Open Climate Fix: GSoC 2025 Project Ideas

Open Climate Fix was founded in 2019 by Jack Kelly and Dan Travers. Our vision is a world powered by clean and sustainable energy systems. We believe in the power of harnessing open source technology to drive forwards the adoption of renewable energy as quickly and efficiently as possible.

So far, our open source work has created some of the most accurate forecasting tools available within the UK, and we'd love your help to make them even better, ensuring we reduce barriers to renewable energy adoption.

Any questions?
Head of Technology: peter@openclimatefix.org
Senior Community Manager: emily@openclimatefix.org

Open Data PVnet

Project Description
We're building an open-source solar forecasting pipeline to integrate with OCF's PVNet model, using publicly available Numerical Weather Prediction data to forecast solar generation at the national level, starting with the UK. Currently, our main forecasting tool, Quartz Solar, is trained using a mixture of public and private datasets, and we want to create an effective model that uses 100% open data.

The data is ready to go, but we need a ML engineer to train the model. The aim will be to start with a UK forecast, but then extend to different countries.

Expected Outcome
A UK ML Solar forecast trained on free NWP data with the accuracy benchmarked.

Other Key Information
- ➔ Expected Size: 170hrs
- ➔ Skills: ML knowledge, Pytorch, Python. NWP knowledge is a bonus
- ➔ Difficulty level: Medium
- ➔ Related Reading: https://github.com/openclimatefix/open-data-pvnet, https://github.com/orgs/openclimatefix/discussions/24
- ➔ Potential mentors: Peter, Sukh, please make contact on here

Cloudcasting ML

Project Description

Traditionally, forecasting tools are trained using historical satellite data. As part of an innovative new project, we have been training a model to *predict* satellite images up to 3 hours ahead over the UK. This work is still at early stages, but we have already proved that a satellite forecast using a very simple ML model can improve our solar energy forecast. There is lots of opportunity to improve on this new and unique satellite forecast, from trying different video prediction and AI weather model architectures to training a diffusion based model to stop the satellite forecast being blurry.

Expected Outcome
An improved satellite forecast model

Other Key Information
- ➔ Expected Size: 350hrs
- ➔ Skills: ML, pytorch, python
- ➔ Difficulty level: Hard
- ➔ Related Reading: https://www.openclimatefix.org/projects/cloud-forecasting, https://github.com/openclimatefix/sat_pred, https://github.com/orgs/openclimatefix/discussions/25
- ➔ Potential mentors: James, please make contact on here

Streaming Zarr from Cloud Storage (Ice Chunk + Zarr 3)

Project Description
We use a large amount of Satellite and Numerical Weather Prediction data all saved in Zarr format for training our ML models. We normally have a local copy, rather than using the cloud. We would like to explore using Ice Chunk and Zarr 3. It would be great to create a benchmark when training our PVNet model with data in cloud storage (using modern stack like Ice Chunk + Zarr 3). We could then use this to measure speed and compare it to training using data on disk.

Expected Outcome
A quantitative comparison between the speeds when using these new tools and not.

Other Key Information
- ➔ Expected Size: 175hrs
- ➔ Skills: ML Knowledge, Familiarity with training ML Models with Pytorch, Python, Data Analysis. Zarr is a bonus
- ➔ Difficulty level: Medium
- ➔ Related Reading: https://github.com/openclimatefix/Satip/issues/222, https://github.com/orgs/openclimatefix/discussions/29
- ➔ Potential mentors: sol@openclimatefix.org, peter@openclimatefix.org

Solar Forecast with TZ-SAM: Create Forecasts for Sites or a Grid (Country) Based on Transition Zero's Solar Asset Mapper

Project Description

The Transition Zero Solar Asset Mapper (TZ-SAM) is a new open source tool launched by not-for-profit Transition Zero (TZ). It provides a database of large solar farms, around 60,00 in total, from around the world that have been identified by using advanced AI satellite imagery identification. We want to use the TZ Solar asset mapper dataset to forecast each of the sites it identifies and/or map out total capacities for large sites in each country. Then we can use these capacities to generate solar forecasts for different grids around the world.

Expected Outcome

Basic solar forecasts, using Open Quartz, for every site/grid/country in the TZ-SAM dataset using a global NWP provider. Bonus: Aim to have it in production by the end of the mentoring period.

Other Key Information

- ➔ Expected Size: 175hrs
- ➔ Skills: Running ML models, docker
- ➔ Difficulty level: Medium
- ➔ Related Reading: https://www.transitionzero.org/products/solar-asset-mapper https://github.com/openclimatefix/open-source-quartz-solar-forecast, https://github.com/orgs/openclimatefix/discussions/28
- ➔ Potential mentors: Zak, please make contact on here

Cloudcasting UI

Project Description

We've been working on forecasting clouds using satellite imagery, and we would now like to take the next step and make an exciting visual representation of these cloud forecasts. This is an exciting and innovative project in the weather forecasting space. We would either implement this into our current product, Quartz Solar Forecast, or a separate tool all together.

Expected Outcome

A UI that clearly shows cloud forecasting.

Other Key Information

- ➔ Expected Size: 175hrs
- ➔ Skills: Plotting and visualising in a front end language, understanding of user experience, good problem solving skills.
- ➔ Difficulty level: Medium
- ➔ Related Reading: https://openclimatefix.org/projects/cloud-forecasting
- ➔ Potential mentors: Brad, please make contact on here

Adjuster this! TabPFN as a replacement for the adjuster

Project Description
For the OCF Quartz Solar forecasting model, we have a simple “adjuster” model. It currently looks at the pattern of errors in the last week, and adjust the new forecast. We would like to experiment with a new foundational timeseries forecasting model, TabPFN, with the planned outcome to dynamically adjust our PV forecasting model based on historical errors and improve the forecasting skill. Can compare to current rules based averages method used.

Expected Outcome
An answer to whether a pretrained ML model such as TabPFN can provide better adjustment of solar forecasts than rules based methods. If yes, then the plan would be to incorporate this into OCF's production system.

Other Key Information

- ➔ Expected Size: 90hrs
- ➔ Skills: ML knowledge + familiarity with ML tools, some data processing skills required e.g. pandas
- ➔ Difficulty level: Easy
- ➔ Related Reading: https://github.com/openclimatefix/india-forecast-app/blob/main/india_forecast_app/adjuster.py, https://github.com/PriorLabs/TabPFN , https://github.com/orgs/openclimatefix/discussions/27
- ➔ Potential mentors: Sukh, please make contact on here

Quartz Solar: New data source in ML model

Project Description
Adding new data sources usually gives a boost to the predictive power of our models, and finding innovative ways of extracting information from them can be even more beneficial. We would like to explore ways to improve our solar energy forecast with an ablation study of how much data on features like dust or neighbouring sites can contribute to the precision of the model. The project can be scaled depending on time constraints.

Expected Outcome
A comparative analysis of the effects of auxiliary data sources on the forecast quality.

Other Key Information

- ➔ Expected Size: 175hrs

- ➔ Skills: Familiarity with training ML models, preferably with pytorch. Data analysis skills would be highly beneficial. All our data is processed with xarray, at least basic familiarity is essential.
- ➔ Difficulty level: Medium-Hard
- ➔ Related Reading: https://github.com/openclimatefix/open-data-pvnet, https://github.com/openclimatefix/PVNet, https://github.com/orgs/openclimatefix/discussions/26
- ➔ Potential mentors: Alex, Sukh, please make contact on here

Improving Probabilistic Solar Forecasts

Project Description

Quantile regression approach in probabilistic solar forecasting directly estimates specific quantiles of the predictive distribution. Instead of direct quantity outputs, the proposition is modification of final layers to predict parameters of Gaussian Mixture Model(s) - for each component the mean, STD and coefficient. This should fundamentally capture more complex multimodal uncertainties inherent within probabilistic solar forecasting.

Expected Outcome

Complete continuous probability distribution as opposed to fixed quantiles. Better representation of uncertainties (such as transitions) and better capture of bimodal scenarios. Overall hopefully more accurate modelling and pattern comprehension - stronger uncertainty bounds perhaps.

Other Key Information

- ➔ Expected Size: 90hrs
- ➔ Skills: Academically - statistics / mathematics / engineering to Masters level. Experience implementing GMMs from scratch. ML/DL/PyTorch. Related architectural knowledge.
- ➔ Difficulty level: Easy-Medium
- ➔ Related Reading: https://github.com/openclimatefix/PVNet, https://scikit-learn.org/stable/modules/mixture.html https://github.com/ldeecke/gmm-torch, https://github.com/orgs/openclimatefix/discussions/30
- ➔ Potential mentors: Felix, please make contact on here