

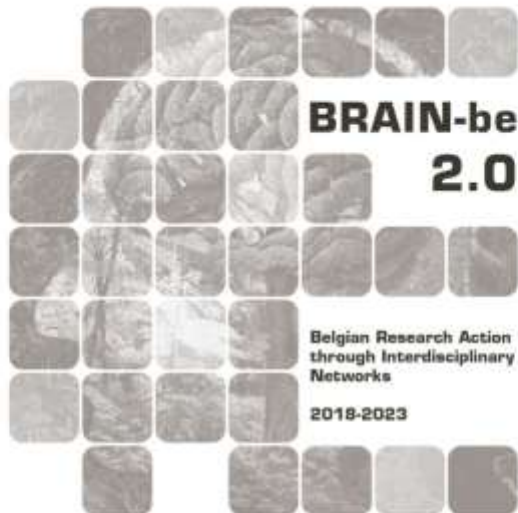
CS-MASK

Crowd-Sourced data for atmospheric Modelling At Sub-Kilometric scale

Prof. Dr. Ir Caluwaerts Steven (RMI/UGent) – Prof. Dr. Termonia Piet (RMI) – Thomas Vergauwen (RMI/UGent)

Pillar 1: Challenges and knowledge of the living and non-living world





NETWORK PROJECT

CS-MASK

Crowd-Sourced data for atmospheric Modelling At Sub-Kilometric scale

Contract - B2/202/P1/CS-MASK

FINAL REPORT

PROMOTORS: Prof. Dr. ir Caluwaerts Steven (RMI/UGent)
Prof. Dr. Termonia Piet (RMI)

Dr. Degrauwe Daan (RMI)

AUTHORS: Prof. Dr. ir Caluwaerts Steven (RMI/UGent)
Prof. Dr. Termonia Piet (RMI)

Thomas Vergauwen (RMI/UGent)





Published in 2026 by the Belgian Science Policy Office

WTCIII

Simon Bolivarlaan 30 bus 7

Boulevard Simon Bolivar 30 bte 7

B-1000 Brussels

Belgium

Tel: +32 (0)2 238 34 11

<http://www.belspo.be>

<http://www.belspo.be/brain-be>

Contact person: Corinne Lejour

Tel: +32 (0)2 238 34 91

Neither the Belgian Science Policy Office nor any person acting on behalf of the Belgian Science Policy Office is responsible for the use which might be made of the following information. The authors are responsible for the content.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without indicating the reference:

Caluwaerts Steven, Termonia Piet, Vergauwen Thomas. Crowd-Sourced data for atmospheric Modelling At Sub-Kilometric scale Final Report. Brussels: Belgian Science Policy Office 2026 – 25p. (BRAIN-be 2.0 - (Belgian Research Action through Interdisciplinary Networks))

TABLE OF CONTENTS

ABSTRACT	5
OBJECTIVES	5
CONCLUSIONS.....	5
KEYWORDS.....	5
1 INTRODUCTION	6
2. STATE OF THE ART AND OBJECTIVES	7
3. METHODOLOGY	9
4. SCIENTIFIC RESULTS AND RECOMMENDATIONS	16
SCIENTIFIC RESULTS	16
RECOMMENDATIONS.....	21
5. DISSEMINATION AND VALORISATION	23
6. PUBLICATIONS	24
7. ACKNOWLEDGEMENTS	25

ABSTRACT

Objectives

The CS-MASK project (Crowd-Sourced data for atmospheric Modelling At Sub-Kilometric scale) investigated the potential of non-traditional, crowdsourced (CS) meteorological observations for advancing sub-kilometric numerical weather prediction (NWP) over Belgium. The project was carried out by the Royal Meteorological Institute of Belgium (RMI).

Conclusions

The main results of CS-MASK are: (1) a hectometric AROME-SURFEX NWP experiment over Belgium for August 2020, demonstrating the feasibility of 700 m runs and providing urban heat island maps for Belgian cities; (2) an improved automated quality control pipeline for crowdsourced temperature observations, introducing priority-group processing and LCZ-informed safety nets; and (3) the MetObs-toolkit, an open-source Python package for processing non-traditional meteorological observations, published in the Journal of Open Source Software and adopted by the international community. In addition, (4) the project contributed to a theoretical framework for the verification of high-resolution NWP models using dense non-traditional observation networks, demonstrating that traditional error-based scores are fundamentally ill-suited for the many-observations-per-grid-cell regime that dense IoT networks create. All of these results will be combined in the PhD thesis dissertation document of Thomas Vergauwen.

Keywords

crowdsourced observations, quality control, hectometric NWP, urban heat island, model verification

1. INTRODUCTION

Accurate prediction of high-impact weather events such as heatwaves and convective storms requires high-resolution NWP models supported by dense, reliable observational networks. The Royal Meteorological Institute of Belgium (RMI), as the federal institution responsible for operational weather forecasting and climate monitoring, has a direct mandate to advance these capacities in support of public safety and evidence-based policy. The CS-MASK project addressed two interconnected challenges at the frontier of hectometric NWP: setting up for the first time a framework to undertake hectometric NWP runs at the RMI and exploring the potential of crowdsourced, non-traditional meteorological observations for high-resolution model verification. The project was executed by Thomas Vergauwen at the RMI as a four-year PhD trajectory in collaboration with Ghent University.

2. STATE OF THE ART AND OBJECTIVES

Scientific context

High-resolution NWP relies on the ability to represent processes at scales where local land-surface characteristics (e.g. urban morphology, vegetation cover, water bodies,...) directly influence near-surface meteorology. At kilometric resolution, these effects are partially captured through tile-fraction approaches. At sub-kilometric scale, individual urban districts, parks, and water bodies start to be explicitly resolved, offering the potential for much more accurate urban climate modelling and for better prediction of convective initiation.

The RMI, as Belgium's national meteorological institute, operates an NWP system and maintains the observational infrastructure underpinning it. Its operational forecasts with the ALARO/AROME models at 1.3 km resolution are among the highest-resolution operational systems in Europe, yet the NWP community was, at the time of submitting the CS-MASK proposal, approaching the threshold of sub-kilometric resolution forecasting. Running models at 700 m resolution over Belgium would promise improved representation of local processes with direct relevance for RMI's responsibilities: early warning for convective storms, urban heat guidance during heatwaves, sea-breeze forecasting along the Belgian coast,...

At the same time, the classical observational infrastructure consisting of WMO-compliant synoptic stations sited in open, rural environments was never designed to evaluate models at these resolutions. A synoptic station in an open field is representative for the weather of a large surrounding area; it carries no information about the urban heat island, the cooling effect of a park, or the moisture flux from a water body. Validating hectometric NWP models that can include such microscale effects with synoptic networks alone is therefore fundamentally limited.

Non-traditional observations have the potential to offer a way forward. Networks of personal weather stations (PWS), citizen science projects, and IoT sensors are producing an ever-growing stream of near-real-time near-surface meteorological data, covering also urban and peri-urban environments that synoptic networks deliberately avoided. Their key weaknesses, such as variable (and in general lower) data quality, incomplete metadata, and heterogeneous siting, demands robust automated quality control (QC) methods before the data can be used scientifically. Internationally, efforts such as TITAN (Båserud et al., 2020) and CrowdQC+ (Fenner et al., 2021) have begun to address this gap, but no approach has yet integrated land-cover information into the quality control logic to mitigate false rejections in heterogeneous environments.

Objectives

The original objectives of CS-MASK were:

- WP1 – Data collection and analysis: (Task 1.1) Build a quality-controlled crowdsourced observational database covering Belgium for August 2020, combining synoptic and non-traditional observations. (Task 1.2) Evaluate available high-resolution land information datasets.
- WP2 – Hectometric runs: (Task 2.1) Establish a framework for 700 m ALARO-SURFEX runs over Belgium and (Task 2.2) produce model data for August 2020 at 4 km, 1.3 km, and 700 m.

- WP3 – Added value: (Task 3.1) Evaluate the hectometric runs against the quality-controlled crowdsourced database. (Task 3.2) Investigate urban parameterisation improvements.

Adaptation of objectives

During the project, the European Destination Earth (DestinE) (<https://destination-earth.eu/>) initiative was launched as a flagship European Commission programme to develop a high-fidelity digital twin of the Earth, including operational hectometric NWP at pan-European scale with substantial dedicated computational infrastructure. This initiative fundamentally changed the strategic landscape for individual national efforts at hectometric model development. Investing project resources in pushing the ALARO model into a resolution regime that DestinE was already covering at even higher resolutions and with far greater resources would have resulted in duplication without strategic added value for Belgium or the broader community.

The project team therefore made the pragmatic decision to shift emphasis: a framework to undertake hectometric runs (Tasks 2.1) was successfully initiated. As a proof-of-concept and to support verification studies, two NWP experiments have been executed (Task 2.2) with AROME-SURFEX at 700 m and 1400 m. A similar 4 km experiment was not produced because the switch from ALARO to AROME precluded a meaningful resolution comparison at that scale. Given all efforts undertaken within DestinE, the originally planned effort on optimising the urban parameterisation scheme and producing fully polished operational-quality hectometric output was de-emphasised. Instead, resources were redirected towards aspects of the proposal that DestinE did not focus on and where CS-MASK could make a genuinely novel contribution: (Task 1.1) the quality control of non-traditional observations and the open-source tooling needed to work with these data, and (Task 3.1) the theoretical and practical frameworks for verifying high-resolution models using dense heterogeneous observation networks. As a result, no contribution was made to (Task 3.2) the urban parameterisation scheme. Finally, ongoing international initiatives also impacted the original Task 1.2 goal to evaluate high-resolution land information datasets. Instead of developing our own dataset Thomas Vergauwen contributed to international initiatives in this domain and made a scientific stay in Met Éireann to work, within the DestinE project, on high-resolution physiography datasets.

3. METHODOLOGY

Observational dataset and quality control

The observational data used in CS-MASK cover Belgium for August 2020. This month was selected for its meteorologically rich character containing a major 12-day heatwave (5–16 August) with strong nocturnal urban heat islands, a few sea-breeze days even reaching Brussels, and several deep convective episodes with intense, localized precipitation patterns.

Data from ten observation networks, as listed in Table 1, have been used for the project.

Table 1: Observational networks used by the CS-MASK project, the type of network and the number of stations for August 2020.

Network	Category	N
Synoptic network (WMO compliant)	Synoptic	29
MOCCA	Regional climate network	6
VLINDER	Regional climate network	59
VMM (air quality)	Air quality monitoring	40
Meetnet Vlaamse Banken	Maritime network	6
Infrabel	Transport monitoring	70
AWV	Transport monitoring	31
Meteoroutes	Transport monitoring	49
WOW	PWS	343
Netatmo	PWS	4003

They include data from different networks with different characteristics and purposes:

- Synoptic network: 29 WMO-compliant stations, high quality, predominantly rural.
- VLINDER network: 59 research-grade automatic weather stations deployed in partnership with secondary schools, covering diverse land cover types (cities, forests, lakes, dunes). Uses Davis Vantage Pro 2 sensors.
- MOCCA network: 6 high-accuracy stations of an urban climate monitoring network in city of Ghent.
- Several networks operated by government agencies and other professional actors: air quality (VMM, 40 stations), maritime (Meetnet Vlaamse Banken, 6 stations), and road/transport infrastructure networks (Infrabel 70, AWV 31, Metroutes 49).

- WOW network: 343 personal weather stations uploading near-real-time data to the Weather Observations Website platform. Variable sensor quality and siting.
- Netatmo: About 4000 low-cost weather stations with very high spatial density in urban areas. This network has the lowest sensor quality and lacks reliable metadata.

Observations were resampled to hourly resolution. Elevation and Local Climate Zone (LCZ) class were assigned to each station using gridded geospatial datasets. An overview of all stations used and their different siting characteristics is given in Figure 1.

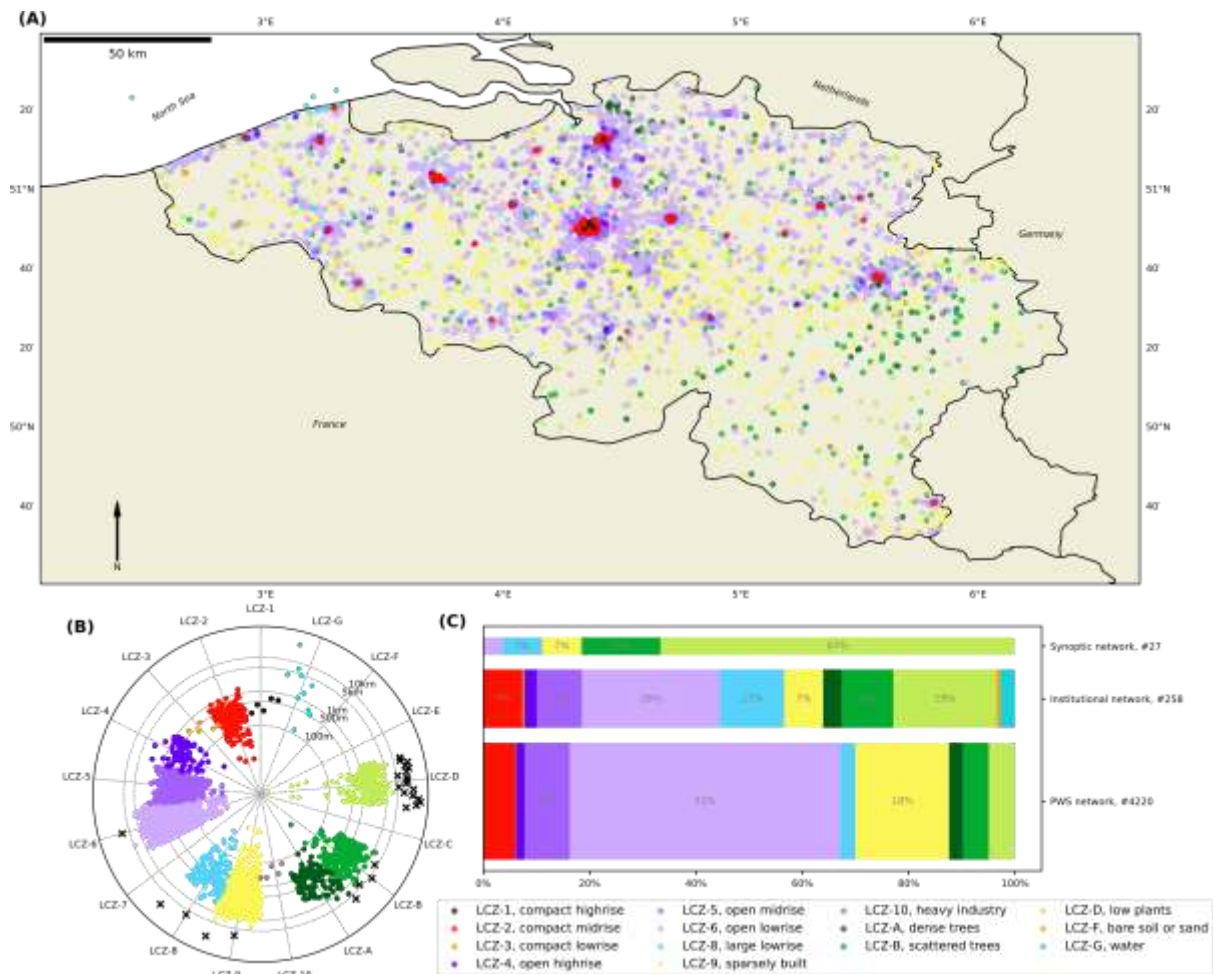


Figure 1 : Complementing synoptic and PWS stations results in a diverse land cover/land use representation. (A) Spatial distribution of the combined stations with indication of the LCZ they are located in. (B) A radar plot showing the distribution of stations across LCZ categories and their local density. The distance to the nearest station is indicated by the radial lines (log-scale). The black crosses represent synoptic stations, for which distances are computed to the nearest synoptic neighbors only, illustrating the sparseness of the synoptic-only network. (C) Percentage of stations located in a particular LCZ for a specific type of observational network. The colors correspond to the LCZ categories.

A dedicated QC treatment is key to be able to use such a diverse observational dataset. The core QC method applied for CS-MASK was implemented in the MetObs-toolkit developed for this project. For this project two important scientific extensions, going beyond what is typically done in similar QC frameworks, have been implemented. These innovations are situated in the so-called buddy check, a traditional step in the QC where each observation is compared with its spatial neighbours (called buddies) to evaluate its plausibility.

Innovation 1 - priority-group processing: The networks were assigned priority labels reflecting their assumed quality. Three priority groups were defined: group 1 containing the WMO compliant stations (synoptic network), group 2 containing other high-quality networks (all monitoring and regional climate networks: MOCCA, VLINDER, VMM, Meetnet Vlaamse Banken, Infrabel, AWW, Meteoroutes), and group 3 collecting all personal weather stations (WOW and Netatmo networks). The MetObs-toolkit was applied iteratively: first to the highest-priority subset, whose surviving observations became a non-rejectable background, called gatekeepers, for the next iteration. This prevented the statistical properties of lower-quality networks (which dominate by number) from corrupting the quality assessment of better-quality, often distinctively sited stations.

Innovation 2 - LCZ safety nets: The conventional buddy check rejects observations that deviate from their spatial neighbours within a search radius of a few kilometres, assuming representativeness at that scale. In urban and peri-urban environments, this assumption fails: a station in a dense urban core is meteorologically different from a station 2 km away in a park or near a water body. Two safety nets have been implemented: the *LCZ-safety net* reevaluates observations rejected by the spatial buddy-check by searching for buddies sharing the same LCZ class within an extended radius; the *LCZ-agg-safety net* applies the same logic using broader aggregated LCZ categories and is used as a fallback when insufficient specific-LCZ buddies are available. If the observation is consistent with its LCZ (or LCZ-aggregated) peers, it is retained. This substantially reduced false rejections of scientifically valuable observations located at unique microenvironments.

Hectometric NWP

A cycled AROME-SURFEX experiment at 700 m resolution was configured over Belgium, downscaled from ARPEGE global analyses. The AROME model (CY48T3) was coupled to SURFEX, which includes the Town Energy Balance (TEB) scheme for urban surfaces and the ISBA scheme for natural surfaces. Surface physiography was based on ECOCLIMAP-SG. Figure 2 illustrates the output of this run for the 2m temperature field at 18 UTC on 5 August 2020 showing an inland progressing sea-breeze front.

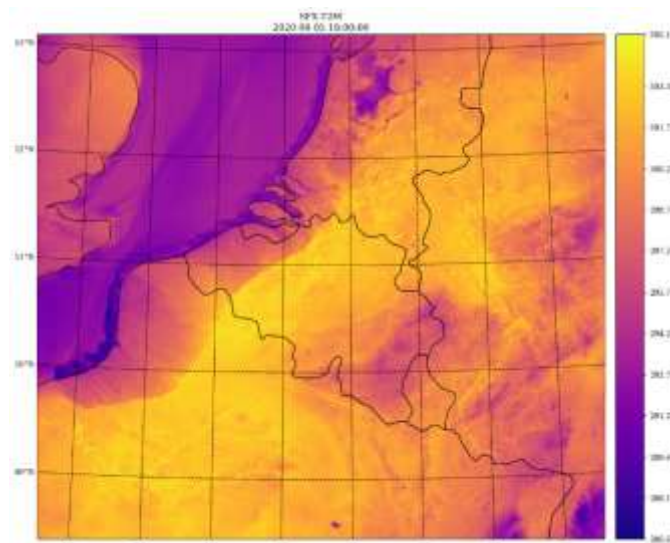


Figure 2. The 2 m temperature field at 18 UTC on 5 August 2020 from the AROME-SURFEX (700 m) experiment. The sea breeze front is visible as a sharp temperature gradient, with cooler temperatures near the coast and warmer temperatures inland.

The model was run for August 2020 in cycling mode with a 24-hour cycling interval and 34-hour forecasts, respecting a 6-hour spin-up period to construct a continuous time series as illustrated in Figure 3.

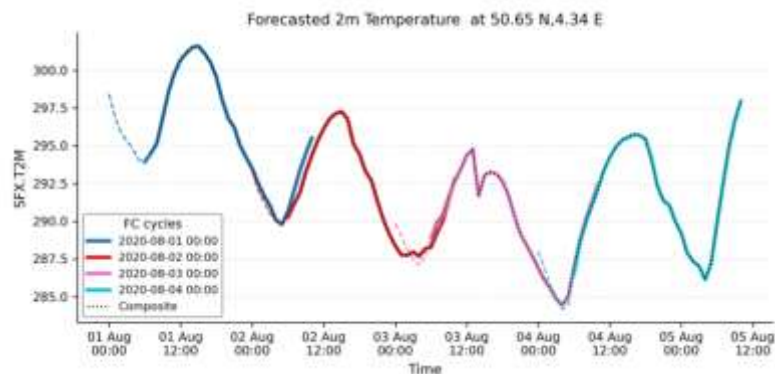


Figure 3. Time stacking of 2 m temperature forecasts from the cycled AROME-SURFEX (700 m) experiment. A continuous time series (black dotted line) is obtained by selecting the output with the shortest lead time while respecting a 6 h spin-up period. Dashed lines indicate the spin-up period of each cycle; colours distinguish individual cycles.

The model experiments were run on the ECMWF HPC machine ATOS. The Deode prototype, developed in the context of the DestinE project, was used as the scripting framework.

Verification framework

Verification of the hectometric runs was carried out at three levels:

Traditional verification used standard error-based scores (bias, RMSE, MAE) computed against the synoptic network.

Traditional verification using crowdsourced data computed the same error-based scores (bias, RMSE, MAE) but against the quality-controlled observations from all networks. The inclusion of non-traditional observations substantially increased the number of verification data points and improved spatial coverage in urban and non-standard environments as can be seen in Figure 4.

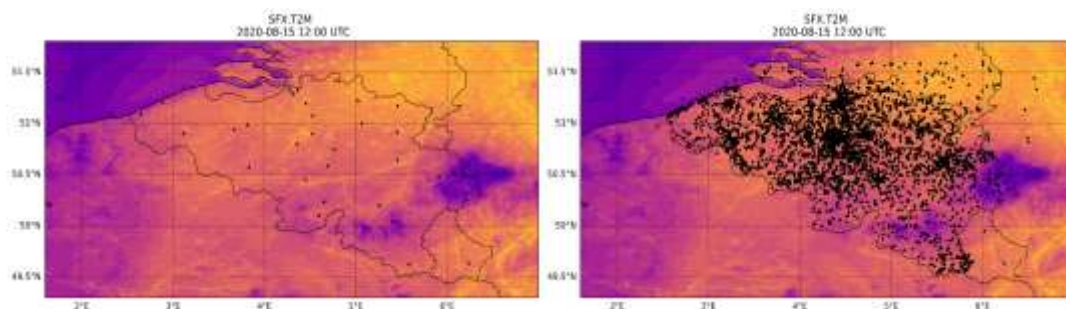


Figure 4. Comparison of the observational density and spatial distribution of traditional synop-only verification (left) and the inclusion of non-traditional observations (right).

Beyond traditional verification addresses a fundamental methodological issue that arose when using dense non-traditional networks for verification. When a model grid point corresponds to multiple observations (the "many-to-one" verification regime illustrated in Figure 5), conventional error-

based scores are affected by two systematic biases: (1) they do not account for subgrid temperature variability, and (2) averaging errors over many observations does not converge to a meaningful model quality measure when the observations sample different microclimates within the same grid cell.



Figure 5. Schematic illustration of the ‘one-to-one’, and the ‘many-to-one’ verification regime. The colors of the diamond shapes represent the characteristic microenvironment of the observation.

A theoretical framework was developed based on sample-based scores (SBS), specifically the Z^2_SBS , which tests whether the forecast value is consistent with the observed distribution within each grid cell rather than penalising it for not matching any individual observation. This theoretical framework is described and discussed in detail in the PhD dissertation of Thomas Vergauwen. The value of SBS if a high observational density is present, was evaluated using a so-called *perfect model setup*. Figure 6 illustrates how a 700 m AROME forecast was used as a surrogate truth, from which artificial forecasts at coarser resolutions (2–7.5 km) were derived, and verification scores were computed. This demonstrated that error-based scores become increasingly misleading in the many-to-one regime, while the sample-based score remains informative across all regimes (Figure 7).

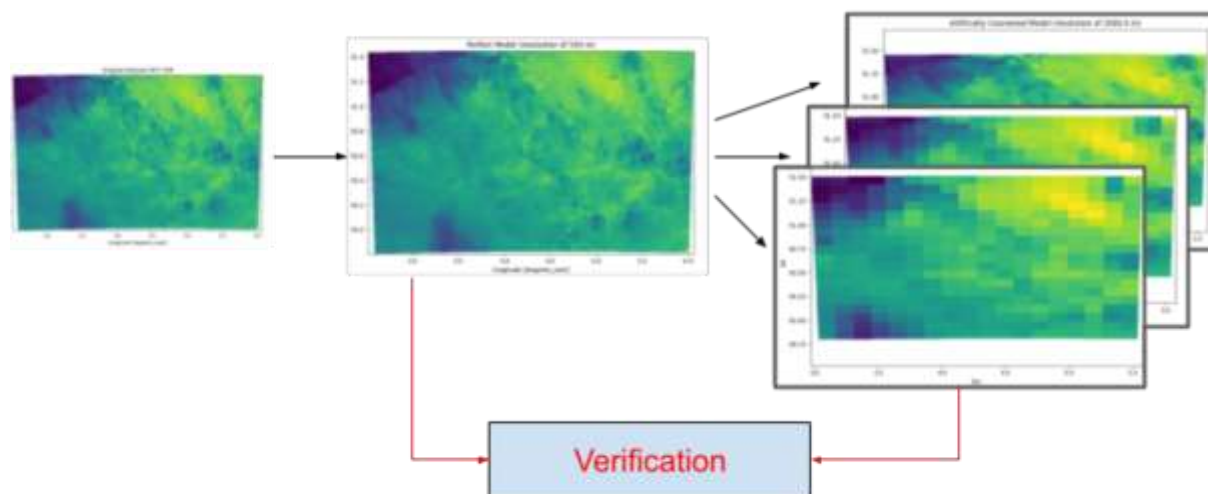


Figure 6: Illustration of the perfect model setup. The perfect model is created by interpolating a 2m temperature field from a 700m AROME-Surfex (cy48t3) run to a 500m grid. Artificial forecasts at different resolutions are generated from this perfect model. Verification is then applied to the artificial forecasts, using the perfect model as if it were a gridded observational dataset.

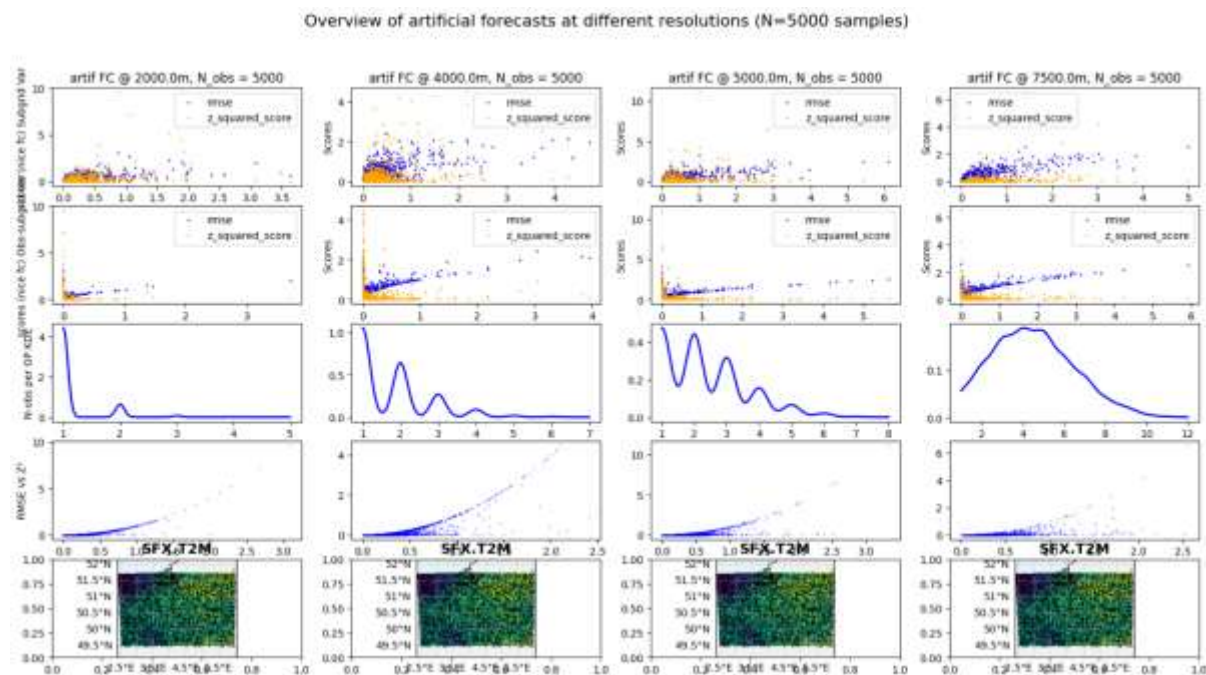


Figure 7: Overview of the effect of subgrid variability on RMSE and Z^2_{SBS} for different artificial forecasts (columns). Scores are based on a random sample of 5 000 observations over the whole domain. The first two rows show the relationship between scores (per grid point) and subgrid variability. In the first row, subgrid variability is defined as the sample variance from the perfect model; in the second row, it is defined as the sample variance from the observations. The third row shows kernel density estimates (KDEs) of the average number of observations per grid point, thereby characterising the verification regime. The last row presents scatter plots of RMSE versus Z^2_{SBS} per grid point. For RMSE, scores per grid point are computed by averaging the squared errors over all associated observations and then taking the square root.

MetObs-toolkit

The MetObs-toolkit is an open-source Python package developed during CS-MASK to address the practical barriers to using non-traditional meteorological data for science. As schematically summarized in Figure 8, the toolkit provides an integrated workflow from raw data ingestion to analysis-ready output, covering: templated data import for heterogeneous formats, time series resampling and synchronisation, automated quality control, gap filling using interpolation or ERA5 reanalysis data, and analysis utilities including landcover correlation estimates via the Google Earth Engine API. The toolkit is designed for ease of installation (pure Python, available via PyPI and Conda-Forge) and is documented on Read the Docs.

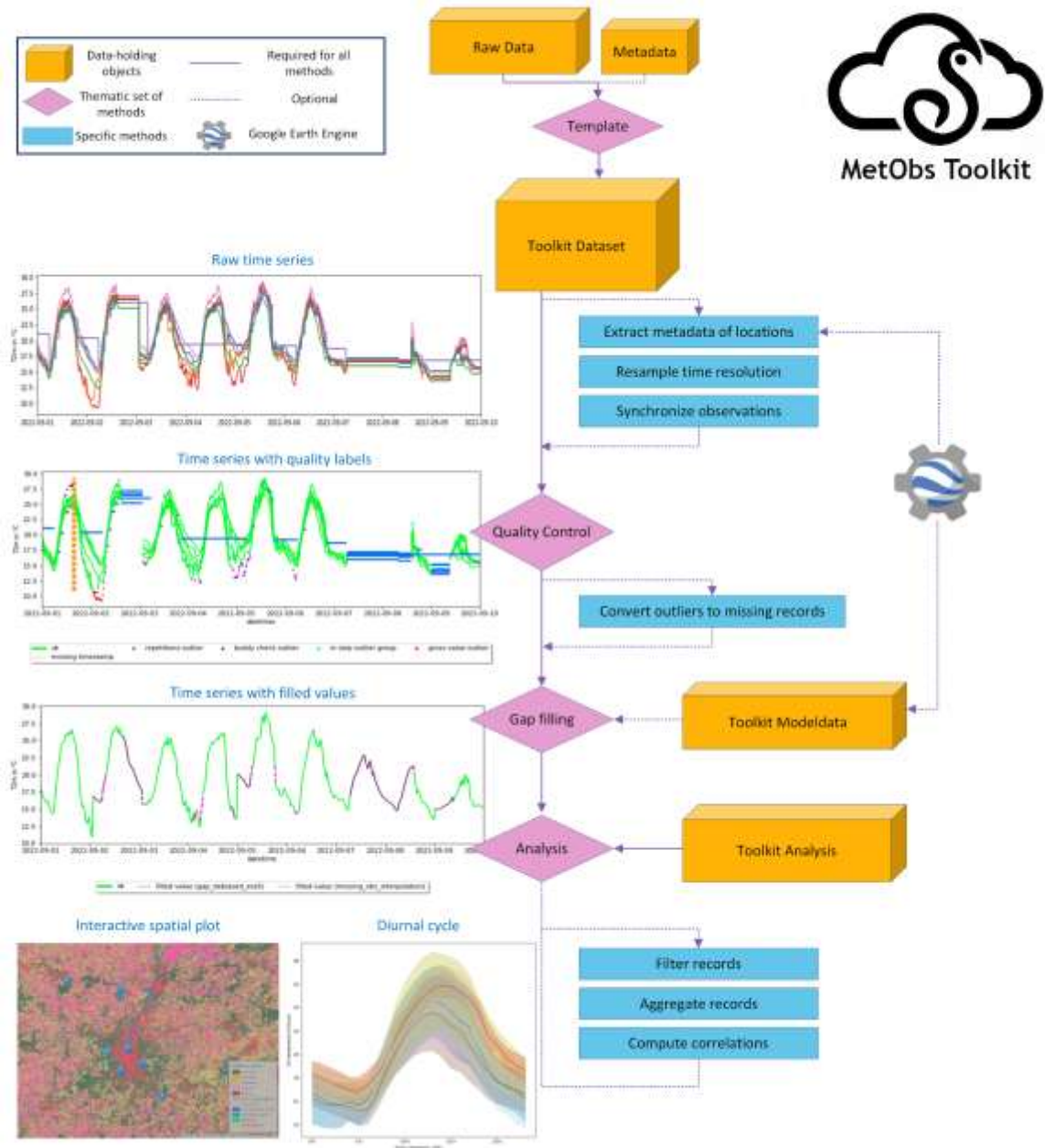


Figure 8: Overview of the MetObs-toolkit framework showing the data processing pipeline from raw data to analysis-ready dataset.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

Scientific results

Quality control innovations

The two QC innovations (priority-group processing and the LCZ safety net) were validated for hourly 2 m temperature observations over Belgium during the August 2020 heatwave. The improved pipeline rejects a comparable total number of observations to the conventional approach, but produces qualitatively superior results:

- Observations in unique microenvironments (water bodies, city centers, forests,...) are correctly retained rather than incorrectly flagged as outliers. In the test case of the Eeklo-Zelzate region, VLINDER station 25 (located in the middle of a large drinking water basin, LCZ G - Water) and VLINDER station 24 (inside a small forest, LCZ A - Dense trees) were consistently misclassified as outliers by the conventional pipeline and correctly retained by the improved version. VLINDER 25, as a gatekeeper from priority group 2, is retained by the priority-group mechanism, while VLINDER 24 is saved by the *LCZ-agg-safety net*. Its specific LCZ-A class lacks sufficient buddies, but the aggregated vegetation category results in sufficient buddies.
- As shown in Figure 9 the nocturnal UHI intensity and its spatial pattern in Mechelen is substantially better preserved by the novel pipeline, with a nocturnal UHI intensity of 4.5°C compared to 3°C for the conventional pipeline, and clearer LCZ-dependent stratification. In Antwerp, with its higher station density, both pipelines produce nearly identical city-scale UHI statistics; the novel pipeline adds value only at the level of individual stations in distinct microenvironments (e.g., urban parks), increasing local spatial variation.
- Regional biases introduced by the numerical dominance of low-quality CS stations are reduced, because the iterative background approach prevents these stations from anchoring the spatial statistics.

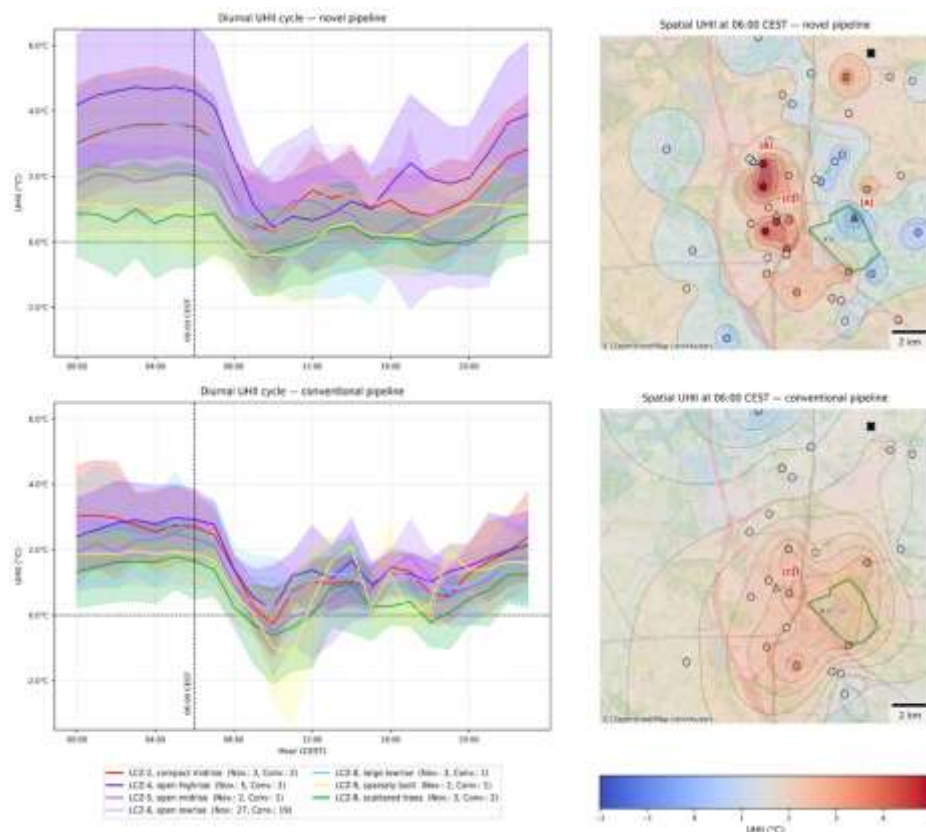


Figure 9: Diurnal UHI intensity per LCZ class for Mechelen (heat wave 5–16 August 2020), comparison of improved (top) vs. conventional QC pipeline (bottom); spatial Kriging interpolation of mean nocturnal UHI intensity at 06:00 UTC, showing a more realistic picture with stronger gradients

An important insight is that no single QC configuration is universally optimal: applications prioritising measurement quality (e.g., data assimilation) may accept more stringent rejection, while urban climate research benefits from retaining microenvironmentally distinct observations even at somewhat higher noise levels.

Hectometric NWP over Belgium

The hectometric AROME-SURFEX experiment demonstrated that 700 m NWP over Belgium is technically feasible within the ACCORD modelling framework. A few key outcomes:

- Figure 10 confirms that the model shows a pronounced diurnal UHI pattern with a clear nocturnal UHI. UHI intensities in Ghent reach up to 5°C on average during the heatwave period, with spatial patterns following the urban morphology. A cold daytime anomaly associated with a large drinking water basin approximately 11 km north of the city centre is correctly reproduced
- Comparable UHI patterns were found for Brussels and Sint-Niklaas.
- Traditional verification against synoptic stations shows biases of the order of 1°C and RMSE values in the range of 1.5–2.5°C (Figure 11). These scores clearly indicate the need for model tuning when operating at non-standard spatial scales.

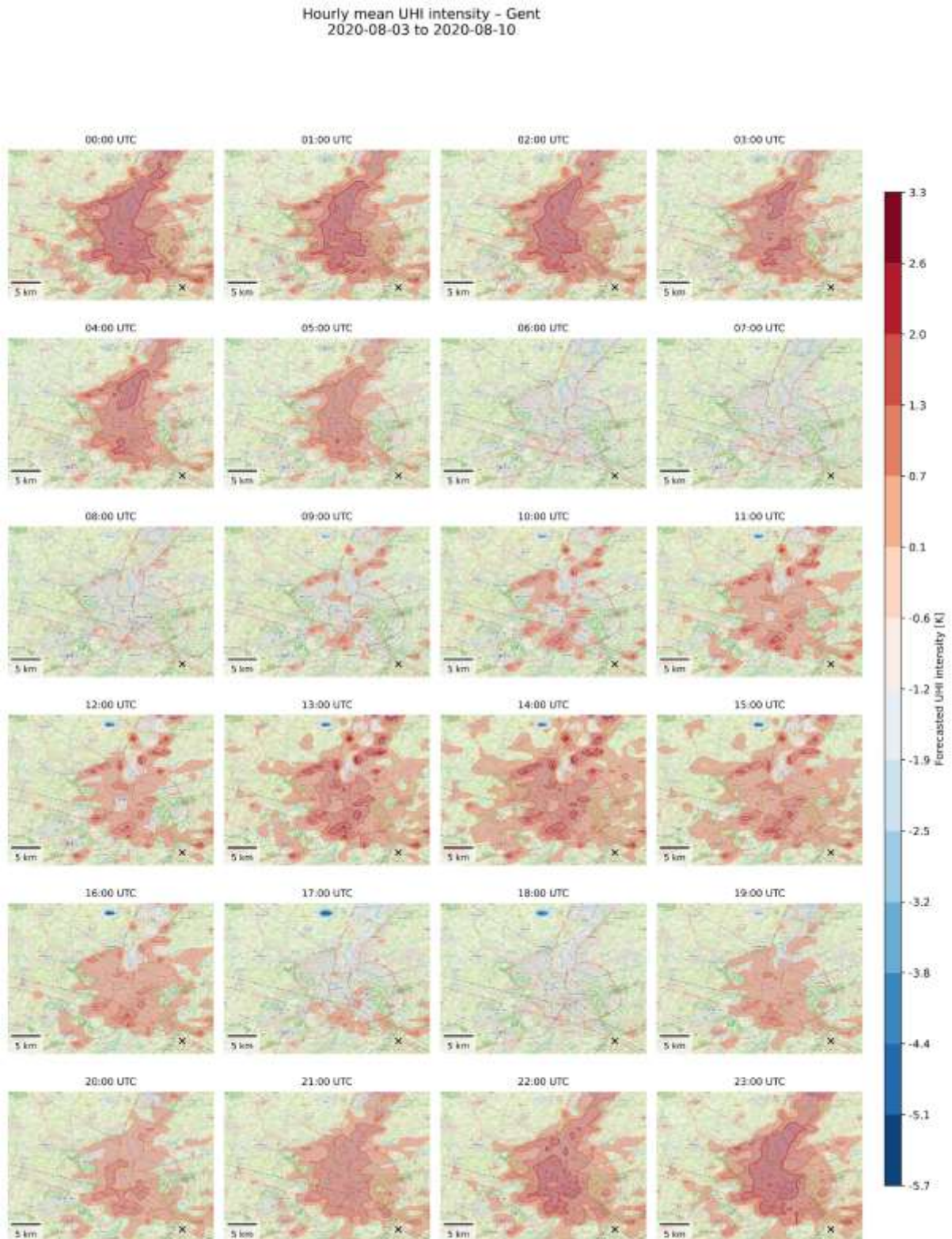


Figure 10: Average forecasted UHI intensity for Ghent per diurnal hour. The UHI intensity is the difference between a grid point's 2 m temperature and the ISBA 2 m temperature at the rural reference point (marked by X). Period: 3 - 10 August 2020, selected for favourable UHI conditions.

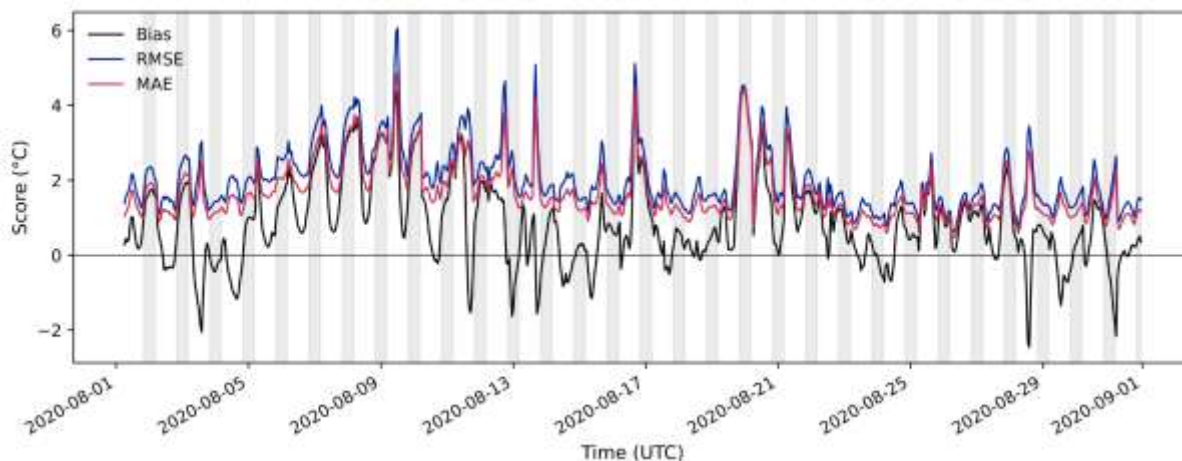


Figure 11: Temporal evolution of average bias, MAE, and RMSE for 2m temperatures from the 700m AROME run, using all observations that passed the QC procedure. Scores are computed by averaging over all observational stations for each day and hour in the period 1 - 31 August 2020.

As explained earlier, the broader goal of demonstrating added value of 700 m over 1.3 km resolution was not pursued with the same intensity as originally planned, given that DestinE was already investing heavily in this direction. The contribution of CS-MASK to hectometric NWP should therefore be understood as a Belgian demonstration study, rather than a contribution to model development.

Verification of high-resolution NWP with non-traditional observations

Several contributions were made to the verification methodology. We present some key results:

Added observational value: Including non-traditional observations for verification significantly increases spatial coverage, particularly in urban and peri-urban environments that are systematically underrepresented in synoptic networks. In the CS-MASK domain, the density of verification points increased by more than an order of magnitude when non-traditional observations were included. This enables verification of urban-specific model aspects (UHI, cool island effects) that remain invisible to synoptic-only verification.

Land cover diversity in verification: The size and land cover diversity of the non-traditional observational dataset enables a form of grouped verification that is entirely out of reach with synoptic stations alone: evaluation of model performance stratified by Local Climate Zone (LCZ) classes. By computing temperature anomalies relative to a common synoptic reference station and grouping results by the LCZ category of each observing site, it is possible to assess whether the model correctly captures the systematic influence of land cover on near-surface temperature. Applied to the 700 m AROME-SURFEX experiment over Belgium for August 2020, Figure 12 reveals that the nocturnal stratification of temperature by LCZ class is well reproduced: urban classes exhibit higher temperatures than rural classes during the night, with inter-class differences that closely match the observed magnitudes, indicating that the model representation of systematic land cover effects (primarily through the Town Energy Balance scheme) is adequate on average.

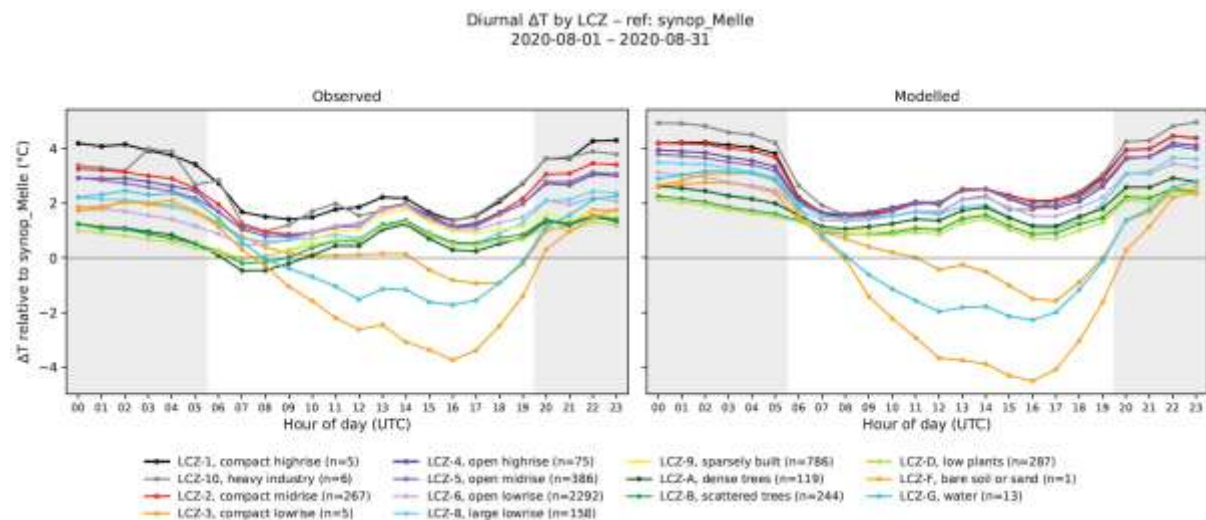


Figure 12: Diurnal cycle of temperature differences relative to the reference observations at the synoptic station in Melle, grouped by LCZ class based on station location. All observations that passed the QC pipeline are included. The left panel shows temperature differences from observations; the right panel shows the corresponding temperature differences from AROME-SURFEX (700m) output at the station locations.

Many-to-one regime: As observation density increases, the many-to-one verification regime becomes the norm. In this regime, error-based scores such as RMSE are affected by subgrid temperature variability that the model cannot resolve, causing artificially inflated errors even for a perfect model. The Z^2 _SBS score, which tests consistency with the observed subgrid distribution, performs robustly across both regimes and provides a more informative measure of forecast quality when dense networks are used. In extreme cases (thousands of observations per grid cell), traditional error-based scores become essentially uninterpretable, while sample-based scores retain their discriminative capacity.

MetObs-toolkit

The MetObs-toolkit was developed and described in a peer-reviewed software paper in the Journal of Open Source Software (Vergauwen et al., 2024). Since this publication and its release, it has been adopted by researchers in multiple European countries. The toolkit originated from the practical framework at the COST FAIRNESS summer school (Ghent, 2023), where participants from across Europe provided feedback that substantially shaped its further development. It has since grown into a community tool with multiple contributors and is actively maintained.

Other outcomes of the project

Beyond the primary scientific outputs, CS-MASK generated significant side results:

PhD training: Thomas Vergauwen followed a PhD trajectory, acquiring expertise in meteorological data processing, automated quality control, hectometric NWP, and verification methodology, a set of skills directly relevant to RMI's operational mission. PhD defense at Ghent University is planned by end of 2026.

Supervised student research: Several master and bachelor theses at Ghent University were co-supervised under CS-MASK, resulting in peer-reviewed international publications and conference presentations related to the use of non-traditional meteorological data for urban climate. A land-

cover-aware temperature bias correction for mobile urban measurements was developed (Vieijra et al., 2023) and the MetObs toolkit was used to investigate novel gap filling techniques designed for urban climate datasets (Jacobs et al., 2024).

Open-source infrastructure: The MetObs-toolkit (Vergauwen et al., 2024) constitutes a reusable, community-maintained software asset for any research group working with non-traditional meteorological observations. Its adoption extends well beyond the CS-MASK consortium.

International collaboration: CS-MASK strengthened existing ties with the ACCORD NWP consortium (VHR-NWP group), the COST FAIRNESS Action (CA20108), and Met Éireann. At the same time some of the research done in CS-MASK was complementary to the ongoing Destination Earth EC Flagship project. In this way the project supported the position of RMI in several important international research collaborations.

Educational material: Online learning paths and heat-stress workshops were developed for secondary schools across Belgium, using VLINDER observations as a pedagogical resource for urban meteorology and data literacy. Additionally, a wide set of examples on working with non-traditional observations and the MetObs-toolkit is provided and maintained on the corresponding documentation website.

RECOMMENDATIONS

Further research recommendations

CS-MASK addressed several research questions but, like always, many other challenges remain open. We present a selection of interesting follow-up activities.

The QC results presented are specific to hourly 2 m temperature observations over Belgium during August 2020. The generalisability of the LCZ-informed safety net approach to other variables (e.g., precipitation, wind speed), other seasons, or other geographic domains has not been formally tested. The effectiveness of the safety nets is contingent on the distinct interactions between the local landcover and the variable of interest. The priority-group assignment is based on an a priori network ranking, not on individual station quality assessments, which introduces some degree of subjectivity.

For the hectometric NWP results, the verification was limited to a single month and runs over Belgium. As such systematic biases cannot be distinguished from episodic errors without a multi-year evaluation over a larger region. The 700 m AROME-SURFEX model has not been tuned for operational use and conclusions drawn from absolute verification scores should therefore be interpreted with caution. The sample-based verification score framework is theoretically motivated and validated through a perfect-model setup, but its discriminative power in operational settings with real model errors remains to be demonstrated across multiple case studies.

It could be of interest to address the previous issues in a framework related to the hectometric runs done in Destination Earth.

Contribution to federal competencies and recommendations for decision support

As the national meteorological service for Belgium the RMI has responsibility for operational NWP, climate monitoring, and early-warning services in Belgium. The results of CS-MASK contribute indirectly to these responsibilities in several ways:

Operational quality control: Dedicated quality control will be a necessary condition if the RMI would decide to integrate CS meteorological observations into its operational chain (e.g., data assimilation, model verification, observational products, ...). The expertise gained during CS-MASK (e.g. the priority-group and LCZ safety-net innovations) are of significant importance in this context. Integrating heterogeneous non-traditional observations into operational workflows requires precisely the type of robust, metadata-aware quality control developed here. Both the methodology and the implementation are now available as open-source code within the MetObs-toolkit.

Urban climate monitoring: The VLINDER citizen science network, maintained partly through municipality partnerships, provides the dense urban coverage that synoptic networks cannot offer. The CS-MASK QC framework enables this network to be used with greater confidence for urban heat island monitoring, urban climate research at RMI and other institutes and services of direct relevance to public health authorities and urban planners operating under federal and regional mandates.

Hectometric NWP readiness: The first demonstration of 700 m AROME-SURFEX feasibility over Belgium, combined with the verification methodology developed to evaluate such runs, strengthens the position of the RMI to contribute meaningfully to DestinE and other international initiatives paving the way for hectometric NWP.

Based on CS-MASK a few recommendations w.r.t. hectometric NWP and CS data can be made:

One should consider setting-up a multi-network observation ingestion pipeline with the LCZ-informed QC developed here as the quality gate, enabling these data to become available for operational purposes. Having at least some reliable, high-quality observations within cities is crucial for an effective QC of PWS observations in urban environments.

Urban policies targeting heat resilience should be underpinned by harmonised observational datasets combining modelled and measured UHI information. The methods developed in CS-MASK provide a reproducible protocol for generating such datasets.

The verification of future high-resolution NWP systems should adopt sample-based score frameworks alongside classical error-based metrics, particularly when non-traditional observation networks are used, to avoid artefacts introduced by the many-to-one regime.

Leverage the CS-MASK study and verification framework for RMI's contribution to DestinE.

Continue maintaining the MetObs-toolkit as shared European infrastructure for non-traditional meteorological observations. New innovations in QC and gap-filling can rapidly reach international applications via the MetObs-toolkit.

5. DISSEMINATION AND VALORISATION

Scientific dissemination

The main scientific results were disseminated through peer-reviewed publications (see Section 6), conference presentations, and contributions to international workshops. Thomas Vergauwen presented at COST (FAIRNESS) workgroup meetings, the 12th International Conference on Urban Climate (ICUC12, 2025), and contributed to the organisation of the VHR-NWP ACCORD workshop in Norrköping (2023), which included a practical session using the Belgian hectometric setup.

The MetObs-toolkit was presented at multiple conferences, and during the COST FAIRNESS summer school (Ghent, 2023) it was used as the practical framework, directly motivating beta-testing by more than 30 early-career researchers from across Europe. The toolkit is distributed via PyPI and Conda-Forge and is documented on Read the Docs, with a community forum for issue tracking and feature requests on GitHub.

Thomas Vergauwen was one of the authors of the fourth chapter (Quality Control and Recovery of Meteorological Data) in the Springer handbook on Micrometeorological Measurements (ISBN 978-3-032-03883-8). Here, the MetObs-toolkit is presented to support good practice when processing meteorological observations.

Valorisation towards stakeholders

The project contributed to the RMI workshop "How do we make cities resilient against heat waves?" (2024), which brought together over 50 scientists, urban policy makers, and crisis managers of major Belgian cities. The CS-MASK team presented urban heat island maps for Belgian cities derived from the hectometric model experiment and the quality-controlled observation networks.

Education and training

Thomas Vergauwen developed and maintains online educational material (VLINDER learning paths) available to secondary school teachers and pupils across Belgium, covering urban meteorology, temperature analysis, and data science skills using real VLINDER observations. Heat stress workshops were organised for secondary schools, combining theoretical and practical components. Several master and bachelor theses at Ghent University were co-supervised within the framework of CS-MASK, resulting in peer-reviewed publications (Vieijra et al., 2023; Jacobs et al., 2024).

A wide set of examples and exercises related to the use and analysis of meteorological observations is documented on the MetObs-toolkit website.

International collaboration

CS-MASK strengthened existing collaborations with the ACCORD NWP consortium (through the VHR-NWP group), the COST FAIRNESS Action (CA20108), and Met Éireann (scientific visit for machine-learning-based improvement of ECOCLIMAP-SG physiography). The MetObs-toolkit serves as a shared infrastructure for multiple European research groups working on non-traditional meteorological observations.

6. PUBLICATIONS

Peer-reviewed publications (first author)

Vergauwen, T., Vieijra, M., Covaci, A., Jacobs, A., Top, S., Dewettinck, W., Vandelanotte, K., Hellebosch, I., and Caluwaerts, S. (2024). MetObs – a Python toolkit for using non-traditional meteorological observations. *Journal of Open Source Software*, 9(95), 5916.

<https://doi.org/10.21105/joss.05916>

Vergauwen, T., Top, S., Degrauwe, D., Termonia, P., and Caluwaerts, Retaining Microclimate Signals Trough Gatekeeper-Based and LCZ-Informed Quality Control of Personal Weather Station Data. (in preparation)

Peer-reviewed publications (co-author)

Vieijra, M., Vergauwen, T., Top, S., Hamdi, R., and Caluwaerts, S. (2023). Land cover aware temperature correction of bicycle transects: A case study of mapping the air temperature in two Belgian cities. *Urban Climate*, 50, 101578.

Jacobs, A., Top, S., Vergauwen, T., Suomi, J., Käyhkö, J., and Caluwaerts, S. (2024). Filling gaps in urban temperature observations by debiasing ERA5 reanalysis data. *Urban Climate*, 58, 102226.

Lalic, B., Stapleton, A., Vergauwen, T., Caluwaerts, S., Eichelmann, E., and Roantree, M. (2024). A comparative analysis of machine learning approaches to gap filling meteorological datasets. *Environmental Earth Sciences*, 83(24), 679.

Vandenalotte, K., Du Moulin B., Puynen, S., Vergauwen, T., Top, S., Caluwaerts, S., Too hot to race? Demonstrating heat stress modelling as an adaptation tool for endurance events. 2026. (Under review)

Book chapter (co-author)

Petrić, M., Lalic, B., Domonkos, P., Weidinger, T., Vergauwen, T., Koči, I., and Foken, T. (2026). Quality Control and Recovery of Meteorological Data. In: *Micrometeorological Measurements: An Introduction for Beginners*. Springer Nature Switzerland, pp. 131–161.

Conference contributions

Vergauwen, T., Top, S., Jacobs, A., Covaci, A., Dewettinck, W., Vandelanotte, K., Hellebosch, I., and Caluwaerts, S. (2024). Fast analysis of urban meteorological observations with the user-friendly MetObs-toolkit. EGU General Assembly 2024, doi:10.5194/egusphere-egu24-3829.

Vergauwen, T., Vieijra, M., Covaci, A., Jacobs, A., Top, S., Dewettinck, W., Vandelanotte, K., Hellebosch, I., and Caluwaerts, S. (2025). MetObs, to streamline your crowdsourced data processing. 12th International Conference on Urban Climate (ICUC12), doi:10.5194/icuc12-361.

7. ACKNOWLEDGEMENTS

This project was funded by BELSPO under the BRAIN-BE 2.0 programme (contract B2/202/P1/CS-MASK). The research was carried out at the Royal Meteorological Institute of Belgium (RMI) during a PhD-trajectory together with Ghent University.

The research benefited from the COST Action CA20108 (FAIRNESS) and from computational resources provided by RMI and the HPC infrastructure in Bologna (ECMWF). We thank the ACCORD community and the VHR-NWP working group for scientific exchanges on hectometric NWP. We also thank all contributors to the MetObs-toolkit, and the participants of the COST FAIRNESS summer school (Ghent, 2023) for beta-testing and feedback.

We would like to thank all the follow-up committee members for their valuable input and remarks: Valery Masson (CNRM), Patrick Samuelsson (SHMI), Xiaohua Yang (DMI), Mohamed Dahoui (ECMWF), Johan Brouwers (VMM), Nathalie Stephenne (SPW), Veronique Verbeke (Env Brussels), Griet Lambrechts (Stad Antwerpen).