## EuroSun 2026

# TOWARDS FOUNDATION MODELS FOR SOLAR THERMAL SYSTEMS

---


### Summary

This work introduces a two-stage masked autoencoding framework tailored for solar thermal systems (STS), representing a step toward foundation models for small-scale energy systems. Using a Vector quantized variational autoencoder (VQVAE) based on one dimensional convolutional neural networks (1D-CNN) for tokenization and a Vision Transformer (ViT) for masked prediction, the model learns complex cross-correlations across eight sensors from the PaSTS dataset. Our preliminary results demonstrate that this scalable architecture enables zero-shot performance on diverse downstream tasks: achieving a Mean Absolute Error (MAE) of 1.88 °C in sensor estimation and an F1 score of 0.72 in anomaly detection. These findings suggest that foundation model architectures can effectively address the challenges presented by sustainable heating and cooling applications such as missing data and system variability.



*Keywords: Solar thermal systems, deep learning, foundation models, machine learning*

*Field of application for the findings: Sustainable Heating and Cooling Systems for Buildings*

*Addressed audience: Researchers in energy system modelling and ML practitioners in the renewable energy sector.*


---

### 1. Motivation and Problem Definition

Deep learning is increasingly defined by foundation models (FMs), large scale architectures trained on diverse data that are applicable to numerous downstream tasks. These FMs have revolutionised the area of language processing and vision (See Chatgpt, Gemini or other commercially available models), as well as other fields, such as climate and earth system modelling ([1, 2]). However, the trend to FMs has not yet been adapted to energy systems. Solar Thermal Systems (STS) present the typical challenges of small-scale energy systems. Real-world installations are diverse and rarely match standardized guidelines, frequently missing sensors, experiencing intermittent dropouts and unique boundary conditions.

The classical approach of developing individual simulations or models for every small-scale system is economically unfeasible. Consequently, a data-driven approach must be capable of generalizing across system realisations and handling incomplete data. Furthermore, it must be adaptable to diverse downstream tasks. In this work, we address these gaps by applying a scalable masked autoencoding framework to the public PaSTS dataset [5]. This represents a critical step toward a unified foundation model for sustainable heating systems, capable of performing diverse tasks like sensor estimation and fault detection without task-specific training.

### 2. Related Works

**Machine Learning in STS Applications:** Recent work in machine learning for solar thermal systems has focussed on specific applications like fault detection [6], short-term forecasting, control and load management [7]. These approaches often rely on simulation data or individual case studies, struggling to generalize across diverse system realizations. While Ebmeier et al. [6] has utilized the multi-system PaSTS dataset [5] to address generalization, studies typically employ standard convolutional neural networks (CNN) and long short-term memory (LSTM) architectures rather than scalable foundation-style models.

**Foundation Models:** The state-of-the-art in generative modelling has shifted toward Transformer-based architectures and discrete tokenization via Vector Quantization (VQVAE [9]). While autoregressive models dominate language processing, masked autoencoding has emerged as a more recent alternative in foundation models [3, 8, 4]. These models are explicitly suited for engineering applications, as they inherently handle missing sensor data and variable input modalities by treating them as masked during inference. Additionally, specific downstream tasks can be framed as specific masks in inference, without specifying those tasks during training.

**Data:** While STS sensor data is increasingly logged, public datasets remain scarce. We utilize the PaSTS dataset [5], which contains diverse real-world domestic STS installations. We follow the reannotated fault labels and evaluation splits defined in Ebmeier et al. [6], processing 8 sensor time series in daily windows.

## 3. Methods

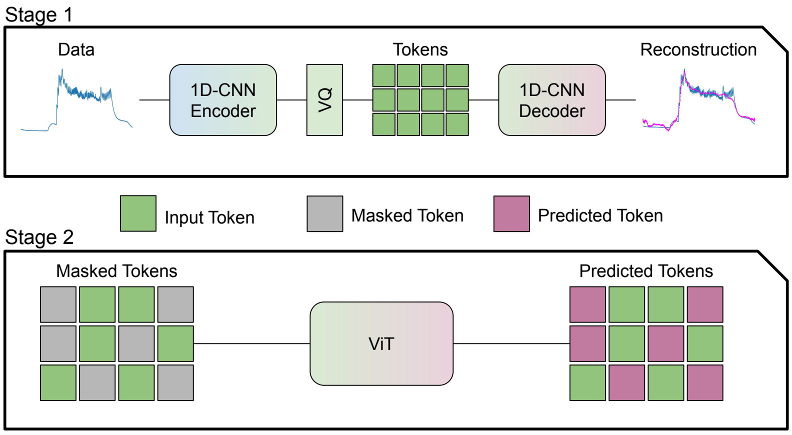


**Fig. 1:** Schemata of the two-stage model. In stage 1 a standard CNN based VQVAE is trained. Stage 2 is a ViT based masked autoencoder.

Following the framework of MaskGIT [3] and MD4 [8], our model utilizes a two-stage approach trained exclusively on nominal data from the PaSTS dataset. The overall structure can be seen in Fig. 1.

**Stage 1: Univariate Tokenization.** A VQVAE with 1D-ConvNext blocks and strided convolutions tokenizes 1440 daily time steps into $L$ discrete tokens. To ensure the model learns local dynamics before cross-correlations, each of the $V = 8$ sensors is processed independently as a univariate series. Once trained via standard reconstruction and vector quantization loss, Stage 1 weights are frozen.

**Stage 2: Masked Latent Modelling.** We employ a Vision Transformer (ViT) architecture to model the $V \times L$ token grid. During training, tokens are randomly replaced using a cosine masking schedule. The model is optimized via Cross Entropy Loss, rescaled by the derivative of the masking schedule to approximate the Evidence Lower Bound (ELBO) of the clean data distribution $p_{\text{clean}}$ [8].

**Downstream Tasks**. Tasks are framed as inference-time masking problems:

- **Sensor Estimation/Forecasting:** Defined by applying a static mask to specific variables or future time steps. The ViT predicts the missing tokens in a single forward pass or via iterative decoding.
- **Anomaly Detection:** Treated as an out-of-distribution task. By integrating the loss over the entire masking schedule, we approximate the variational lower bound $\mathcal{L}_{\text{vlb}}(x)$. Samples yielding high $\mathcal{L}_{\text{vlb}}(x)$ values relative to a threshold are classified as faulty.

## 4. Preliminary Results

**Tab. 1:** Preliminary performance on downstream tasks.

| Task | Metric | Value |
|---|---|---|
| Sensor Estimation | MAE [°C] | 1.88 |
| Anomaly Detection | Overall F1 Score | 0.72 |
| Anomaly Detection | System-wise F1 Score | 0.36 |

Two primary downstream tasks were evaluated to test the zero-shot capabilities of the architecture. Results are summarized in Table 1.

**Sensor Estimation:** The model effectively estimates the individual temperature sensors using the remaining sensors, achieving a MAE of 1.88 °C. Visual inspection (Fig. 2) of the collector temperature estimation

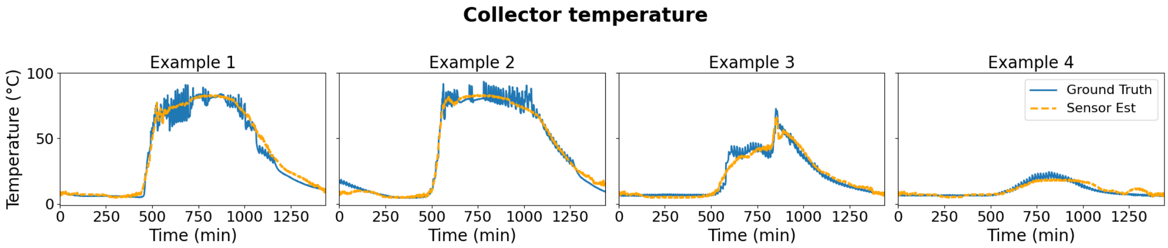


**Fig. 2:** Example of sensor estimation task. The collector temperature is estimated based on the other sensor values. We see a good mean prediction with a smoothing effect.

confirms the good reconstruction. We attribute the observed smoothing effect to the stage 1 compression rate and deterministic modelling.

**Anomaly Detection:** By approximating the $\mathcal{L}_{\text{vlb}}(x)$, the model successfully identifies anomalous days without having seen fault data during training. While the overall F1 score is promising (0.72), system-wise performance (0.36) indicates good generalisation between systems. Preliminary tests suggest that introducing block-masking significantly improves detection by forcing the model to rely on cross-correlations rather than local temporal interpolation. Additionally, we observed a strong sensitivity to the Stage 1 compression rate and masking priors.

## 5. Outlook

Our findings demonstrate that a scalable, two-stage masked autoencoder can match the performance of single-purpose baselines in a zero-shot setting. Ongoing efforts focus on: (i) extending Stage 1 to probabilistic reconstructions to capture uncertainty, (ii) investigating the effect of Stage 1 compression rate on downstream performance, and (iii) implementing task-specific masking priors (e.g., block-masking) to enhance task specific performance.

Overall, the model could be further extended towards a multi-domain dataset encompassing different types of medium-scale heating systems and incorporating more (meta-)data. Because such richer data are not openly available, we plan to acquire and publish such data (similar to [5]). This will allow to fully develop our approach as a viable foundation for general-purpose energy system modelling.

## References


[1] Cristian Bodnar, Wessel P Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A Weyn, Haiyu Dong, et al. A foundation model for the earth system. *Nature*, 641(8065):1180–1187, 2025.

[2] Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.

[3] Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu, and William T. Freeman. MaskGIT: Masked generative image transformer. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

[4] Google DeepMind. Gemini diffusion: Scalable non-autoregressive language modeling, 2025. URL `https://deepmind.google/models/gemini-diffusion/`. Accessed: 2026-02-19.

[5] Florian Ebmeier, Nicole Ludwig, Georg Martius, and Volker H. Franz. PaSTS an operational dataset for domestic solar thermal systems. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, e-Energy '24, page 529–534, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704802. doi: 10.1145/3632775.3662159. URL `https://doi.org/10.1145/3632775.3662159`.

[6] Florian Ebmeier, Nicole Ludwig, Jannik Thuemmel, Georg Martius, and Volker H Franz. Fault detection in solar thermal systems using probabilistic reconstructions. *arXiv preprint arXiv:2511.10296*, 2025.

[7] Mashud Rana, Subbu Sethuvenkatraman, Rahmat Heidari, and Stuart Hands. Solar thermal generation forecast via deep learning and application to buildings cooling system control. *Renewable Energy*, 196:694–706, 2022. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2022.07.005. URL `https://www.sciencedirect.com/science/article/pii/S0960148122010011`.

[8] Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis Titsias. Simplified and generalized masked diffusion for discrete data. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 103131–103167. Curran Associates, Inc., 2024. doi: 10.52202/079017-3277.

[9] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.