EuroSun 2026

Physics-based regression modeling for temperature prediction in large-scale water pit thermal energy storage

Summary

Global decarbonization efforts highlight the critical role of large-scale water pit thermal energy storage (PTES) in district heating systems. However, existing numerical and physical models require significant expertise for model setup and face challenges in accurately defining physical parameters. This paper proposes a physics-based regression approach based on the governing equations of water thermal dynamics. The model is developed and validated using monitoring data from the Marstal PTES in Denmark. The results show that the model achieves good predictive accuracy and maintains stable long-term predictions, with a mean absolute error of around 2.4 °C. In addition, the deviation in energy content is below 1%, and the charging/discharging energy is predicted with errors within 10%. The proposed approach provides an alternative for rapid system assessment without the need for complex calibration procedures.

Keywords: PTES; Physics-informed modeling; temperature prediction; monitoring data; long-term prediction

Field of application for the findings: Water pit thermal energy storage systems, for rapid simulation and control-oriented applications.

Addressed audience: Researchers and engineers in thermal energy storage, district heating systems, and data-driven or physics-informed modeling

1. Introduction

The transition to sustainable energy systems requires integrating high shares of variable renewable energy sources such as solar and wind. District heating networks play a key role in decarbonizing the heating sector [1], where seasonal thermal energy storage (TES) helps balance the mismatch between summer heat supply and winter demand [2]. Among TES technologies, large-scale pit thermal energy storage (PTES) stands out due to its low cost and suitability for seasonal load shifting [3].

Currently, the design, optimization, and operation of PTES rely heavily on numerical or physical models [4]. These range from computational fluid dynamics (CFD) models for detailed physical phenomenon investigation [5], to simplified system-level models, such as TRNSYS [6], Semi-analytical models [7], or equation-oriented object models (Modelica) [8].

Traditional modeling approaches have a high application threshold for non-experts. They require accurate identification of dynamic and uncertain parameters, leading to complex calibration processes [9]. Moreover, detailed numerical models are computationally intensive, limiting their use in real-time control and long-term simulations [10].

Regression-based methods offer an alternative to conventional modeling approaches by learning system dynamics directly from operational data. In this study, a physics-based regression model is developed based on the governing equations of water temperature dynamics. The model is further assessed through long-term predictions using multi-year monitoring data, demonstrating its ability to capture the main thermal behavior of the system with reasonable accuracy and stable performance. By focusing on real operational data, this work provides a practical approach for temperature prediction in PTES systems, with reduced reliance on complex model calibration.

2. PTES description and data preprocessing

The Marstal PTES in Denmark is selected as the case study. The storage pit, serving as a pilot site for large-scale seasonal water storage, has a total volume of 75,000 m^3. The pit is designed in an inverted pyramidal shape with a top width of 88 m, a top length of 113 m, and a depth of 16 m. It features an insulated floating cover to minimize heat losses to the environment and is integrated into the district heating network, including solar collectors and heat pumps. Three diffusers are located at the top, middle, and bottom, facilitating charging and discharging operations at different depths. For temperature monitoring, Class A PT100 sensors are installed vertically along the central axis of the pit, providing water temperature measurements for 32 layers [11].

For the regression model development, monitored data from the Marstal PTES (2013–2016) are used. Data quality control is applied to address noise and sensor anomalies. Outliers are detected using a sliding window with a one-month span, where values deviating from the median by more than three times the standard deviation are removed. The resulting missing data are filled by interpolation. For long gaps exceeding 24 hours, interpolation across adjacent layers is applied instead of linear interpolation to preserve temperature stratification [12].

3. Methodology

3.1 Physics-based constrained regression

To describe the thermal dynamics of the water pit, the governing equation for a representative layer j, with temperature T_j, volume V_j, density ρ, and specific heat capacity c_p, can be expressed as:

$$\rho c_p V_j \frac{dT_j}{dt} = U_\mathrm{c} A_{j-1,j}(T_{j-1} - T_j) + U_\mathrm{c} A_{j+1,j}(T_{j+1} - T_j) + U_\mathrm{b} A_\mathrm{b}(T_\mathrm{b} - T_j) + \rho c_p \sum_k \dot{m}_{k\to j}(T_k - T_j) \quad (1)$$

where T_{j-1} and T_{j+1} are the temperatures of the adjacent layers, T_b is the boundary temperature (soil or air at that depth), U_c and U_b are effective heat-transfer coefficients, $A_{j-1,j}$ and $A_{j+1,j}$ are the interfacial areas between adjacent layers, A_b is the boundary heat-transfer area, and $\dot{m}_{k\to j}$ denotes the mass inflow rate from node k to layer j. After combining like terms, Eq. (1) can be rewritten in a linear-in-parameters form:

$$\begin{aligned}\rho c_p V \frac{dT_j}{dt} = &-\left(U_\mathrm{c} A_{\mathrm{c},j+1} + U_\mathrm{c} A_{\mathrm{c},j-1} + U_\mathrm{b} A_\mathrm{b}\right) T_j - \rho c_p m_{j+1} T_j - \rho c_p m_{j-1} T_j \\ &+ U_\mathrm{c} A_{\mathrm{c},j+1} T_{j+1} + U_\mathrm{c} A_{\mathrm{c},j-1} T_{j-1} + U_\mathrm{b} A_\mathrm{b} T_\mathrm{b} + \rho c_p \dot{m}_{j+1} T_{j+1} + \rho c_p \dot{m}_{j-1} T_{j-1} \end{aligned} \quad (2)$$

Based on Eq. (2), a physics-guided candidate library $\Theta(\mathbf{T})$ is constructed, including temperature, boundary temperature, flow, and inflow temperature terms. Specifically, the candidate functions are:

$$\Theta(\mathbf{T}) = \left[T_{j-1},\ T_j,\ T_{j+1},\ T_\mathrm{b},\ \dot{m}_{k\to j},\ \dot{m}_{k\to j}T_k,\ \dot{m}_{k\to j}T_j\right] \quad (3)$$

The system dynamics can then be written in a compact matrix form as:

$$\dot{\mathbf{T}} = \Theta(\mathbf{T})\Xi \quad (4)$$

where $\dot{\mathbf{T}}(t)$ denotes the time derivative of the water temperature vector, and Ξ is the coefficient matrix to be identified. The coefficients are estimated using sparse relaxed regularized regression (SR3).

$$\mathrm{SR3}(\Xi, W) = \min_{\Xi, W} \frac{1}{2}\left\|\dot{\mathbf{T}} - \Theta(\mathbf{T})\Xi\right\|_2^2 + \lambda R(W) + \frac{1}{2\nu}\left\|\Xi - W\right\|_2^2 \quad (5)$$

where the first term corresponds to the residual error, measured as the squared difference between observations and predictions. The second term, $R(W)$, is an L_2 regularization (Ridge) term added to reduce model complexity. The L_2 penalty shrinks coefficients, thereby reducing multicollinearity and improving numerical stability. The third term softly couples the fit variable Ξ to the regularized variable W, which yields numerically stable Ξ-updates. The parameter ν controls the balance: smaller ν tightens the coupling and strengthens stabilization; larger ν weakens the coupling and increases the effective sparsifying action of the proximal step.

To ensure physical consistency, the following constraints are imposed:

- Constraint I: The coefficient in front of T_j equals the sum of the coefficients in front of T_{j+1}, T_{j-1}, and T_b;
- Constraint II: The coefficient is the same for terms with the same flow rate;
- Constraint III: The coefficient is the same for terms involving T_j and a flow rate;
- Constraint IV: Terms that are not included in the governing equation are set to zero.

3.2 Key performance indicators

To evaluate the thermal performance of the PTES, three key performance indicators (KPIs) are considered, including the energy content, charging energy and discharging energy.

The charging/discharging energy is defined as the net power exchanged through the inlet and outlet flows:

$$Q_{\mathrm{ch/dis}} = c\,(\dot{m}_{\mathrm{in}}T_{\mathrm{in}} - \dot{m}_{\mathrm{out}}T_{\mathrm{out}}) \tag{6}$$

where $\dot{m}_{\mathrm{in}}$ and $\dot{m}_{\mathrm{out}}$ are the mass flow rates at the inlet and outlet, respectively, and T_{in} and T_{out} are the corresponding temperatures. A positive value of $Q_{\mathrm{ch/dis}}$ indicates charging, while a negative value indicates discharging.

The energy content of the storage is evaluated with respect to a reference temperature T_{ref} (set to 10 °C in this study):

$$E_{\mathrm{sto}} = \sum_{j=1}^{N} \rho c V_j \left(T_j - T_{\mathrm{ref}}\right) \tag{7}$$

where N is the number of layers, V_j is the volume of layer j, and T_j is the temperature of layer j.

4. Model performance

The model is configured with five temperature layers, including three layers associated with the diffusers and two intermediate layers between them. It is found that excluding Constraint II leads to improved performance; therefore, only Constraints I, III, and IV are enforced. Under this configuration, the model achieves minimum MAE values of 2.33 and 2.40 for the 30-min and 1-h datasets, respectively.

The comparison of water temperatures in five representative layers is shown in Fig. 1. The predicted temperatures agree well with the measurements, indicating that the model captures the main thermal dynamics across different depths. Moreover, the model remains stable over a three-year prediction period, reproducing long-term temperature variations without divergence. For the upper and middle layers, the model accurately reproduces both the seasonal trends and short-term variations. In contrast, larger discrepancies are observed in the bottom diffuser layer. This is mainly due to the presence of spikes and abrupt fluctuations in the measured data, which are difficult for the regression model to capture. These fluctuations are likely caused by strong mixing and flow disturbances near the bottom diffuser, leading to more complex local dynamics that deviate from the underlying modeling assumptions. As a result, the prediction errors in the bottom layer are higher compared to other layers.

The temperatures of the five selected layers are interpolated to reconstruct the full 32-layer temperature profile, taking advantage of the thermal stratification of the water. Based on this reconstructed profile, the energy content and charging/discharging energy are calculated and compared with the measurements, as shown in Fig. 2. The results show that the deviation in energy content is less than 1%, indicating that the thermal stratification is well captured and that the interpolation to 32 layers is reliable. The charging energy is also well predicted, with errors below 5%. However, a relatively large discrepancy is observed in the discharging energy in 2016, which may be attributed to the degradation of the floating lid during that period [9]. The damaged lid led to increased heat losses and lower actual water temperatures than expected, resulting in an overestimation of the discharging energy by the model.

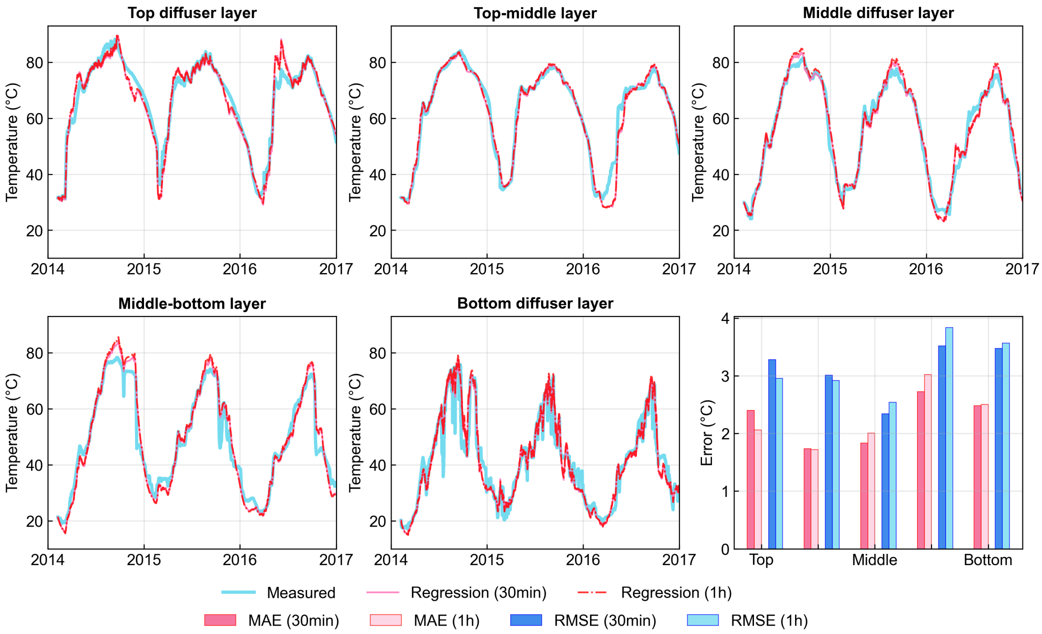


Fig. 1: Comparison between predictions and measurements. Shown is the 5-layer prediction enforcing Constraints I, III, and IV, and employing one year of training data. "Top–middle layer" refers to the layer located between the top and middle diffusers, and "middle–bottom layer" refers to the layer between the middle and bottom diffusers.

5. Conclusion

This study proposes a physics-based regression model for water temperature prediction. The model is validated using a long-term thermal storage system and demonstrates stable performance over extended prediction periods. With five temperature layers and the enforcement of Constraints I, III, and IV, the model achieves minimum MAE values of 2.33 and 2.40 for the 30-min and 1-h datasets, respectively, while the deviation in energy content remains below 1% and the error in charging/discharging energy is within 10%. The model requires minimal implementation effort, as it avoids parameter tuning and complex calibration. Although only five layers are used, temperatures in other layers can be reasonably reconstructed through interpolation due to thermal stratification. Future work will focus on validation on other PTES and further improving prediction accuracy and robustness.

6. Acknowledgments

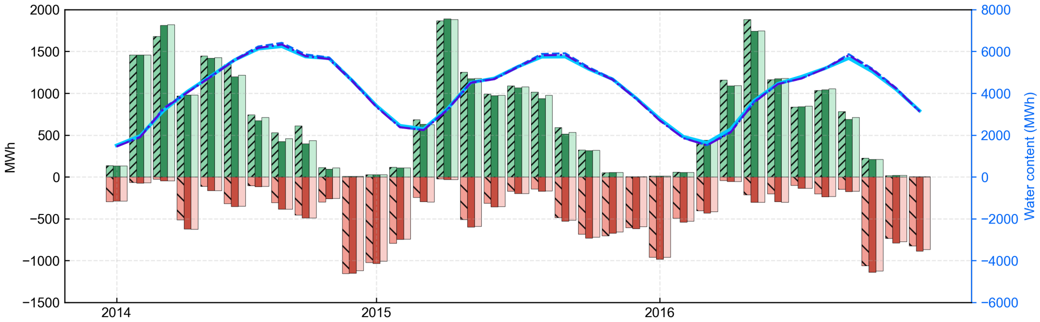
The authors would like to acknowledge the financial support and contributions from EU Horizon Europe TREASURE project (reference number: 101136095). The manager of Marstal Fjernvarme, Lasse Kjærgaard Larsen, kindly provided the monitoring dataset. Their contributions are valued and appreciated.

References

[1] Ioannis Sifnaios, Daniel Møller Sneum, Adam R Jensen, Jianhua Fan, and Rasmus Bramstoft. The impact of large-scale thermal energy storage in the energy system. *Applied Energy*, 349:121663, 2023.

[2] Abdulrahman Dahash, Fabian Ochs, and Alice Tosatto. Advances in modeling and evaluation of large-scale hot water tanks and pits in renewable-based district heating. In *Proceedings of 13th International Conference on Solar Energy for Buildings and Industry (EuroSun 2020)*, 2020.

[3] Amaya V Novo, Joseba R Bayon, Daniel Castro-Fresno, and Jorge Rodriguez-Hernandez. Review of seasonal heat storage in large basins: Water tanks and gravel–water pits. *Applied Energy*, 87(2):390–397, 2010.



Year	Charge energy (1h)	Discharge energy (1h)	Water content (1h)	Charge energy (30min)	Discharge energy (30min)	Water content (30min)
2014	-4.9%	8.0%	0.9%	-3.1%	6.9%	-0.1%
2015	-4.0%	5.0%	0.5%	-3.1%	3.2%	-0.4%
2016	-4.0%	10.8%	-0.3%	-3.1%	9.0%	-0.9%

Fig. 2: Comparison of charging/discharging energy and energy content between predictions and measurements.

[4] Michael Reisenbichler-S, Ioannis Sifnaios, and Franz Wotawa. Validation of a pit thermal energy storage model: Demonstration of a comprehensive approach. *Journal of Energy Storage*, 116:115680, 2025.

[5] Chun Chang, Zhiyong Wu, Helena Navarro, Chuan Li, Guanghui Leng, Xiaoxia Li, Ming Yang, Zhifeng Wang, and Yulong Ding. Comparative study of the transient natural convection in an underground water pit thermal storage. *Applied energy*, 208:1162–1173, 2017.

[6] Xinyu Pan, Yutong Xiang, Meng Gao, Jianhua Fan, Simon Furbo, Dengjia Wang, and Chao Xu. Long-term thermal performance analysis of a large-scale water pit thermal energy storage. *Journal of Energy Storage*, 52:105001, 2022.

[7] Meng Gao, Shuyang Shao, Yutong Xiang, Dengjia Wang, Simon Furbo, and Jianhua Fan. Semi-analytical model of a large-scale water pit heat storage for the long-term thermal applications. *Energy and Buildings*, 307:113975, 2024.

[8] Michael Bayer, Curtis Meister, Philipp Schuetz, Willy Villasmil, Heimo Walter, and Abdulrahman Dahash. Development of a reduced-order dynamic model for large-scale seasonal thermal energy storage applications. *Energy*, 333:137379, 2025.

[9] Pengcheng Wang, Meng Gao, Weiqiang Kong, Zhiyong Tian, Dengjia Wang, Zhongbing Liu, and Jianhua Fan. Experimental and numerical study of a large-scale water pit thermal energy storage: Quantifying the impact of progressive lid insulation degradation. *Journal of Energy Storage*, 144:119702, 2026.

[10] Guozhi Xu, Lei Hu, Yongqiang Luo, Zhiyong Tian, Jie Deng, Guofeng Yuan, and Jianhua Fan. Numerical modeling and parametric analysis of thermal performance for the large-scale seasonal thermal energy storage. *Energy and Buildings*, 275:112459, 2022.

[11] Ioannis Sifnaios, Adam R Jensen, Simon Furbo, and Jianhua Fan. Performance comparison of two water pit thermal energy storage (ptes) systems using energy, exergy, and stratification indicators. *Journal of Energy Storage*, 52:104947, 2022.

[12] Ioannis Sifnaios, Geoffroy Gauthier, Daniel Trier, Jianhua Fan, and Adam R Jensen. Dronninglund water pit thermal energy storage dataset. *Solar Energy*, 251:68–76, 2023.