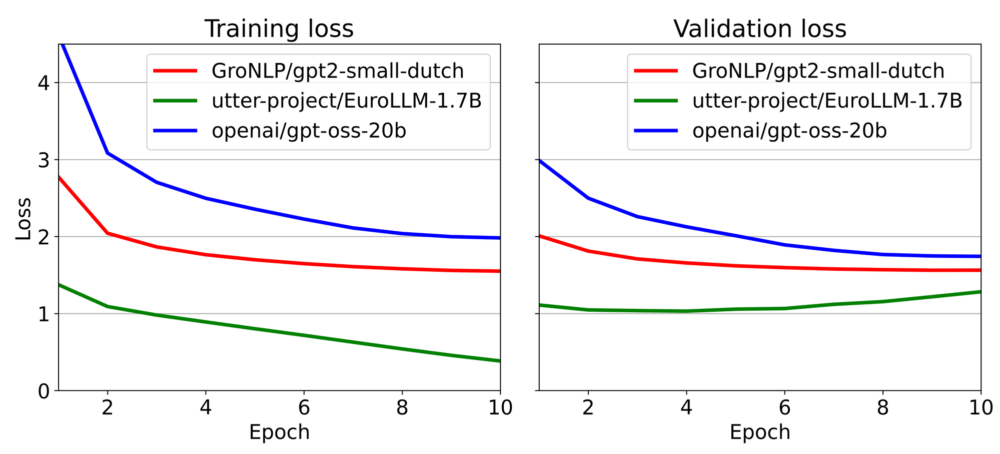

Figure 5: *Cross-entropy loss during training (left) and validation (right).*

Conclusion

This study tackles the initial challenge of generating Dutch weather forecasts from structured NWP data, with a focus on linguistic accuracy and stylistic conformity. A core contribution is the exploration of a context selection-based approach, where contextually similar weather scenarios can be selected to guide generation, which is expected to improve both accuracy and adherence to KNMI's stylistic norms. The ability to identify similar forecast texts given current weather conditions is an important first step towards steering text generation. Results show that selecting stylistically similar examples using NWP data is certainly possible. Modeling then focuses on adding the factual content from the NWP data. This makes context selection in this manner a promising research direction for controlled data-to-text generation.

References

[1] Delobelle, P., Winters, T., Berendt, B. (2020). RobBERT: a Dutch RoBERTa-based Language Model. arXiv:2001.06286.

[2] Lin, S., Wang, W., Yang, Z., Liang, X., Xu, F. F., Xing, E., Hu, Z. (2019). Data-to-text generation with style imitation. arXiv:1901.09501.

[3] Martins, P. H., Fernandes, P., Alves, J., Guerreiro, N. M., Rei, R. et al. (2024). EuroLLM: Multilingual Language Models for Europe. arXiv:2409.16235

[4] OpenAI, Agarwal, S., Ahmad, L., Ai, J., Altman, S. et al. (2025). gpt-oss-120b & gpt-oss-20b Model Card. arXiv:2508.10925

[5] Reimers, N., Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv:1908.10084.

[6] de Vries, W., Nissim, M. (2020). As good as new. How to successfully recycle English GPT-2 to make models for other languages. arXiv:2012.05628.

[1] R&D Observations and Data Technology, Royal Netherlands Meteorological Institute. [2] Department of Weather and Climate Services, Royal Netherlands Meteorological Institute.