REQUEST FOR A SPECIAL PROJECT 2025–2027

MEMBER STATE: Portugal

Principal Investigator[1]**:** Ana Oliveira

Affiliation: CoLAB +ATLANTIC

Address: IPL, Rua do Conhecimento no. 4, Peniche, Portugal

Other researchers:

Bruno Marques, Caio Fonteles, Fabíola Silva, Inês Girão, João Paixão, Manvel Khudinyan, Maria Castro, Rita Cunha

Project Title:

ML4EUcities – Machine Learn for European cities

To make changes to an existing project please submit an amended version of the original form.)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** ………….	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2025	
Would you accept support for 1 year only, if necessary?	YES ■	NO ☐

Computer resources required for project year:		**2025**	**2026**	**2027**
High Performance Computing Facility	[SBU]			
Accumulated data storage (total archive volume)[2]	[GB]			

EWC resources required for project year:		**2025**	**2026**	**2027**
Number of vCPUs	[#]	35	35	35
Total memory	[GB]	350	350	350
Storage	[GB]	10000	15000	25000
Number of vGPUs[3]	[#]	2	2	2

Continue overleaf.

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.
[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.
[3]The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

Principal Investigator: Ana Oliveira

Project Title: ML4EUcities - Machine Learning for European cities

Extended abstract

Introduction

As climate change prospects point towards the pressing need for local adaptation strategies (IPCC, 2022), exposure to extreme weather events becomes one of the most important aspects in determining our society's resilience in the future (Sagan et al., 2022; UNDRR, 2023). Globally, we are already experiencing changing patterns of exposure to certain types of extremes (e.g., wildfires in high latitudes, droughts in midlatitudes, flash floods in riverine and coastal areas) (IPCC, 2022; Perkins et al, 2020); and, at the European level, recent historical weather measurements are already showing a changing climate, where heatwaves (HW) become longer, more frequent and intense, while cold waves (CW) show only minor or non-significant changes (EEA, 2018; Hooyberghs et al., 2019; C3S, 2022). This amplitude increase between the temperature extremes is a major challenge to our strongly urbanised (UN-HABITAT, 2022) but steadily ageing society (EUROSTAT, 2020), in several aspects: from excess mortality and hospital admissions in the most vulnerable population segments, to asymmetries in household's energy poverty versus unmet peak electricity demand for indoor acclimatisation (Ebi et al, 2021; WHO, 2021; ; Ballester et al., 2023;), both the health and energy sectors are deeply affected by weather conditions, particularly air and land surface temperatures which are, at the local level, strongly influenced by the energy exchanges between the lower atmosphere and our deeply artificialized urban surfaces - and is aspect is of utmost importance while measuring and analysing the specificities of what is our urban climate (Oke, 1983, Oke, 2017).

Within our cities, by changing the overall surface albedo through the introduction of darker and more energy absorbing materials while reducing natural vegetation - thermal properties changes - as well as through the introduction of 'urban canyons' - aerodynamic changes - our cities' trap more radiation received from the sun (during the day) compared to the countryside Oke, 2017). At the urban canopy layer level (UCL, i.e., the air volume beneath the average height of the buildings) this excess heat storage at the surface is later released back to the atmosphere as 'sensible heat flux' producing a positive urban 'anomaly', the so-called the Urban Heat Island (UHI) effect. Described since the last century, the UHI is one of the greatest by-products of the anthropogenic-induced changes over the landscape, by interfering with the underlying thermodynamic physical processes which can be quantified through the daily cycle of the Urban Energy Balance (UEB) components, which can be inferred from Earth Observation imagery (Bastiaanssen et al., 1997; Parlow, 2003; Zhou et al., 2019; Parlow, 2021). Hence, background synoptic/regional weather conditions are effectively modulated within our cities, which in turn has a mensurable impact on the accuracy of estimating short term and long-term temperature-related risks. And, from a risk assessment and response standpoint, this means that the level of detail and precision that we offer when mapping urban weather and climate-induced hazards is of utmost importance to reliably prioritise, in space and time, preventive and early warning actions, as well as when testing the alternative climate adaptation measures, something not currently offered in what concerns operational weather forecasting systems (Baklanov at al., 2018; Grimmond, 2020).

Problem Statement

Considering these overarching aspects, it is well established that there are sufficient societal drivers, i.e., an existing need, for advancing the level of detail and precision of the air and land surface temperature predictions by accounting for the daily cycle of the UHI effect, particularly to (1) anticipate the neighbourhoods most exposed to air temperature extremes, in the short-term, (2) simulate the cooling/heating effect of alternative urban development pathways in the future climate change scenarios, and (3) quantify the sensitivity of the sectoral impacts to the temperature extremes. In this regard, the Digital Twin City (DTC) concept, comes forth as an integrated user-driven solution towards situational awareness goals, by promoting the symbiose between in-situ sensors (connected in real-time via internet-of-things (IoT) technologies), remotely sensed data and Earth Observation (EO) imagery through more efficient data-driven impact prediction approaches - notoriously Artificial Intelligence (AI) and Machine Learning (ML). However, there is limited access to air temperature or land surface data in high-resolution (i.e., at a sub-kilometric scale) due to a lack of urban-resolved observational data (i.e., most cities do not have proper mesoscale weather networks installed, and the temporal scarcity of high-resolution satellite thermal observation acquisitions. As described in (Oke et al., 2017), there are four main methods to obtain city-level UHI data: (i) numerical and physical models; (ii) field observations; (iii) remote sensing/satellite observations; and (iv) empirical models. The same author highlights why options (i) to (iii) have serious constraints for the final users, as a stand-alone tool:

- firstly, numerical and physical models provide quasi-controlled experiments, which are accurate and very flexible to test scenarios, but their usage is typically restricted to academic research, i.e., not available for urban managers and decision makers (Oke et al., 2017), due to the heavy computational demand and expertise required to run them (let alone to put them into operational forecast);
- secondly, air temperature measurements are easy to interpret, even by non-specialists - these depict air temperature observations as discrete values in space (i.e., data points) - but high density weather networks are not available in most cities (Meier et al., 2015; Oke et al., 2017); while, some studies have been using citizen acquired data to fill in this need, this option implies proper handling of its quality limitations of this data, and is subject to non-homogeneous spatial densities (e.g., excluding outliers and correcting systematic positive bias) (Napoly et al., 2018; Nipen et al., 2020);
- and thirdly, while there have been many EO missions encompassing optical and thermal payloads (e.g., Landsat, Sentinel 2 and 3, MODIS, METEOSAT), their data is strongly dependent on the spatial versus temporal resolution nexus and atmospheric column visibility (optical depth, cloud coverage); for example, geostationary satellites provide hourly or sub-hourly data, which allows to assess the full daily cycle of the UHI, but lack the metric resolution required for urban applications, while sun-synchronous satellites offer single shots at a specific time of the day only, in some cases having several days between revisits.

Hence, when considering these methods separately, they fail to offer end-users an effective, cost-efficient, and easily interpretable data supply service tackling urban climate-dependent issues (Oke at al., 2017). This lack of appeal hinders their adoption by local communities and private entities interested in understanding the impacts of Urban Heat Islands (UHI) on spatial planning, public health, or energy management. Conversely, by employing Machine Learning and Artificial Intelligence algorithms to multidimensional collocated EO, numerical modelling and in-situ observations data, we can generate spatio-temporal data fusion models to predict downscaled synthetic versions of the key Essential Climate Variables (ECVs) describing the UHI effect. In doing so, these methods successfully overcome the previously mentioned limitations, meeting sector-specific continuity and resolution requirements, and enabling the generalisation of results elsewhere, at a much lower operational cost.

Scientific Plan

ML4EUcities aims to pioneer the development of Machine Learning (ML) and Artificial Intelligence (AI) models designed to downscale air and land surface temperature predictions in urban areas by a factor of at least ten. This initiative serves as a preliminary step towards the implementation of cost-effective Integrated Urban Climate and Weather components into local Digital Twin Systems. By leveraging crowdsourced data from the citizens, earth observation and weather forecasting models, we aim to offer spatio-temporal data fusion models that can solve the unmet need for a low-cost, efficient, and scalable Urban Climate prediction system. To achieve this, ML4EUcities aims to tailor its solution to the requirements of local early warning systems, and as a tool for evaluating climate adaptation measures, namely the impact of green infrastructures on the Urban Heat Island effect. ML4EUcities will provide a coupled ML-based near-surface Air Temperature (henceforth, T2m) and Land Surface Temperature (LST) downscaling system targeting several cities in Europe, proving the concept's reliability and scalability to further urban regions.

This goal is only attainable by using large sets of historical data as the target response variables, containing sufficient observations to depict the full statistical distribution of the T2m and LST variables, namely the extreme values of the time series. By training several AI/ML data fusion techniques, ML4EUcities will downscale Numerical Weather Predictions (NWP) of air temperature forecasts and spaceborne satellite-derived LST observations into a sub-kilometric (between 100 and 200m) grid of temporally-resolved T2m and LST predictions, using quality controlled weather data acquired consistently through the networks of citizen-owned stations, as well as high resolution spaceborne thermal imagery. To achieve this, ML4EUcities will leverage the framework from Lowry (1977) which states that a given weather element is the result of: (i) the 'background' regional climate (e.g., Mediterranean Climate), (ii) the effects of local landscape (e.g., topography, coastal proximity), and (iii) the effects of local urbanisation (e.g., impermeabilization, compactness). In addition, ML4EUcities will also consider the well-known concept of the UEB which allows to tie together the UHI and the SUHI, considering these diverse types of source data:

- the substantial number of existing citizen-owned weather stations across the main European (and global) metropolitan areas, continuously measuring our cities' temperatures and humidity - particularly, hundreds of NetAtmo and Lonobox weather stations are available in Denmark's metropolitan areas; these are also readily available in many urban regions over Europe, as well as globally, ensuring replicability of the modelling approaches;
- the comprehensive database of EO imagery depicting the most important determinants of our cities' UEB - while optical imagery is the input for accounting land cover classes, surface properties (e.g., albedo, emissivity, which are needed to compute the urban surface heat flux components), the thermal imagery is of utmost important also to verify the UEB and the UHI of the urban surfaces, the so-called SUHI which has a distinct daily cycle but is deeply coupled with the atmospheric conditions;
- the reliable and temporarily resolved synoptic weather forecasts from the national and European services - these provide a historical database of gridded data that resolves the background weather conditions (i.e., those that arise at global and synoptic scales) with much greater temporal detail (hourly intervals, or less), and physically

The overarching goal is to improve the level of detail offered by the existing urban climate (and weather) monitoring and prediction solutions, while keeping foreseen operational (i.e., computational) cost minimal, to make the case for a follow-up operational EO-based decision-support tool. Hence, ML4EUcities aims to provide higher resolution and temporally-resolved geospatial data to detect and characterise the UHI and SUHI in European cities, while ensuring that ML models respect the physical processes involved in the daily cycle of the UEB. In addition, ML4EUcities aims to highlight the added value of using downscaled geospatial information to anticipate climate risks arising from changing temperatures, particularly HW and CW, in support of knowledge-based decision making. The project goals can be further refined into specific goals (Sx), as follows:

- S1. To develop, implement, and validate an improved version of the existing methods for downscaling T2m and LST based on ML methods for data fusion, accounting for the methodological opportunities and shortcomings highlighted in the previous section and ensure a reliable representation of extreme values, namely during heat a cold temperature extremes;
- S2. To demonstrate the feasibility of blending EO remote sensing data with voluntary networks of in-situ weather measurements (e.g., from NetAtmo, Lonobox) and NWP model outputs (e.g., HARMONIE-AROME), through ML techniques that successfully downscale these ECVs and depict the UHI and SUHI, providing insights into the spatio-temporal urban climate response during HW and CW events;
- S3. To advance current scientific capacity to deliver knowledge-based operational tools to downscale operational weather predictions and simulate future climate change scenarios at a sub-kilometric (between 100 and 250m) spatial resolution, especially for early-warning and decision support during HW and CW occurrence, but also as a tool for assessing different urban planning climate adaptation measures.

Justification for the Resources Required

For each city, to attain spatio-temporal reliability of model performance, one is required to gather a significant amount of data. For instance, each FUA site usually covers an equivalent to 500.000 data points in space (at 0,002x0,002 grid resolution), which multiplies in time in an hourly basis. For comprehensive assessment, hourly data over multiple years is desirable. Hence, if considering 3-years-worth of data, up to 13.140.000.000 data points are generated, per city. Accordingly, we have been facing several challenges regarding our computational capacity. Therefore, we require a high-performance computing cluster with at least 350 vCPUs, 350 GB of RAM, 25 TB of storage and 2 vGPUs, and the possibility to be able to develop advanced ML software.

Technical characteristics of the Code

Data will be pre-processed using Python in Linux environment. Results will be validated against known benchmarks. The EDA shall be conducted in Python using Jupyter notebooks implementations, and standard libraries for climate and geospatial data processing. In terms of environment requirements, currently, the code is based on Python 3.9, runs on Linux, Ubuntu 20.04 or later, and the following libraries are required:

- xarray==2023.10.1
- scikit-learn==1.0.2
- matplotlib==3.8.0
- pandas==2.1.2
- absl-py==2.0.0
- geopandas==0.14.0
- earthpy==0.9.4
- tqdm==4.62.2
- cdo==1.6.0

Reference

Ballester, Joan, Marcos Quijal-Zamorano, Raúl Fernando Méndez Turrubiates, Ferran Pegenaute, François R. Herrmann, Jean Marie Robine, Xavier Basagaña, Cathryn Tonne, Josep M. Antó, and Hicham Achebak. 2023. “Heat-Related Mortality in Europe during the Summer of 2022.” Nature Medicine 29 (7): 1857–66. https://doi.org/10.1038/s41591-023 02419-z.

Baklanov, A., C. S.B. Grimmond, D. Carlson, D. Terblanche, X. Tang, V. Bouchet, B. Lee, G. Langendijk, R. K. Kolli, and A. Hovsepyan. 2018. “From Urban Meteorology, Climate and Environment Research to Integrated City Services.” Urban Climate 23 (March): 330–41. https://doi.org/10.1016/J.UCLIM.2017.05.004.

Bastiaanssen, W.G.M., Henk Pelgrum, Jiemin Wang, Yaoming Ma, J. F. Moreno Moreno, G. J. Roerink, and Tamme van der Wal. 1998. “A Remote Sensing Surface Energy Balance Algorithm for Land (SEBAL).” Journal of Hidrology, 198–212. https://www.researchgate.net/publication/245099273_A_remote_sensing_surface_energy_balance_algorithm_for_lan d_SEBAL.

C3S. 2022. “EUROPEAN STATE OF THE CLIMATE: SUMMARY 2022.”

Ebi, Kristie L., Anthony Capon, Peter Berry, Carolyn Broderick, Richard de Dear, George Havenith, Yasushi Honda, et al. 2021. “Hot Weather and Heat Extremes: Health Risks.” The Lancet 398 (10301): 698–708. https://doi.org/10.1016/S0140-6736(21)01208-3.

EEA. 2018. “Unequal Exposure and Unequal Impacts: Social Vulnerability to Air Pollution, Noise and Extreme Temperatures in Europe,” no. 22. https://doi.org/10.2800/324183.

EUROSTAT. 2020. “Ageing Europe LOOKING AT THE LIVES OF OLDER PEOPLE IN THE EU 2020 Edition.” Luxembourg. https://doi.org/10.2785/628105.

Grimmond, Sue, Veronique Bouchet, Luisa T. Molina, Alexander Baklanov, Jianguo Tan, K. Heinke Schlünzen, Gerald Mills, et al. 2020. “Integrated Urban Hydrometeorological, Climate and Environmental Services: Concept, Methodology and Key Messages.” Urban Climate 33 (September): 100623. https://doi.org/10.1016/J.UCLIM.2020.100623.

IPCC. 2022. “Climate Change 2022: Impacts, Adaptation and Vulnerability .” UK and New York. https://doi.org/10.1017/9781009325844.

Lowry, William P. 1977. “Empirical Estimation of Urban Effects on Climate: A Problem Analysis in: Journal of Applied Meteorology and Climatology Volume 16 Issue 2 (1977).” Journal of Applied Meteorology and

Climatology 16 (2): 129 35. https://journals.ametsoc.org/view/journals/apme/16/2/1520-0450_1977_016_0129_eeoueo_2_0_co_2.xml.

Napoly, Adrien, Tom Grassmann, Fred Meier, and Daniel Fenner. 2018. "Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data." Frontiers in Earth Science 6 (August): 382573. https://doi.org/10.3389/FEART.2018.00118/BIBTEX.

Nipen, Thomas N., Ivar A. Seierstad, Cristian Lussana, Jørn Kristiansen, and Øystein Hov. 2020. "Adopting Citizen Observations in Operational Weather Prediction." Bulletin of the American Meteorological Society 101 (1): E43–57. https://doi.org/10.1175/BAMS-D-18-0237.1.

Oke, T. R. 1987. Boundary Layer Climates. Vol. 2nd Edition. https://www.routledge.com/Boundary-Layer Climates/Oke/p/book/9780415043199.

Oke, Timothy R., Gerald Mills, Andreas Christen, and James A. Voogt. 2017. "Urban Climates." Urban Climates, September, 1–525. https://doi.org/10.1017/9781139016476.

Parlow, Eberhard. 2003. "The Urban Heat Budget Derived from Satellite Data." Geographica Helvetica 58 (2): 99–111. https://doi.org/10.5194/GH-58-99-2003.

Parlow, Eberhard, and Stefania Bonafoni. 2021. "Regarding Some Pitfalls in Urban Heat Island Studies Using Remote Sensing Technology." Remote Sensing 2021, Vol. 13, Page 3598 13 (18): 3598. https://doi.org/10.3390/RS13183598.

Perkins-Kirkpatrick, S. E., and S. C. Lewis. 2020. "Increasing Trends in Regional Heatwaves." Nature Communications 11 (1). https://doi.org/10.1038/s41467-020-16970-7.

Hooyberghs, H., J., Berckmans, F., Lefebre, and K. De Ridder. 2019. "Heat Waves and Cold Spells in Europe Derived from Climate Projections." 2019. https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-heat-and-cold-spells?tab=overview.

Sagan, Anna, Scott L Greer, Erin Webb, Martin Mckee, Natasha Azzopardi Muscat, Suszy Lessof, Isabel De La Mata, Josep Figueras, and Martin Mckee Is Co-Director. 2022. "STRENGTHENING HEALTH SYSTEM RESILIENCE IN THE COVID 19 ERA." Eurohealth. Vol. 28.

UNDRR. 2023. "GAR Special Report 2023: Mapping Resilience for the Sustainable Development Goals." United Nations Office for Disaster Risk Reduction. https://www.undrr.org/gar/gar2023-special-report.

UN-HABITAT. 2022. "Envisaging the Future of Cities."

WHO. 2021. "Heat and Health in the WHO European Region: Updated Evidence for Effective Prevention." 2021. https://apps.who.int/iris/handle/10665/339462.

Zhou, Decheng, Jingfeng Xiao, Stefania Bonafoni, Christian Berger, Kaveh Deilami, Yuyu Zhou, Steve Frolking, Rui Yao, Zhi Qiao, and José A. Sobrino. 2018. "Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives." Remote Sensing 2019, Vol. 11, Page 48 11 (1): 48. https://doi.org/10.3390/RS11010048.