The Journal of Open Source Software

MetObs - a Python toolkit for using non-traditional meteorological observations

Thomas Vergauwen[1,2], **Michiel Vieijra**[1], **Andrei Covaci**[3], **Amber Jacobs**[2], **Sara Top**[2], **Wout Dewettinck**[2], **Kobe Vandelanotte**[1,2], **Ian Hellebosch**[2,4], and **Steven Caluwaerts**[1,2]

1 Royal Meteorological Institute of Belgium, Brussels, Belgium **2** Ghent University department of Physics and Astronomy, Ghent, Belgium **3** Vrije Universiteit Brussel (VUB), Brussels, Belgium **4** VITO, Mol, Belgium

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

In-situ meteorological observations are highly important for weather and climate research. The evolution towards more affordable sensor technology and data communication has resulted in the emergence of novel meteorological networks alongside the traditional high-quality measurement networks of meteorological institutions. Examples include urban measurement networks intended to study the impact of cities (Caluwaerts et al., 2020) and networks consisting of devices of weather enthusiasts (Muller et al., 2015). However, exploiting the data of such non-traditional networks comes with significant challenges (Meier et al., 2017). Firstly, sensors and data communication protocols are usually low-cost, and this in general results in an increase of measurement errors, biases and data gaps. Secondly, data storage formats and temporal measurement frequencies are often not consistent or compatible. Finally, metadata, such as land use around a station and elevation, are not easily accessible or documented.

The MetObs-toolkit is a Python package developed to address these issues and facilitate the use of non-traditional observations. The package provides automated quality control (QC) techniques to identify and flag erroneous observations, and includes methods to fill data gaps. Additionally, the package offers tools for analyzing the data, e.g. linkage with popular land-use datasets (Demuzere et al., 2022; Zanaga et al., 2022) is included such that microclimate effects can be investigated with the MetObs-toolkit.

Statement of need

The primary objective of the MetObs-toolkit is to enable scientists to process meteorological observations into datasets ready for analysis. The data cleaning process involves three steps:

1. resampling the time resolution if necessary,
2. identifying erroneous and missing records, and
3. filling the missing records.

Sophisticated software such as TITAN (Båserud et al., 2020) and CrowdQC+ (Fenner et al., 2021) exists for identifying erroneous observations (QC), which is one aspect of cleaning a dataset. These packages offer a wide range of functionalities for this specific task, while MetObs aims to provide a framework for the entire flow from raw data to analysis. Moreover, researchers often face the challenge of coding scripts that can generate analyses, particularly when using geographical datasets such as landcover datasets. Traditionally, this requires the installation of numerous packages, storage of geographical datasets, and GIS manipulations (often manually done with specific GIS software). The toolkit implements one user-friendly framework for creating various plots, generating analysis statistics, and incorporating GIS data through the use of the Google Earth engine. By using the toolkit, scientists can set up a pipeline to process raw data into analysis in an easy-to-use (and install) manner. Additionally, the developed pipeline can be directly applied to other datasets without any formatting issues.

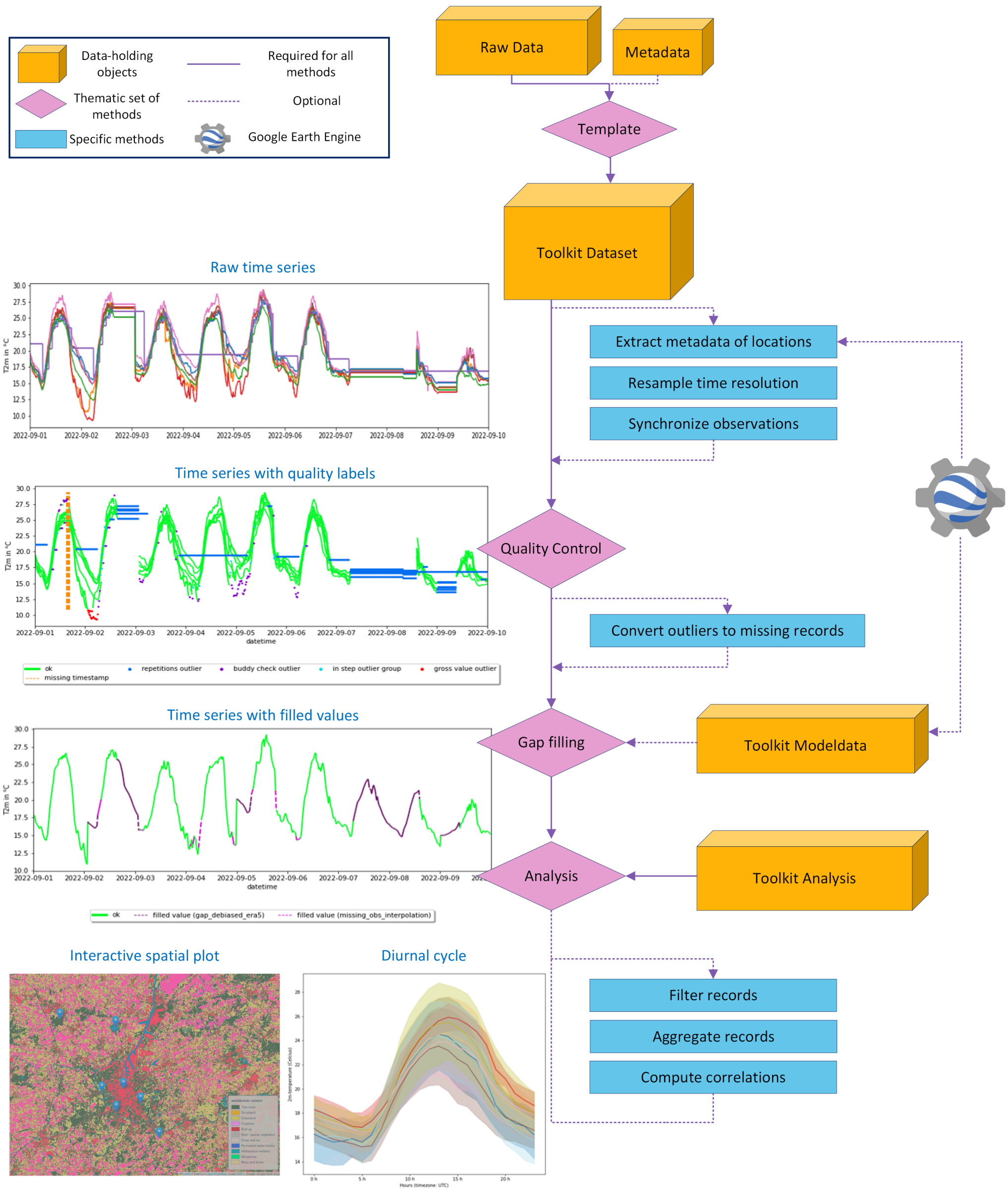

Figure 1: A schematic overview of the main MetObs-toolkit functionalities.

Technical implementation

The MetObs-toolkit provides a comprehensive framework for scientists to process raw meteorological data for analysis by making intensive use of the pandas (the pandas dev team, 2023) and geopandas (Jordahl et al., 2020) functionalities. The process consists of the following steps, visualized in the Figure 1.

Firstly, the raw data is mapped to the toolkit standards by use of a template. Once the raw data is imported into the Toolkit Dataset, missing observations are identified and methods to resample and synchronize observations can be used.

Quality control is performed in the form of a series of checks. These checks are designed to examine data types, irregular timestamps, max-min thresholds, repetitions criteria, spike tests, allowed variation in time windows and spatial tests. Advanced quality control methods are available through the implementation of TITAN into the toolkit. The user can choose to keep the outliers or convert them to missing records (which can be filled).

Gap filling is applied by using interpolation methods and/or importing ERA5 reanalysis (Hersbach et al., 2020) time series to fill the gaps. The latter is stored as a Toolkit Modeldata, which has a set of methods to directly import the required time series through the use of the Google Earth engine API. The user obtains a cleaned-up dataset ready for analysis. A set of typical analysis techniques such as filters, aggregation schemes, and landcover correlation estimates are implemented in the Toolkit-Analysis class.

Figure 1 gives an overview of the main framework of the MetObs-toolkit, but it is an evolving project that responds to the community's needs and input. As an example, the development of a graphical user interface (GUI) for the toolkit is planned. A GUI would increase the ease of use by enabling to create templates, adjust QC settings and plot data interactively.

Acknowledgments

The authors would like to thank all participants of the COST FAIRNESS (CA20108) summer school 2023 in Ghent for their role as beta testers. The input, ideas and feedback from these scientists, dealing with microclimate datasets in many European countries, were instrumental in improving the MetObs-toolkit.

No specific funding has been obtained to build the MetObs-toolkit, but the authors have been supported by different Belgian and Flemish scientific grants.

FWO: Sara (fellowship 1270723N) and Wout (fellowship 1157523N)

BELSPO: Kobe (B2/223/P1/CORDEX.be II), Thomas (B2/202/P1/CS-MASK), Michiel (B2/212/P2/CLIMPACTH) and Steven (FED-tWIN Prf-2020-018_AURA)

Andrei (VUB, SRP74/LSDS, OZR3893, Innoviris-Brussels ILSF-2023-12) and Ian (VITO, UG_PhD_2202)

References

Båserud, L., Lussana, C., Nipen, T. N., Seierstad, I. A., Oram, L., & Aspelien, T. (2020). TITAN automatic spatial quality control of meteorological in-situ observations. *Advances in Science and Research*, *17*, 153–163. https://doi.org/10.5194/asr-17-153-2020

Caluwaerts, S., Hamdi, R., Top, S., Lauwaet, D., Berckmans, J., Degrauwe, D., Dejonghe, H., De Ridder, K., De Troch, R., Duchêne, F., Maiheu, B., Van Ginderachter, M., Verdonck, M.-L., Vergauwen, T., Wauters, G., & Termonia, P. (2020). The urban climate of Ghent, Belgium : A case study combining a high-accuracy monitoring network with numerical simulations. *Urban Climate*, *31*, 19. https://doi.org/10.1016/j.uclim.2019.100565

Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I. D., Vliet, J. van, & Bechtel, B. (2022). *Global map of local climate zones* (Version 1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6364594

Fenner, D., Bechtel, B., Demuzere, M., Kittner, J., & Meier, F. (2021). CrowdQC+—a quality-control for crowdsourced air-temperature observations enabling world-wide urban climate applications. *Frontiers in Environmental Science*, *9*. https://doi.org/10.3389/fenvs.2021.720747

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., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Jordahl, K., Bossche, J. V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *Geopandas/geopandas: v0.8.1* (Version v0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Meier, F., Fenner, D., Grassmann, T., Otto, M., & Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, *19*, 170–191. https://doi.org/10.1016/j.uclim.2017.01.006

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., & Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *International Journal of Climatology*, *35*(11), 3185–3203. https://doi.org/10.1002/joc.4210

the pandas dev team. (2023). *Pandas-dev/pandas: pandas* (Version v2.0.0). Zenodo. https://doi.org/10.5281/zenodo.7794821

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., & Arino, O. (2022). *ESA WorldCover 10 m 2021 v200* (Version v200) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7254221