The Alan Turing Institute

CLIM-RECAL: open recalibrated climate data repository

1. **Name of Theme Lead making submission:**

   Gavin Shaddick

2. **Names of other contributors and partner institutions if relevant:**

   Contributors:
   - Gavin Shaddick
   - Mark Birkin
   - Ruth Bowyer
   - Fernando Benitez-Paez
   - Jennifer Ding

   Partner Institutions:
   - The Met Office
   - The Joint Centre for Excellence in Environmental Intelligence (JCEEI)

3. **Name of theme or CTP:**

   Environment & Sustainability, Ecosystems of Digital Twins

4. **Title of project or initiative:**

   CLIM-RECAL: open recalibrated climate data repository

5. **Link to project website (if relevant):**

   - Turing project page: https://www.turing.ac.uk/research/research-projects/impacts-climate-change-and-heat-health
   - Open project repo: https://github.com/alan-turing-institute/dymechh

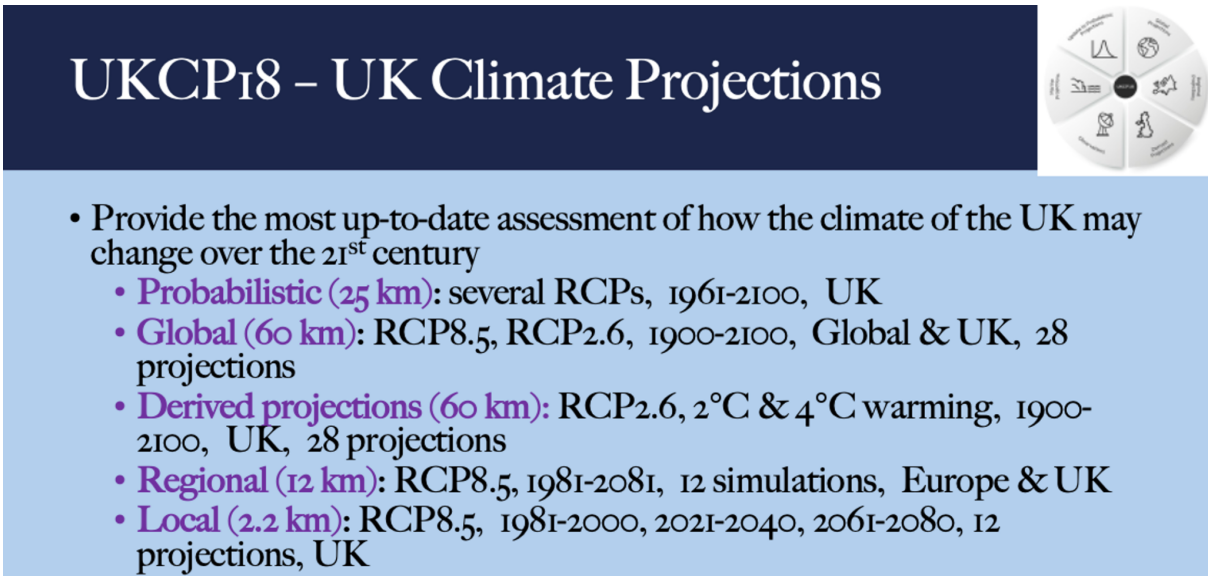
6. **Project start (earliest and latest possible) / end dates and duration of ASG funding (if different):**

Earliest possible start date: September 2022
Latest possible start date: October 2022
End date: March 2023

7. **Briefly describe the project and if already active the progress to date (please provide links to any outputs as relevant) Max 150 words**

UKCP18 Met Office climate projections data exists at multiple temporal-spatial resolutions and Representative Concentration Pathways (RCPs). While these datasets are widely available, there is a crucial step missing that many stakeholders (e.g. local councils, public health officials) really need to apply the data for their purposes: data localised to their areas. Additionally, climate projections are not just available for the future, but also for past and present. *Re-calibration* at local levels is required to bias adjust data in these different periods using on-the-ground weather measurements.

This project proposes extending existing DyME-CHH work, which necessitates re-calibrating local climate projections for our stakeholders across a national scale. This builds upon on expertise and collaborations from existing ASG projects (RAMP, DyME, QUANT, SPENSER, Synthetic Population Catalyst) and further develops collaborations with **Met Office/JCEEI** stakeholders to facilitate easy access and use of re-calibrated data by academia, industry, and not for profit organisations.



8. **We are particularly interested in projects that contribute to the vision and showcase: (1) data integration and/or model linking; (2) good data science practice and RRI; (3) effective visualisation and/or communication to stakeholders. How would the proposed project development highlight these areas?**

<u>Data integration and/or model linking</u>

Access to high-quality, directly usable, climate data will allow more people working in the environmental and climate change spaces to achieve more. This project is fundamentally a data integration and model linking project as it aligns data from the 1) Met Office climate projections dataset with 2) Had-UK climate observations in addition to other models from these providers. To

improve usability of climate data in a form that is accessible for stakeholders, particularly non-specialists, we plan to host data outputs in a range of spatial formats and CRS (coordinate reference system).

This recalibration process is currently performed adhoc by climate scientists for specific areas and time periods, rather than at scale due to computational requirements for manipulating and processing large datasets and the domain-expertise needed to apply sophisticated *downscaling* techniques. As such, when it has been performed, datasets, and knowledge of the implementation of methods, tend to be restricted to small user groups and shared directly within that network rather than from a single source of truth that all can directly access. This leads to issues with reproducibility of findings, transparency of methods, and lack of meaningful comparison, or more generally the ability to perform comparisons, of the range of bias adjustment methods used for calibrating projections with local climate data.

This project will also provide direct data inputs to the on-going work with the **NERC Digital Solutions Hub** and plans to develop a use case of the Data Integration Model for Air Quality (DIMEX), a SPF Clean Air funded project, which shows how digital twins (DTs) representing the interaction between environmental stressors, human activities and health can be built by integrating models from different domains. This work also draws on Dave Topping's Turing 2.0 project (Towards DT of the Urban Environment, which refers to DIMEX) and recent work on an Urban Transport DT, funded by DfT. The proposed open repo will dramatically shorten the time to development of DTs that are linked to urban-scale climate effects, such as in this case focussing on the interaction with air quality.

Effective visualisation and/or communication to stakeholders

The project output will include creating RShiny dashboards for visualisation of the re-calibrated climate data. A key component of the project will be to run training sessions and workshops (and associated materials) with stakeholders to ensure that the end result meets their needs.

We will also aim to provide climate projections translated to different global warming levels (e.g. 2C, 4C above pre-industrial levels) rather than the high-emission scenario (RCP8.5) to aid non-domain expert stakeholders in understanding and using the data. We will also provide this data in relative form compared with more recent baselines (e.g. user's year of birth, today) for easy integration into other kinds of climate reporting activities.

9. **How would the proposed project/initiative contribute to cross-theme impact?**

This project builds upon past ASG work with RAMP/DyME, QUANT, SPENSER, Synthetic Population Catalyst (SPC), and DIMEX and brings together key researchers across Turing ASG CTPs to build a valuable open resource that can benefit from the track record of collaboration that ASG has already established. Innovations in calibration, data integration and

visualisation can be leveraged to the benefit of other ASG stakeholders e.g. within the Digital Twin Urban Analytics mobility theme (DfT, TMF) and the DT-UA land use theme (Geospatial Commission, RTPI).

Specifically, this collaboration will involve PIs/PDRAs affiliated with Ecosystems of Digital Twins, Urban Analytics, Shocks and Resilience, Health, and Environment & Sustainability.

10. How do you intend to engage end users to ensure real-world impact?

The RAM will work to ensure that stakeholder needs are embedded in the shaping of the research output through the following activities:

- Stakeholder map of core UK climate data stakeholders and identifying relevant partners including The Met Office and target local councils
- User interviews with stakeholders throughout the project
- User testing before deployment
- Stakeholder trainings & workshops to promote the tool and generally upskill UK climate sector on working with key UK climate datasets

11. What financial resources are needed for the project development? (approximately, full costings will be requested if proposal is successful)

Estimated costings: **£15,000**

- **Azure infrastructure: £10,000**
  - £7000 - During project: 7 months of funding for GPUs to enable data processing work
  - £3000 - After project: funding for data storage and RShiny dashboard access for up to 24 months after project ends
- **Workshops and trainings: £5,000**
  - Create open documentation and training/workshop materials on climate data and the CLIM-RECAL repo targeting non-experts
  - Deliver 5+ stakeholder workshops with the Met Office and with local authorities in partnership with the LCAT team

12. What human resources / skills are required? Do you have someone in mind already and when would they be available to start? Does this person consider themselves to be a member of an underrepresented group in STEM / Data Science?

The core research team is composed of two female researchers and one Hispanic male researcher, both underrepresented categories in STEM/DS. The core skills the research team bring are in health and urban analytics research, microsimulation, agent-based modelling, and stakeholder engagement. These should be sufficient to develop a solid MVP that can be improved with technical and domain-specific support, as detailed below.

The technical skills needed include data science and software engineering, in order to scale up the MVP to a production-level open data repository. Existing REG time allocated to the DyME-CHH project from Fall 2022 - Spring 2023 should meet these skills requirements.

The domain-specific skills needed can be fulfilled by climate scientist advisors from organisations like the Met Office, that the theme lead has working relationships with. We have already had conversations with these stakeholders and hope to establish a stronger collaboration through this project.

13. **What are the risks to the project and how would you mitigate these? In particular please note risks related to any recruitment.**

The representative climate pathways (RCPs) climate projections data are held in the Centre for Environmental Data Analysis (CEDA) data archive. CEDA requires registration and institutional affiliation for data access and we will need to ensure that they are on board with the processing and repackaging of data they are gatekeepers of. We will reach out to our contacts at CEDA, and the Met Office - the origin of the data - and engage them within the project so that we can ensure their input will shape the final output.

An additional consideration is that different research groups may feel ‘ownership’ of the different methods available that have been used in re-calibration and which have generated much discussion within the literature The use of an open repo will mitigate this by allowing specialists to contribute in a transparent manner, and will be designed to not be prescriptive to users.

Finally, by developing an open data resource the research team and Turing will need to develop supporting infrastructure for the community of users such as project governance, code of conduct, and sustainability plans. We hope to leverage the expertise within the TPS team to tackle these challenges.