

# CS-MASK Project

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

BRAIN-be 2.0 — Belgian Science Policy (BELSPO)

Royal Meteorological Institute of Belgium (RMI)

## Context, Objectives and Methodology

Sub-kilometric numerical weather prediction (NWP) allows urban morphology, park cooling and sea-breeze dynamics to be modelled explicitly — effects that remain invisible at coarser resolutions. The RMI operates one of Europe's most advanced operational NWP systems (AROME at 1.3 km), yet WMO-compliant synoptic stations at open, rural locations are not designed to evaluate models at sub-kilometric scales. Non-traditional, crowdsourced (CS) observations — personal weather stations (PWS), citizen-science networks and IoT sensors — provide high spatial density in precisely the urban environments avoided by synoptic networks, but require robust automated quality control (QC) before scientific use.

The CS-MASK project was carried out as a four-year PhD trajectory for Thomas Vergauwen (RMI / Ghent University), funded by BELSPO under BRAIN-be 2.0. Three work packages were defined: (WP1) construction and QC of a CS database for Belgium (August 2020); (WP2) hectometric AROME-SURFEX runs at 700 m and 1 400 m; and (WP3) evaluation of the model runs against the quality-controlled observations. Following the launch of the European Destination Earth (DestinE) initiative — which already targets pan-European hectometric NWP — resources were strategically redirected from optimising urban parameterisations towards QC innovation, open-source tooling (MetObs-toolkit) and a new verification framework for dense, heterogeneous observation networks.

**Observational dataset.** Ten networks were assembled for Belgium for August 2020, chosen for its meteorological richness: a 12-day heat wave, intense nocturnal urban heat islands (UHI), sea-breeze events and convective showers. The dataset combines 29 WMO synoptic stations with ~4600 non-traditional stations, regional monitoring networks (~250) and Netatmo PWS (~4000). Local Climate Zones (LCZ) were assigned per station to categorise the local environment.

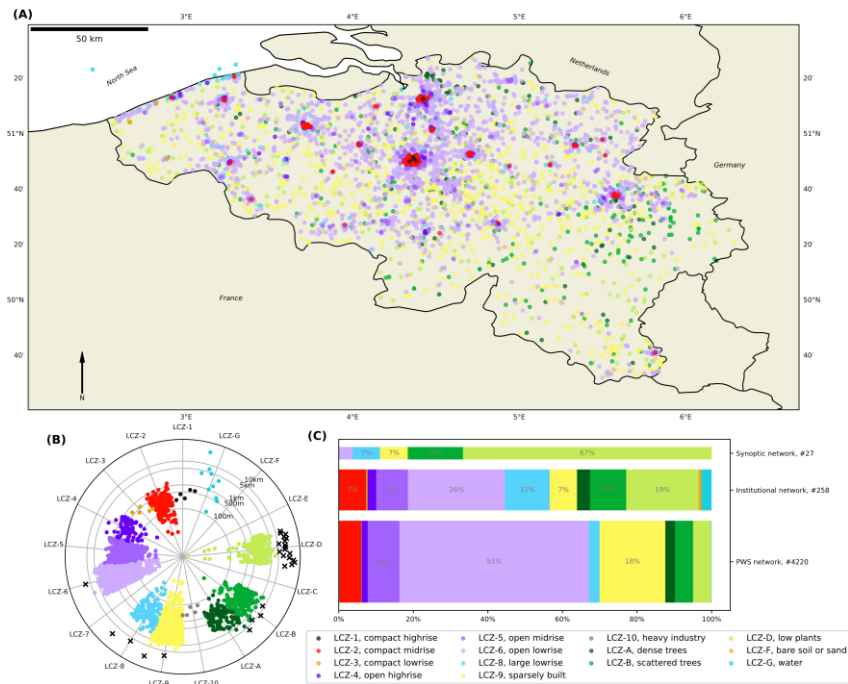


Figure 1: Spatial distribution of all stations with LCZ classification; radar plot of LCZ diversity per network group.

**Quality control.** Two innovations were built on top of traditional QC algorithms: (1) *Priority groups* are used to execute quality control iteratively; higher-priority observations (synoptic, then monitoring networks) form a non-rejectable "gatekeeper" background that prevents CS biases from corrupting QC statistics for lower-priority data. (2) *LCZ safety nets* re-evaluate observations that deviate from neighbouring observations but are plausible when the effect of the local environment is taken into account.

**Hectometric NWP.** For the first time, AROME-SURFEX was run at 700 m over Belgium for August 2020, with the Town Energy Balance (TEB) scheme active for urban environments.

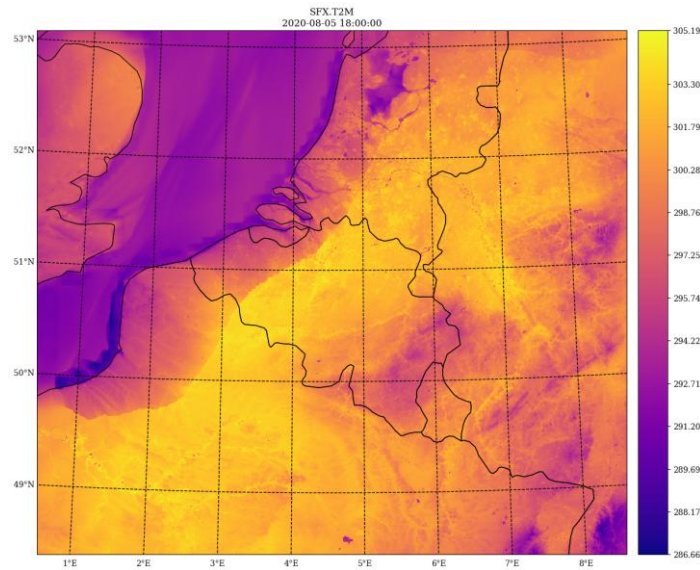


Figure 2: 2 m temperature field at 18 UTC on 5 August 2020 from the 700 m AROME-SURFEX run, showing the sea-breeze front.

**Verification.** In addition to standard error scores (bias, RMSE, MAE) applied to synoptic and non-traditional observations, a new *sample-based score* ( $Z^2\_SBS$ ) was developed for the many-to-one verification regime — in which multiple observations fall within a single model grid point — a situation that becomes the norm when using crowdsourced observations and personal weather stations.

**MetObs-toolkit.** An open-source Python package covering the full workflow from raw, heterogeneous data input to analysis-ready output: synchronisation, automated QC, gap-filling, metadata enrichment and analysis tools with Google Earth Engine integration.

## Results, Conclusions and Recommendations

**Quality control.** The improved pipeline has a comparable total number of rejections to the conventional approach, but the quality of the retained dataset is considerably better. Two key insights emerge from the case study. First, the new pipeline is deliberately more conservative in rejecting observations: every rejection made by the new pipeline is also made by the conventional one, but not vice versa. Second, the added value is strongly dependent on observation density. In regions with low station density, such as Mechelen, the new pipeline produces a substantially larger maximum nocturnal UHI intensity (4.5°C versus 3°C for the conventional approach) and clearer stratification per LCZ class. Stations in unique micro-environments — such as a park (LCZ-B), a school green space (LCZ-6) and a compact city centre (LCZ-2) — are correctly retained rather than erroneously rejected, resulting in sharper spatial temperature gradients. In regions with high station density, such as Antwerp, there are no significant

differences in the temporal evolution or the diurnal UHI cycle per LCZ class: the new pipeline does not degrade performance where conventional QC already works well.

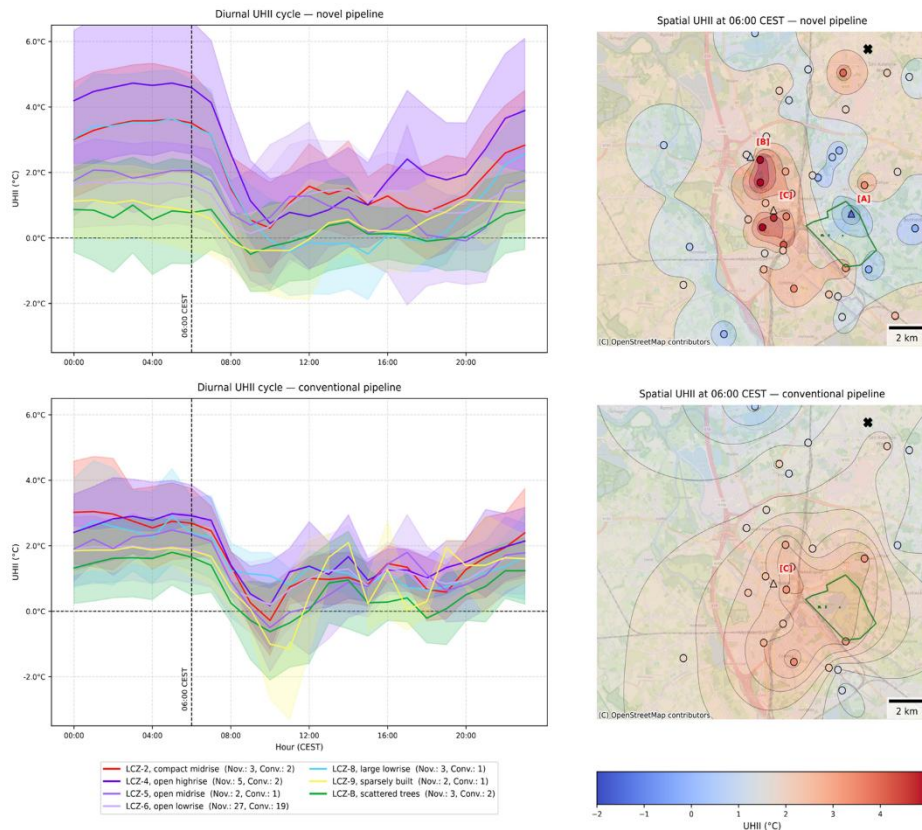


Figure 3: Diurnal UHI intensity per LCZ class for Mechelen (heat wave 5–16 August 2020), comparison of improved (top) vs. conventional QC pipeline (bottom); spatial Kriging interpolation of mean nocturnal UHI intensity at 06:00 UTC, showing a more realistic picture with stronger gradients.

**Hectometric NWP.** The 700 m experiment confirmed the technical feasibility of hectometric NWP over Belgium, yielding UHI intensity maps for Belgian cities and capturing sea-breeze and convective shower events. A systematic warm bias was identified for nearly all synoptic stations, demonstrating that sub-kilometric NWP requires careful model tuning. The CS-MASK contribution to hectometric NWP serves as a Belgian demonstration study providing a foundation for future DestinE contributions.

**Verification with non-traditional observations.** Including CS observations increases verification coverage by more than an order of magnitude. LCZ-stratified verification of the 700 m forecast shows that nocturnal temperature stratification per LCZ class is correctly reproduced — urban classes are warmer than rural classes, with intensities consistent with the observations — indicating adequate mean performance of the TEB scheme.

**Sample-based verification scores.** In the many-to-one regime, RMSE and MAE exhibit three systematic problems: double counting of grid points, autocorrelation between error terms and unaccounted sub-grid variability, making them misleading at high observation density. The  $Z^2$ \_SBS score avoids these problems by testing, per grid point, whether the forecast value is consistent with the observed sub-grid distribution. In a perfect-model experiment with synthetic forecasts at 2–7.5 km resolution,  $Z^2$ \_SBS remains informative across all regimes, while error scores become uninterpretable at high observation densities.

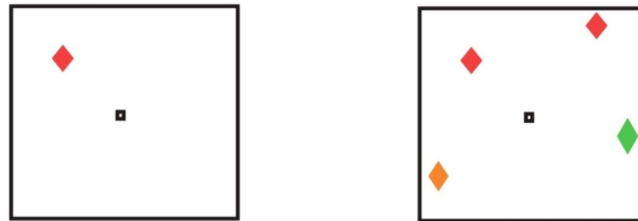


Figure 4: Illustration of one-to-one vs. many-to-one verification regime; RMSE, MAE and  $Z^2$ \_SBS as a function of forecast resolution and observation density.

**MetObs-toolkit.** Published in the *Journal of Open Source Software* (Vergauwen et al., 2024), adopted by groups across Europe and used as a practical framework at the COST FAIRNESS summer school (Ghent, 2023). The toolkit also forms the basis for a QC chapter in the *Springer Handbook on Micrometeorological Measurements* (2026).



Figure 5: Logo MetObs-toolkit

### Recommendations.

1. Integrate the priority-group and LCZ safety-net innovations into QC algorithms when local land-use effects are relevant and when crowdsourced or PWS observations are used.
2. Use  $Z^2$ \_SBS for verification when the observational density is substantially greater than the spatial resolution of the model.
3. Leverage the CS-MASK demonstration study and verification framework for the RMI's contribution to DestinE.

4. 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.

**Keywords:** crowdsourced observations · quality control · hectometric NWP · urban heat island · model verification