



Check for
updates

NIST Technical Note NIST TN 2265

Comparison of Ice-on-Coil Thermal Energy Storage Models

Kalyan Ram Kanagala
Amanda Pertzborn, PhD

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2265>

NIST Series Technical Note

NIST TN 2265

Comparison of Ice-on-Coil Thermal Energy Storage Models

Kalyan Ram Kanagala
PREP Associate

Amanda Pertzborn, PhD
*Building Energy and Environment Division
Engineering Laboratory*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.TN.2265>

September 2023



U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

Certain commercial equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies

[Copyright, Use, and Licensing Statements](#)

[NIST Technical Series Publication Identifier Syntax](#)

Publication History

Approved by the NIST Editorial Review Board on 2023-09-08

How to cite this NIST Technical Series Publication:

Kalyan Ram Kanagala, Amanda Pertzborn, PhD (2023) Comparison of Ice-on-Coil Thermal Energy Storage Models. (National Institute of Standards and Technology, Gaithersburg, MD), NIST TN 2265. <https://doi.org/10.6028/NIST.TN.2265>

NIST Author ORCID iDs

Amanda Pertzborn, PhD: 0000-0002-1473-7500

Contact Information

amanda.pertzborn@nist.gov

Abstract

Data collected from the Intelligent Building Agents Laboratory (IBAL) at the National Institute of Standards and Technology (NIST) are used to develop a physics-based and four machine learning models of ice-on-coil thermal energy storage (TES): linear interpolation, linear regression, neural network, and Gaussian process. Data cleaning considerations are discussed in addition to presenting the results of the five models. For this TES system, which is linear over a significant range of operation, the linear interpolation model performs the best, but there is a thorough discussion of the advantages and disadvantages of each model.

Keywords

HVAC; machine learning; thermal storage.

Contents

1. Introduction	1
2. Data Generation from Laboratory Experiments	3
3. Data Preparation	4
4. Model Descriptions	9
4.1. Physics-Based Model	9
4.2. Interpolation	10
4.3. Linear Regression	10
4.4. Neural Network	10
4.5. Gaussian Process	11
5. Model Results	12
5.1. Physics-Based Model Results	14
5.2. Linear Interpolation Model Results	16
5.3. Linear Regression Model Results	19
5.4. Neural Network Model Results	21
5.5. Gaussian Process Model Results	23
6. Combined Model Results	25
7. Discussion	26
7.1. Convergence	26
7.2. Complexity	27
7.3. Accuracy	27
7.4. Ease of Implementation	27
7.5. Summary of Each Model	28
7.5.1. Physics-Based Model	28
7.5.2. Linear Interpolation	29
7.5.3. Linear Regression	29
7.5.4. Neural Network	30
7.5.5. Gaussian Process	30
8. Conclusions and Future Work	31
References	32

List of Tables

Table 1. Summary of data sets.	3
Table 2. Subset of the downsampled data used to calculate the load.	5
Table 3. Subset of the raw data set.	6
Table 4. Data sets used to develop the charge and discharge models.	12
Table 5. RMSE comparison of the models applied to the discharge mode data.	13
Table 6. RMSE comparison of the models applied to the charge mode data.	13
Table 7. Data sets used for the combined models.	25
Table 8. RMSE of the models developed on a combination of charge and discharge data.	25
Table 9. Model Comparison Table	28

List of Figures

Figure 1. Pictures of the TES used in the IBAL. Left: External view of the TES. Center: Internal view of the TES spiral coil with the foam covering in place. Right: Internal view of the TES spiral coil with the foam covering removed. The pictures show the TES without water.	1
Figure 2. Schematic showing how TES fits in the hydronic system.	2
Figure 3. Schematic showing the approximate locations of the sensors that generate the data used in building the ML models.	4
Figure 4. Change in the ice inventory as a function of integrated load for the (a) discharge data set and (b) charge data set used in training.	7
Figure 5. Multiple charge mode data sets (a) before filtering and (b) with the non-linear data filtered out.	8
Figure 6. Measured and smoothed flow rate through the thermal storage.	8
Figure 7. Multiple discharge mode data sets with (a) the nonlinear data filtered out and (b) the nonlinear data filtered out and the erroneous flow rates smoothed.	9
Figure 8. Discharge mode training and test data before and after smoothing	14
Figure 9. Comparison of the training and test data used for the ML models of the discharge case to the physical model.	15
Figure 10. Comparison of the training and test data used for the ML models of the charge case to the physical model.	15
Figure 11. Prediction vs actual data for the discharge case using both training and test data sets.	16
Figure 12. Prediction vs actual data for the charge case using both training and test data sets.	16
Figure 13(a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the charging mode.	17

Figure 14(a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the discharging mode.	17
Figure 15(a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.	18
Figure 16(a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the charging mode.	19
Figure 17(a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the discharging mode.	19
Figure 18(a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the discharging mode.	20
Figure 19(a) Neural network model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the charging mode.	21
Figure 20(a) Neural network model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the discharging mode.	21
Figure 21(a) Neural network model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.	22
Figure 22(a) GP model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) GP model prediction of the change in ice inventory as a function of integrated load for the charging mode.	23
Figure 23(a) GP model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) GP model prediction of the change in ice inventory as a function of integrated load for the discharging mode.	24
Figure 24(a) GP model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) GP model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.	24

Acknowledgments

This work was funded in part by grant 70NANB18H161. We also thank Farhad Omar, Natascha Milesi-Ferretti, Wai Cheong Tam, and William Healy for their reviews.

1. Introduction

The Intelligent Building Agents Laboratory (IBAL) [1–3] at the National Institute of Standards and Technology (NIST) contains real commercial heating, ventilation, and air conditioning (HVAC) equipment for use in studying advanced control approaches. The equipment includes air handling units (AHUs), variable air volume boxes (VAVs), chillers, and an ice-on-coil thermal energy storage (TES) tank. The IBAL can be used to emulate the operation of a real commercial building. In this context, the air system generates building cooling loads due to lighting, computer equipment, solar gain, occupants, etc. The hydronic system provides cooling for the building based on the thermal comfort and ventilation requirements of the emulated building occupants. The cooling is provided by either a chiller or the TES. The ice-on-coil TES in the IBAL is shown in Fig. 1. A 30 % propylene glycol (PG) solution flows through a plastic spiral coil and water surrounds that coil. The center and right hand pictures show the internal spiral coil. Operation of the TES is the focus of the study detailed in this technical note.



Figure 1. Pictures of the TES used in the IBAL. Left: External view of the TES. Center: Internal view of the TES spiral coil with the foam covering in place. Right: Internal view of the TES spiral coil with the foam covering removed. The pictures show the TES without water.

The process of making ice in the TES is called charging. The charging process uses a chiller to cool the PG below the freezing point of water. As the PG flows through the TES, the water surrounding the spiral coil freezes. The process of using the TES instead of the chiller to provide cooling for the building is called discharging. In this case, warm PG passes through the TES, where it transfers heat to the ice, melting it, before returning to the cooling coil in the AHU, where it picks up the heat from the air system, cooling the air and meeting the building cooling load. A simple control approach is to charge the TES when electrical rates are low, typically overnight, and discharge it when electrical rates are high, typically in the afternoon. Figure 2 is a schematic diagram depicting how the TES and the chillers work together and separately in meeting the building load.

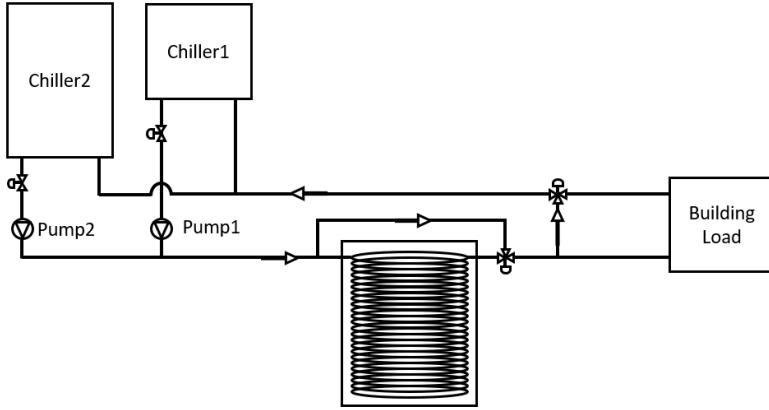


Figure 2. Schematic showing how TES fits in the hydronic system.

The IBAL was built to study advanced control approaches in commercial HVAC systems. These control approaches may require data driven models of equipment and subsystems. In this report we discuss several different data driven or machine learning (ML) approaches used to model the TES using data from the IBAL. We chose to focus on the TES first because it is a relatively simple system and there is a physics-based modeling approach for the TES that can be used as a baseline for comparison.

Physics-based models of ice-on-coil systems include those presented in Refs. [4–6]. A comprehensive literature review did not return ML models of ice-on-coil TES; most ML work for ice-on-coil TES focuses on the control aspects of the system [7–11]. The goals of the present study are to:

1. apply different ML approaches to model TES to understand the advantages and disadvantages of the different approaches,
2. develop an understanding of how to evaluate the models,
3. determine how much the raw data has to be cleaned prior to building a model,
4. create TES models for use in future studies, and
5. create a strategy that can be used to develop and evaluate models for other pieces of equipment.

The details of how the data are generated is discussed in Sec. 2. Each ML model uses a set of training data to learn its parameters and a separate set of test data to evaluate how well the model performs. Section 3 discusses how the raw data is first formatted into the right form for use by the models (raw), then how bad data points are replaced in a smoothing process (smoothed), and, finally, how nonlinear data are filtered out (filtered). Models are built using raw, smoothed, and filtered data sets to evaluate how sensitive each model is to the data set. The goal is to understand how well each model copes with messy data.

The modelling approaches used in this study, described in Sec. 4, are:

- physics-based,
- linear interpolation,
- linear regression,
- neural network, and
- Gaussian process.

Initially, individual models are built for the charging and discharging modes of operation of the TES using the raw, smoothed, and filtered data sets, as presented in Sec. 5. In Sec. 6 the results of single models that capture both the charging and discharging modes of operation are presented. Section 7 is a detailed comparison of all the models and recommendations about which approaches to use and when to use them. Section 8 is a discussion of future work.

Table 1 is a summary of the data sets used in this study. By convention each data set name is the date on which the data were acquired (e.g., 2018_05_17 is May 17, 2018). The S and F indicate if the data set was smoothed or filtered, which is explained in more depth in Secs. 2 and 3. Combined indicates that the model is built using both charging and discharging data sets (Sec. 6).

Table 1. Summary of data sets.

Mode	Training	Testing
Discharge	2018_05_17	2018_05_09
	2018_05_17 S	2018_05_09 S
	2018_05_17 S and F	2018_05_09 S and F
Charge	2019_10_08	2019_10_22
	2019_10_08 F	2019_10_22 F
Combined	(2018_05_17+2019_10_08) F	(2018_05_09+2019_10_22) F
	(2018_05_17 S+2019_10_08) F	(2018_05_09 S+2019_10_22) F
Combined	(2018_05_17+2019_10_08) F	2018_05_09 F
	(2018_05_17 S+2019_10_08) F	2018_05_09 S and F
Combined	(2018_05_17+2019_10_08) F	2019_10_22 F
	(2018_05_17 S+2019_10_08) F	2019_10_22 F

S = Smoothed; F = Filtered.

2. Data Generation from Laboratory Experiments

Although the IBAL contains a variety of HVAC equipment and sensors, this current study focuses on the TES, so the other equipment will not be discussed in detail. The target of

the models is to predict the ice inventory, which is a measure of the amount of ice in the tank as a percentage of the maximum possible amount. The laboratory data that are used to build the predictive models of ice inventory are:

- ts_f [m^3/s or lpm] = the flow rate of PG through the TES,
- ts_{meter} [%] = the ice inventory (0 % to 100 %),
- ts_{in} [C] = temperature of the PG entering the TES, and
- ts_{out} [C] = temperature of the PG exiting the TES.

The locations where these data are measured are shown in Fig. 3.

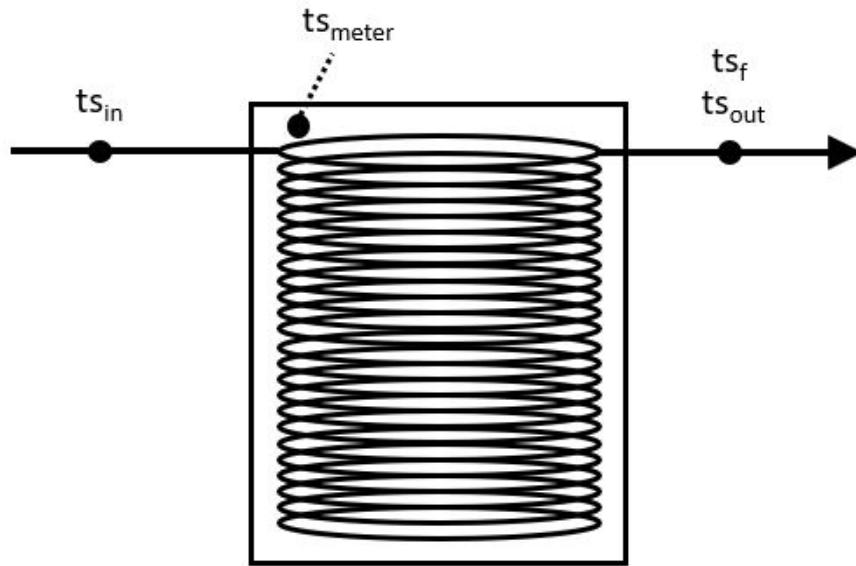


Figure 3. Schematic showing the approximate locations of the sensors that generate the data used in building the ML models.

The data are recorded every 10 s and can be accessed from the [IBAL database](#). Select the Experiments page and search the Experiment/Metadata drop-down using the name of the test (e.g. 2019_10_08). The interested reader can access other data sets from this database interface [12].

3. Data Preparation

The ML models presented in this study predict ice inventory given the measurement data. A predictive model of ice inventory is useful for making control decisions. In the case of

TES, control decisions take place at multiple time scales. A decision could be made about whether or not to discharge the TES in the next few hours or the next few days. Decisions about when to discharge also impact decisions about when the TES will need to be charged, which could also be in a few hours or in a few days.

One approach to building a model of ice inventory is to treat the data as a time series and perform time series forecasting using techniques such as recurrent neural networks or auto-regression. These models look at the recent history and predict a few hours into the future. In this work, however, we are interested solely in predicting what the ice inventory will be in the future given a cumulative load over a given time interval and an initial inventory. This type of forecasting is interested in a bulk behavior rather than the more detailed behavior captured by a time series model. In general, the time series model is useful for predictions on the order of a few hours, while the approach detailed here can be used to predict over a few hours or over a few days.

These models discussed in this report require the integrated load, the initial ice inventory, and the final ice inventory for that integrated load. The integrated load is calculated from the instantaneous load by use of the trapezoidal rule. Equation 1 is used to calculate the instantaneous load Q , where c_p is the specific heat of PG (3.9 kJ/kg-K) and ρ is the density (1030 kg/m³).

$$Q[kW] = ts_f * \rho * (ts_{in} - ts_{out}) * c_p \quad (1)$$

The initial 10 s data is downsampled to every 10 minutes because the ice inventory does not change significantly in 10 s. This downsampling is performed by calculating the mean flow rate, ice inventory, input and output temperatures for every 10 minute interval (i.e., the mean of 60 datapoints sampled at 10 s). This is performed by using the Pandas resample method and passing “10T” as the argument. Table 2 shows a portion of the downsampled data.

Table 2. Subset of the downsampled data used to calculate the load.

Time [min]	ts _f [lpm]	ts _{meter} [%]	ts _{in} [C]	ts _{out} [C]	Q [kW]
0	0	47.4	17.9	16.6	0
10	4.9	47.6	19.0	15.9	1.0
20	76.5	46.7	0.3	0.2	0.4
30	75.2	47.0	-3.2	-0.3	-14.7
40	74.9	47.4	-3.4	-0.4	-14.8
50	74.9	48.3	-3.5	-0.6	-15.0
60	74.9	49.3	-3.6	-0.6	-15.0
70	74.9	49.9	-3.6	-0.6	-15.0
80	75.3	50.8	-3.6	-0.6	-15.0
90	75.3	51.3	-3.6	-0.6	-14.9

The integrated load is calculated by integrating the downsampled instantaneous load with respect to time for a specified time interval. The input data for the models is a 2D array that contains the initial ice inventory and an integrated load; the output is the final ice inventory resulting from the integrated load applied to the initial inventory. Table 3 shows a subset of the inputs and output; this is the raw data set. The first row shows a trivial result - when the time interval of integration is 0 minutes, the final inventory is the same as the initial inventory. The second row shows that the integrated load from time zero to time 10 minutes is 0 kWh. However, the final inventory is slightly different than the initial inventory. In theory, the final and initial inventories should be identical when the integrated load is zero, but they are different due to measurement uncertainty. The third row shows the results for the load integrated from time 0 to time 20 minutes, the fourth row for time 0 to time 30 minutes, etc. As the absolute magnitude of the integrated load increases, the difference between the final and initial inventories increases.

This table shows only the first 10 rows of data, in which the initial inventory is always 47.4 %, as measured at time 0. This procedure continues until the last time step. Then the procedure starts over from time 10 minutes, so the initial inventory is again 47.4 % (the final inventory from time 0 minutes) and continues to the last time step. This procedure repeats until the final time step so that the full set of inputs and the accompanying output used in training and testing the models is all possible combinations of initial and final ice inventory. For n individual values of ice inventory, there are $n*(n-1)/2$ combinations.

Table 3. Subset of the raw data set.

Time Interval [min]	Initial Inventory [%]	Integrated Load [kWh]	Final Inventory [%]
0	47.4	0	47.4
10	47.4	0	47.6
20	47.4	0.085	46.7
30	47.4	0.2	47
40	47.4	-1	47.4
50	47.4	-3.4	48.3
60	47.4	-5.9	49.3
70	47.4	-8.4	49.9
80	47.4	-10.9	50.8
90	47.4	-13.4	51.3

Figure 4 shows the change in the ice inventory (final inventory minus initial inventory) as a function of the integrated load for the charge and discharge modes. These are the raw data sets used to train the models. Similar data sets are used to test the trained models.

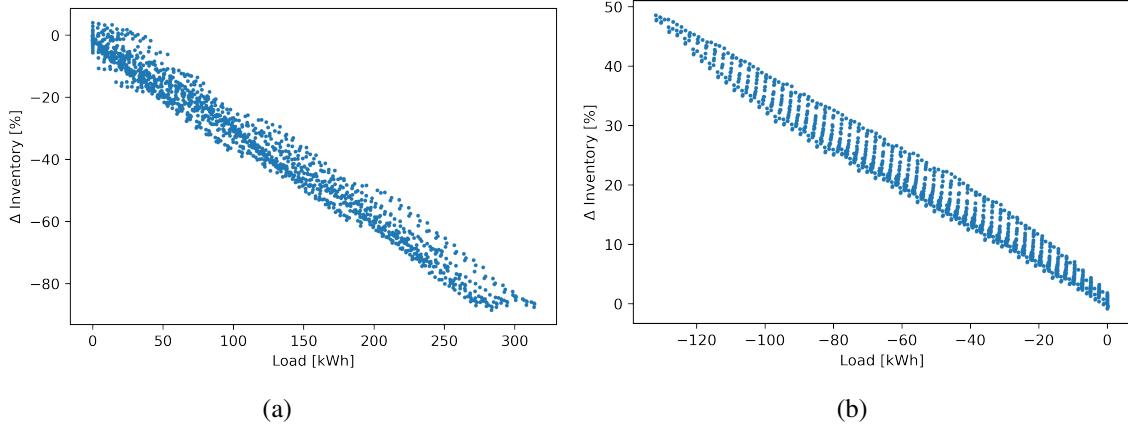


Figure 4. Change in the ice inventory as a function of integrated load for the (a) discharge data set and (b) charge data set used in training.

The data used in this study show distributional shifts. A distributional shift means that data sets generated from multiple experiments appear to have different probability distributions. We identified two reasons for this shift. First is measurement uncertainty. In Fig. 4, each x-value (load) corresponds to multiple y-values (change in ice inventory). In an ideal scenario, the relationship between load and the change in ice inventory should be a straight line. The presence of multiple y values for the same x value is due in part to measurement uncertainty that includes the typical uncertainty associated with the sensors and variations in the experimental setup.

The second major cause is a mix of linear and nonlinear data in the data set. Between ice inventory levels of approximately 30 % and 85 % the processes of charging or discharging are linear, and the models all assume an underlying linear function. Figure 5 shows five different charging data sets. All of these data sets should have the same slope, but in Fig. 5(a), which includes nonlinear data, the slopes differ substantially. When the nonlinear data are filtered out (Fig. 5(b)), the slopes are similar within uncertainty.

If the four data sets labeled “dataset” were used to train a model and the data set labeled “Test Data” was used to test that model, the model built using the unfiltered data would perform worse than the model built using the filtered data. The unfiltered data look like they come from different underlying distributions, but ML models generally assume that the train and test data have the same underlying statistical properties. When the data display this kind of distributional shift, there is no statistical guarantee on the performance of an ML model.

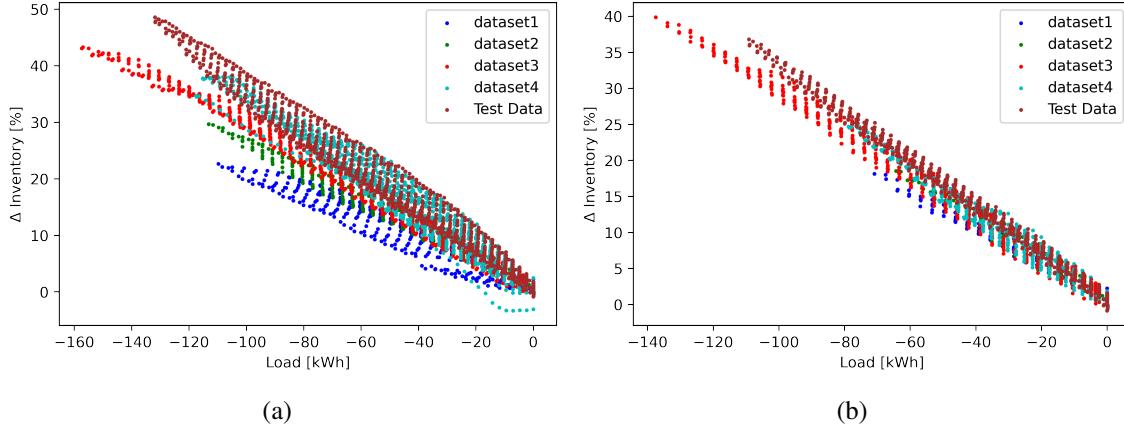


Figure 5. Multiple charge mode data sets (a) before filtering and (b) with the nonlinear data filtered out.

An additional source of noise is bad data. Plots of ts_f for the raw data set for the discharging mode reveal that there are outliers due to faulty measurement results. Faulty measurements are not an uncommon occurrence with real experimental data; to better understand the impact of this messy data on the final model, we cleaned the data using a “smoothing” process. The smoothing process replaces the erroneous flow data with the median of the correct measurements. This is the smoothed data set. Figure 6 shows the difference between the raw flow rate and the smoothed flow rate for the data sets used in training and testing.

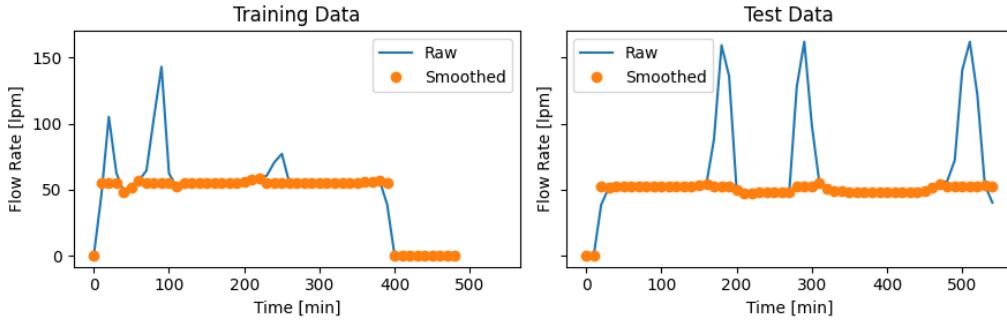


Figure 6. Measured and smoothed flow rate through the thermal storage.

Figure 7 shows data for the discharging mode. In this case, the left figure shows the data after filtering and the right figure shows the data after filtering and smoothing. Even after filtering and smoothing, there is more variance in the discharge data than in the charge data. This is an indication that the discharge process is generally less consistent because the flow

rate through the TES can vary during discharge based on the magnitude of the integrated load and the temperature set point.

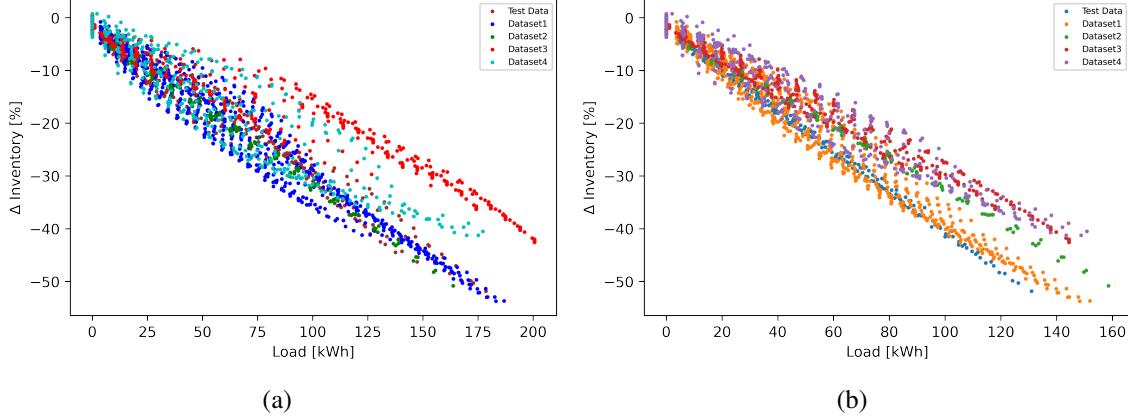


Figure 7. Multiple discharge mode data sets with (a) the nonlinear data filtered out and (b) the nonlinear data filtered out and the erroneous flow rates smoothed.

4. Model Descriptions

The models used in this study are discussed in this section. Later sections will present the results of these models and discuss their performance in more detail.

4.1. Physics-Based Model

The physics-based model is a simple model of the charging and discharging process of an ice-on-coil thermal storage tank that is only concerned with determining the change in ice inventory as a function of the energy added to (discharging mode) or removed from (charging mode) the TES [5]. Equation 2 is the simple equation.

$$\Delta I[\%] = \frac{-Q}{h_f \cdot m} \Delta t \quad (2)$$

The variable $h_f = 334 \text{ kJ/kg}$ is the latent heat of fusion required to change ice to water (or water to ice), $m = 2800 \text{ kg}$ is the total mass of water in the TES, Δt is the time (10 seconds for the laboratory data, the integration interval otherwise) and Q is the heat transfer in kW. The value of Q can be calculated from the laboratory data as shown in Eq. 1. This equation neglects any losses to the surroundings. Over the course of a day these losses are much smaller (on the order of 1 % or less depending on weather conditions) than the load applied to the TES from the system.

In discharge mode, the source of Q is the building load that is met by the ice tank. In this case Q is positive because energy is added to the TES, melting the ice and the change in

ice inventory is negative. In charge mode, the source of Q is the rate of charge of the ice tank. In this mode, Q is negative because energy is removed from the TES and the change in ice inventory is positive. Some models calculate ts_{out} [6] in addition to the change in ice inventory, but this study just considers the bulk behavior of the system for a given load. The following sections describe the ML approaches.

4.2. Interpolation

Interpolation is essentially a lookup table approach for predicting the value of a variable given the values of other variables. It is commonly used with property data. For example, given a measurement of the temperature and pressure of a fluid, interpolation can be used to estimate the thermal properties at that point based on a table of known thermal properties at other temperatures and pressures. For this study, the training data represent a lookup table, so, given an initial ice inventory and integrated load from the test data set, the final ice inventory is interpolated from the table of values provided by the training data. Interpolation is a well known and simple method to apply, but it has the limitation that it requires sufficient data to create the lookup table since it cannot extrapolate beyond the data it has seen. Linear interpolation is used in this study.

4.3. Linear Regression

In linear regression (LR), the output Y is modeled as the linear weighted combination of the input features, the functional form of which is shown in Eq. 3; the additive noise, ϵ , is assumed to be a unit normal distribution, and β are the parameters learned during the training process.

$$Y = \beta * X + \epsilon, \quad \epsilon \sim N(0, 1) \quad (3)$$

Although there are many variations of LR, in this study, ordinary least squares is adopted. One advantage of LR is that the loss function shown in Eq. 4, the mean squared error (MSE), is convex, so minimizing it will result in finding the global optimum. In Eq. 4 the input features X are the initial ice inventory and the integrated load, while the output Y is the final ice inventory. Therefore, this model has two parameters; β_1 is the weight on the initial ice inventory and β_2 is the weight on the integrated load. LR was selected because the change in the ice inventory is linearly proportional to the integrated load. Polynomial regression could be selected for equipment with nonlinear behavior.

$$mse = \frac{\sum (prediction - actual)^2}{N} \quad (4)$$

4.4. Neural Network

For the neural network (NN) the inputs X are the initial ice inventory and the integrated load and the output Y is the final ice inventory. In this study we manually tuned the NN

architecture before settling on the final architecture shown below. Given that the dynamics of the system are linear, with some nonlinearities when the ice inventory is outside the [30 %, 85 %] range, the starting point is a NN with a single layer, but a second layer is added to account for the nonlinear portions of the data sets. The final architecture is an input layer with 15 neurons, one hidden layer with 15 neurons, and the output layer with one neuron. The parameter values learned during the training process are the weights of the connections between neurons.

- Number of hidden layers : 1
- Number of neurons in input layer : 15
- Number of neurons in hidden layer : 15
- Number of neurons in final layer : 1
- Activation function : Rectified Linear Unit
- Optimization technique: Adam optimizer
- Weight Initialization: Glorot uniform
- Loss function: Mean Squared Error

4.5. Gaussian Process

The modeling techniques discussed so far do not include any estimate of the model uncertainty. Their model parameters are point estimates, meaning that when the models are used for prediction, they will always return the same output for a given set of inputs. In probabilistic ML, one technique that does assess model uncertainty is the Gaussian process (GP).

Equation 5 is a general equation for a model of the system, where Y is the observed sensor reading, ε is the measurement uncertainty, and $f(X)$ is the true value of the process being measured. The model is trying to learn $f(X)$. In a perfect world, with infinite data, $f(X)$ can be learned perfectly, but in reality, due to the limitations of the data sets, $f(X)$ is only being estimated by the model. The uncertainty in the estimate of $f(X)$ is the model uncertainty.

$$Y = f(X) + \varepsilon, \quad \varepsilon \sim N(0, 1) \quad (5)$$

When Eq. 3 is used for a GP, the ε is the same as in the LR model. All the models in this study see the impact of the noise in the measurement data, but the GP model also includes model uncertainty as captured in the $f(X)$ term. When Eq. 3 is used in LR, $f(X)$ is $\beta * X$, but the parameters β are point estimates of the slope and intercept. In a GP, $f(X)$ is a distribution that captures the model uncertainty. This is the equivalent of saying that the β is now a random variable with some associated probability distribution rather than a point estimate.

As with all Bayesian methods, the parameters of GP require a prior. Priors can take many forms. In a simple case, a prior is a Gaussian distribution on the model parameter with

some mean and variance [13]. For a GP, choosing the kernel is equivalent to choosing the prior; a linear kernel is chosen because the dynamics of TES are known to be linear.

A GP is specified by its mean function $m(X)$ and covariance, or kernel, function K . The function f can then be written as a GP according to Eq. 8, where X is the set of inputs.

$$m(X) = E[f(X)] \quad (6)$$

$$K(X, X') = E[(f(X) - m(X))(f(X') - m(X'))] \quad (7)$$

$$f(X) \sim GP(m(X), K(X, X')). \quad (8)$$

The mean function $m(X)$ and the parameters of the kernel function K are learned during the training process. The posterior prediction of the distribution of functions on the test data is given in Eq. 9. In applying GP to make a prediction, X^* is the test data, f_* is the random variable representing the function drawn from the probability distribution for the test data, and f is the random variable representing the function trained on the training data.

$$Pr(f_*|f) = F(K_{f^*f}K_{ff}^{-1}y, K_{f^*f^*} - K_{f^*f}K^{-1}K_{f^*f}^T) \quad (9)$$

Key settings for the GP model used in this paper are:

- Kernel: Linear;
- Loss function: Negative Marginal Log-Likelihood.
- Optimizer: BFGS.

5. Model Results

In this section the results are presented for each of the models as applied to charging and discharging data. These results are for the individual models of the charging and discharging modes. The results for the model built using the combined charge and discharge data are in Sec. 6. This approach allowed us to analyze the performance on the specific modes of operation. Table 4 lists the variations of the data sets used by the models discussed in this section.

Table 4. Data sets used to develop the charge and discharge models.

Charge	Discharge
raw	raw
filtered	smoothed
	filtered
	smoothed and filtered

The root mean square error (RMSE) is used to compare the models to the measurement data and is calculated as shown in Eq. 10, where N is the number of samples, prediction is the output from the model, and actual is the measurement. The RMSE for all models is shown in Tables 5 and 6.

$$rmse = \sqrt{\frac{\sum (prediction - actual)^2}{N}} \quad (10)$$

Table 5. RMSE comparison of the models applied to the discharge mode data.

Model	Data set	Raw	Smoothed	Filtered	Smoothed & Filtered
Physics-based	Test	9.9	3.5	6.3	2.4
Interpolation	Test	5	6.8	2.5	4.2
Linear regression	Test	5.5	7.8	2.5	4.1
Neural network	Test	7.6	7.8	2.4	3.8
Gaussian process	Test	4.6	7.8	4.1	1.5

Table 6. RMSE comparison of the models applied to the charge mode data.

Model	Data set	Raw	Filtered
Physics-based	Test	2.2	2.1
Interpolation	Test	1.8	1.5
Linear regression	Test	1.8	1.7
Neural network	Test	1.6	1.5
Gaussian process	Test	1.8	1.5

For the discharge models, the RMSE for the models developed using smoothed data are worse than those developed using the raw training data, which was unexpected. Figure 8 shows that the training and the test data for the smoothed case differ in their slope. As previously mentioned, an underlying assumption in all ML models is that the training and test data come from the same distribution. In the case of the raw data, although the data are not exactly the same, they have a similar slope, but in the case of the smoothed data, the slope of the test data is noticeably different from the training data because these data sets include the nonlinear regions.

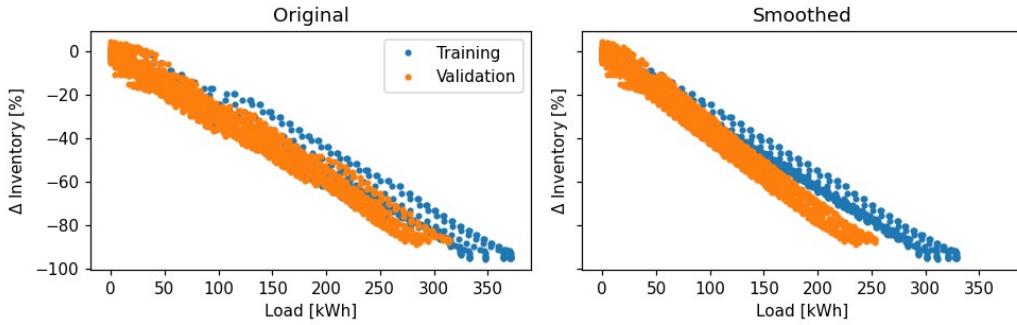


Figure 8. Discharge mode training and test data before and after smoothing

The RMSE of the models developed on filtered data show substantially better performance than those developed on the unfiltered data, but, for the discharge case, the trend where the smoothed data set has worse performance still holds. The overall conclusion is that the performance of a linear model is, unsurprisingly, much improved by removing the nonlinear data.

In the following sections the results are presented in a series of plots for each model. One plot shows the change in the ice inventory (initial - final) as a function of the integrated load for both the charge and discharge modes. A second plot shows the prediction vs the actual final ice inventory. The plots for the filtered data sets are not shown, but the basic trends are the same.

5.1. Physics-Based Model Results

Figure 9 shows the results of applying the physics-based model to data sets for the discharging mode and Fig. 10 shows the same for the charging mode. The model does well for the charging case, but the issue with the erroneous readings from the flow meter is apparent for the discharging case. The plots in Fig. 9 with the word “Smoothed” in the title use the integrated load calculated from the cleaned flow rate. These calculations yield a much lower value for the RMSE.

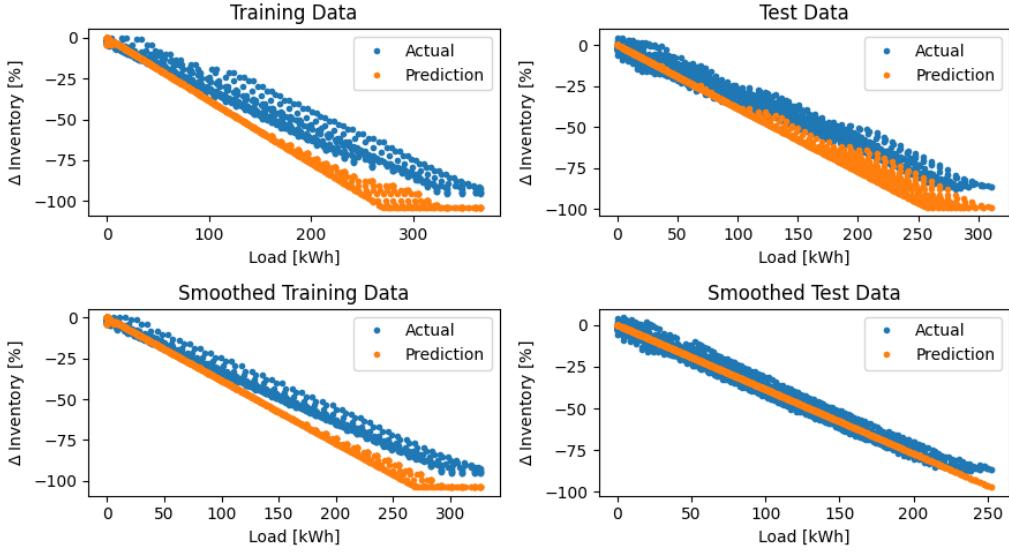


Figure 9. Comparison of the training and test data used for the ML models of the discharge case to the physical model.

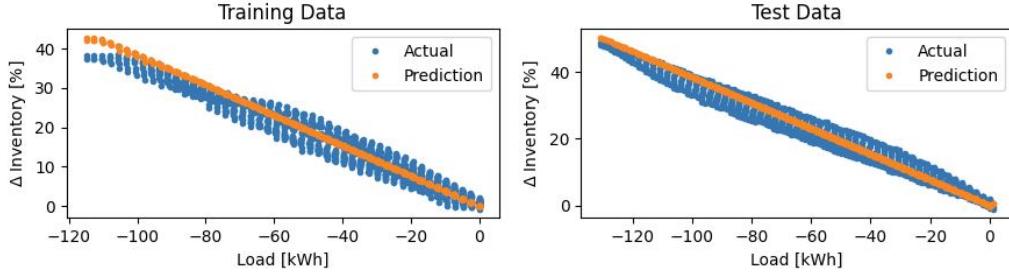


Figure 10. Comparison of the training and test data used for the ML models of the charge case to the physical model.

Figure 11 shows the prediction vs the actual result for the discharge data and Fig. 12 shows the same results for the charge data. In all the figures, the perfect prediction line is also shown. In the discharge case, using the training data, the prediction of the ice inventory for a given initial inventory and integrated load is generally lower than the actual result. This is also true of the test data set, but, when using the cleaned flow rate data, the prediction is much improved for both data sets, though the prediction for the training data set still tends to be low. For the charge data set, the prediction is generally good, centered around the perfect prediction line.

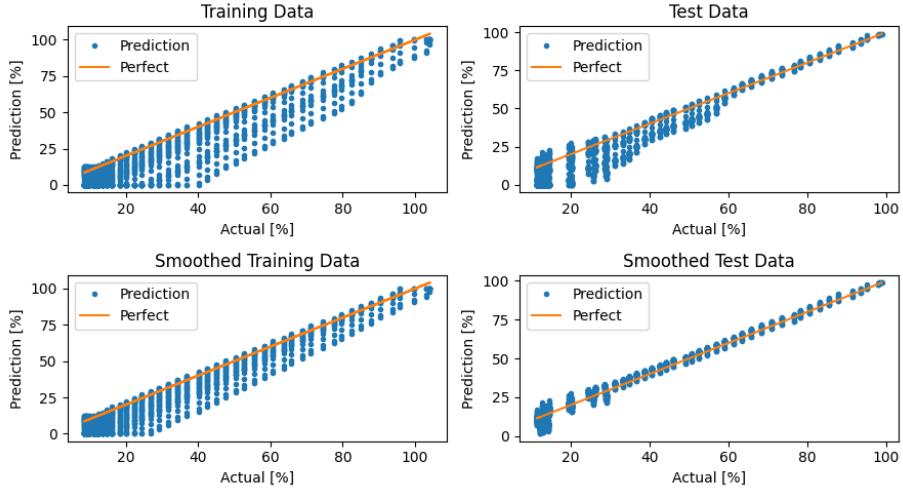


Figure 11. Prediction vs actual data for the discharge case using both training and test data sets.

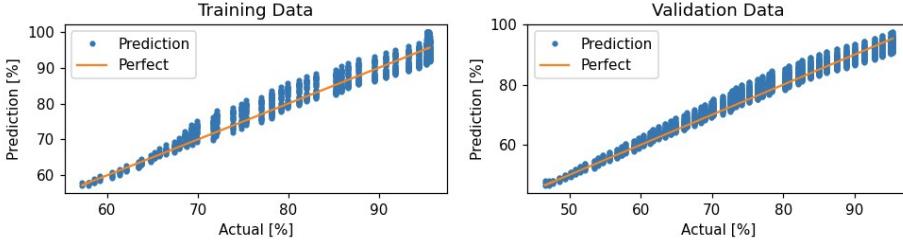


Figure 12. Prediction vs actual data for the charge case using both training and test data sets.

This relatively simple physics-based model provides a baseline for comparison for the ML models that are presented in the following sections.

5.2. Linear Interpolation Model Results

Figure 13 shows the linear interpolation results for predicted ice inventory vs the actual ice inventory and the predicted and actual change in the ice inventory as a function of integrated load for the charging mode. As seen in Fig. 13(a), linear interpolation is generally in good agreement with the actual data. However, in Fig. 13(b), there are missing data points for the prediction when the change in the ice inventory is greater than about 35 %. This is an illustration of a disadvantage of linear interpolation - the model cannot extrapolate beyond the training data. In the training data, the initial ice inventory and integrated load pairs range from [57.5 %, -117.5 kWh] to [95.59 %, 0.25 kWh]. In the test data, their values range from [46.7 %, -131.9 kWh] to [94.9 %, 0.2 kWh]. The test data include points that are not contained within the training set, so the prediction fails for those values. In the case

of the discharge mode, shown in Fig. 14, the test data fall within the range of values used in training, so there are no missing predictions. There are techniques to compute values outside of the training data set, but they are not part of the basic linear interpolation model.

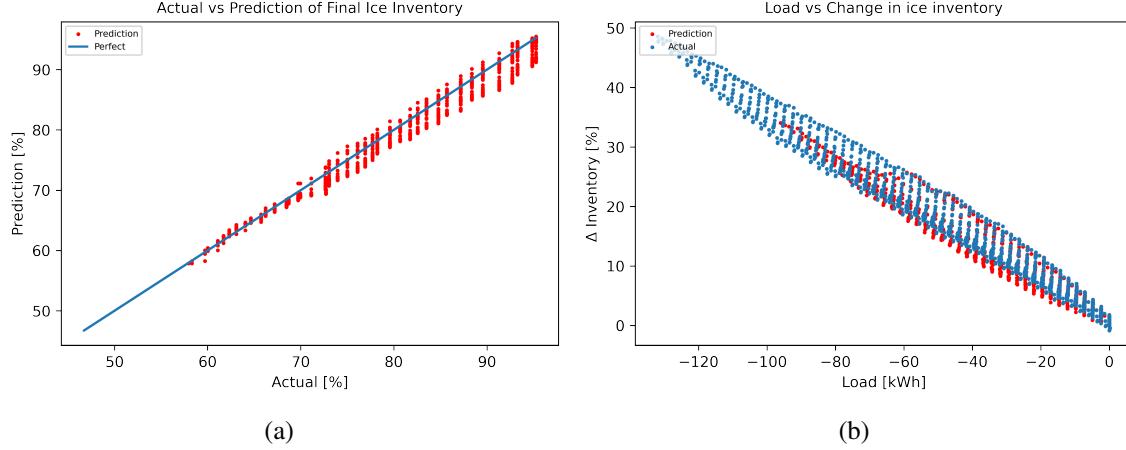


Figure 13. (a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the charging mode.

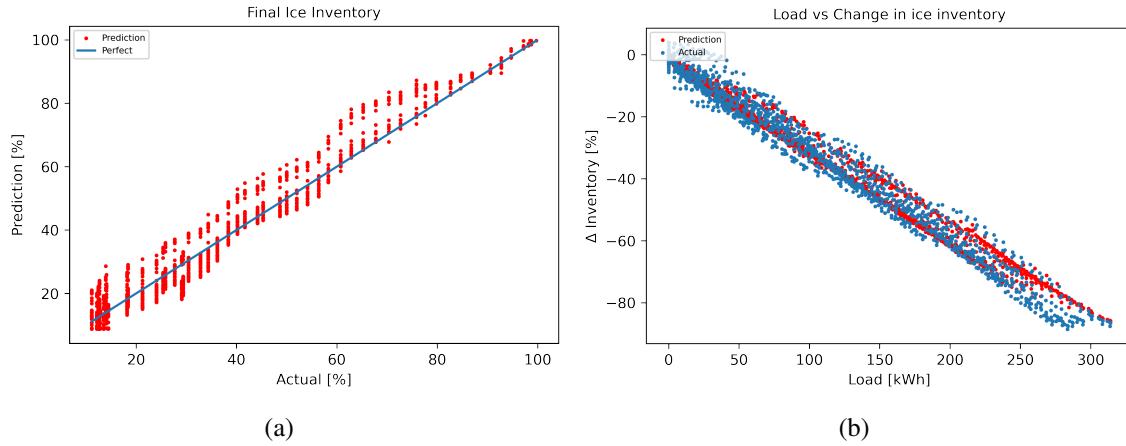


Figure 14. (a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the discharging mode.

Although in discharge mode the model does not have an issue with extrapolation, it does show high variance in the predictions. One reason is measurement noise, particularly in the flow rate, and when the ice inventory is below approximately 30 % the dynamics start

to become nonlinear. This issue does not appear in the charge data because the inventory level does not drop below 30 % and the raw data are cleaner. As seen in Fig. 15(a), smoothing the data does help reduce the variance to an extent, but Fig. 15(b) shows the previously mentioned issue that when smoothed, the prediction accuracy decreases due to the inclusion of the nonlinear regions.

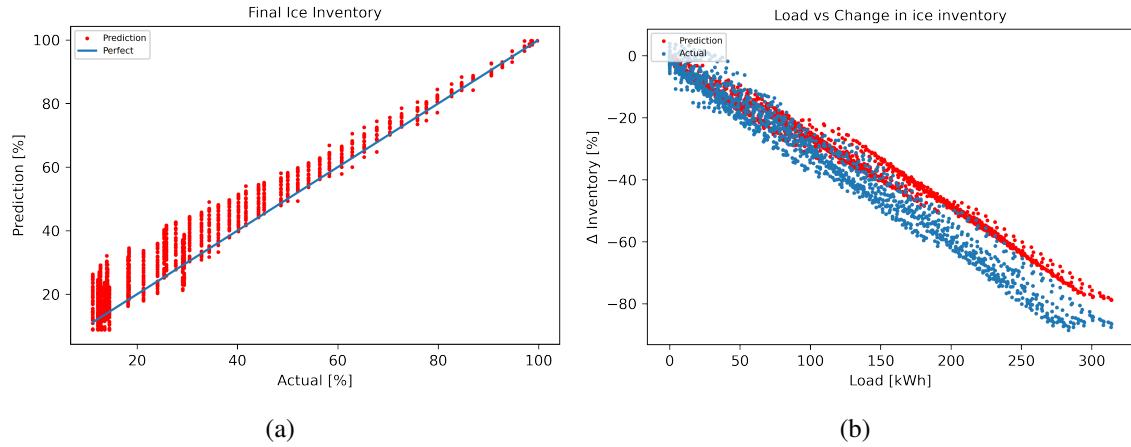


Figure 15. (a) Interpolation model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) Interpolation model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.

5.3. Linear Regression Model Results

Figure 16 shows the results for the LR model applied to the charging mode data. From Fig. 16(a), it appears that the model underpredicts the actual inventory below 60 % and above approximately 85 %. As shown in Fig. 16(b), unlike in the case of the interpolation model, the prediction is not limited by issues with extrapolation. However, consistent with Fig. 16(a), it underpredicts the inventory across the entire range of integrated loads. Figure 17 shows the regression results for the discharge case. There is slightly less variance than in the interpolation data, but the variance is still high when the ice inventory is below 20 %.

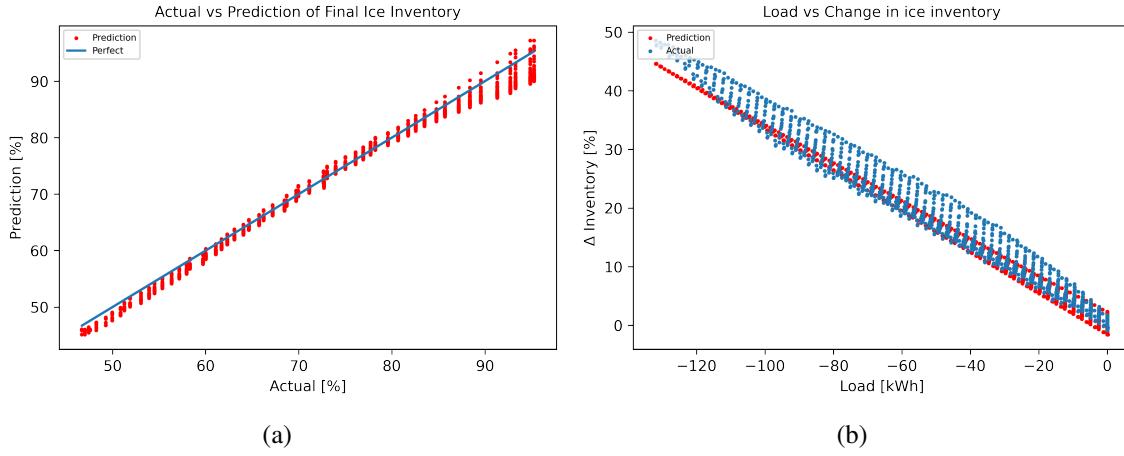


Figure 16. (a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the charging mode.

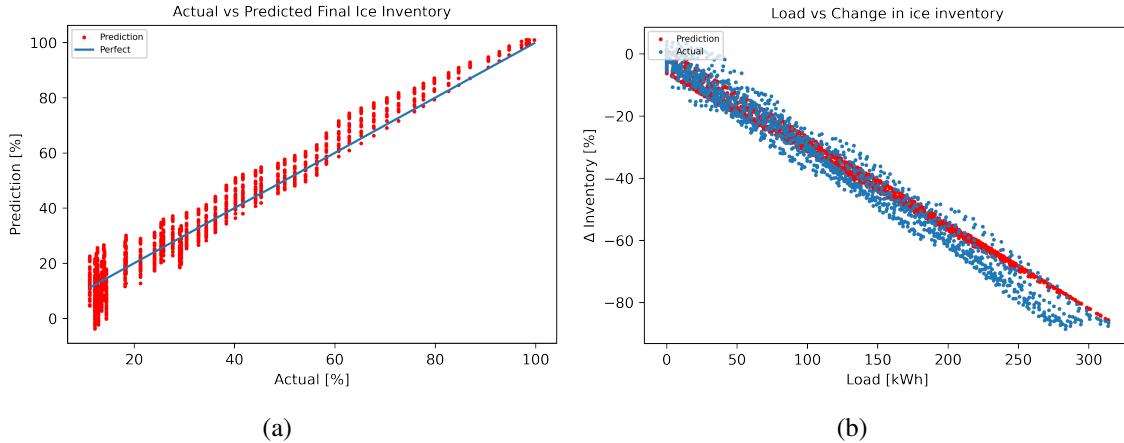


Figure 17. (a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the discharging mode.

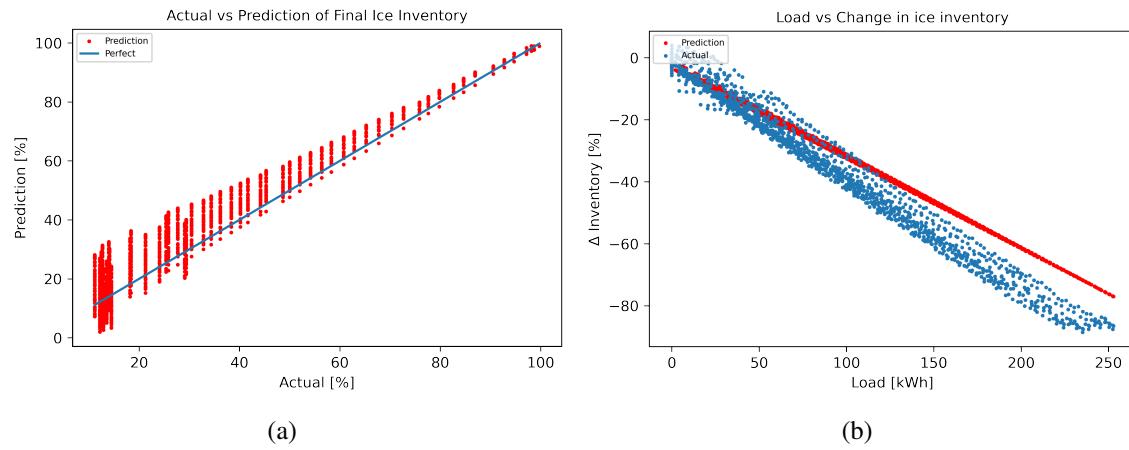


Figure 18. (a) Linear regression model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Linear regression model prediction of the change in ice inventory as a function of integrated load for the discharging mode.

5.4. Neural Network Model Results

Figure 19 shows the results for the charging mode for the NN model. Figure 20 shows the results for the discharging data and Fig. 21 shows the results for the smoothed data. The variance is still relatively high and in this case the model tends to overpredict the inventory. The results for the smoothed data are similar to the other cases, with higher variance at low inventory and lower variance as the inventory increases.

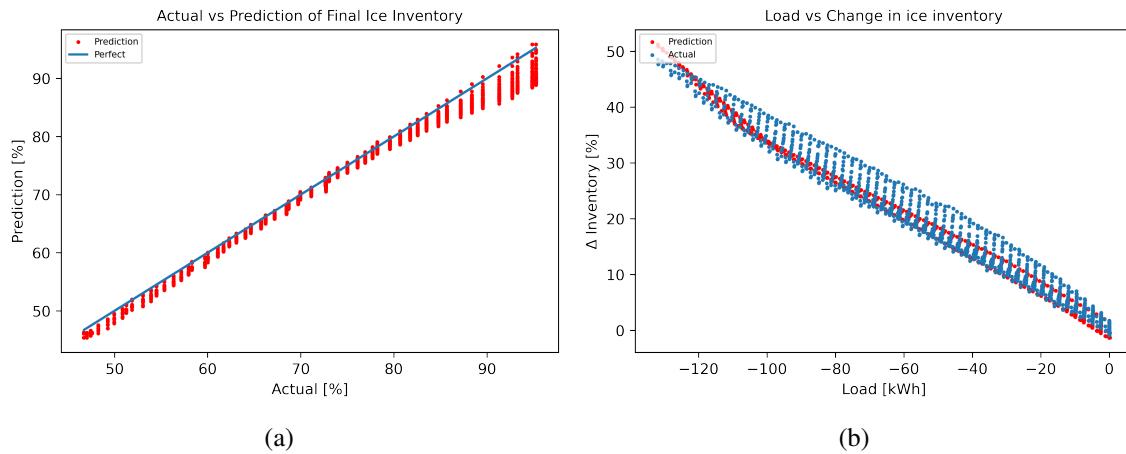


Figure 19. (a) Neural network model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the charging mode.

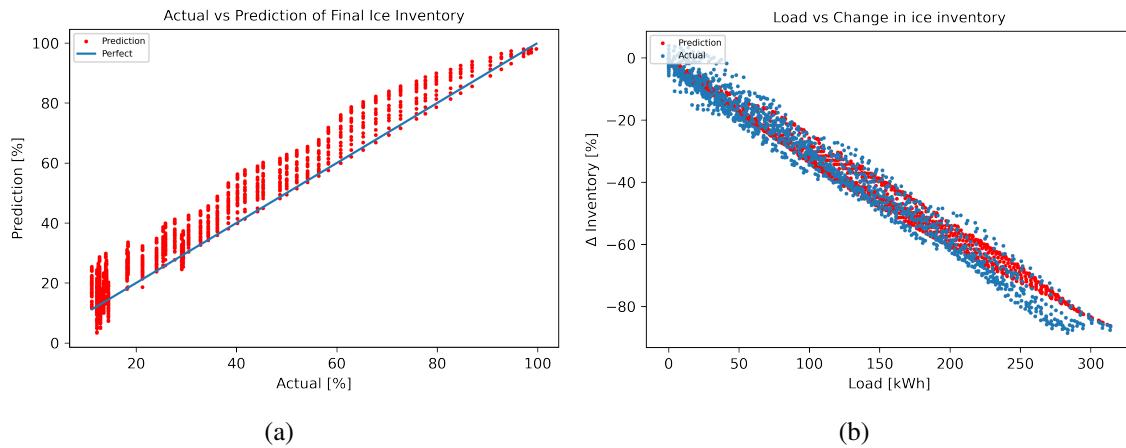


Figure 20. (a) Neural network model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the discharging mode.

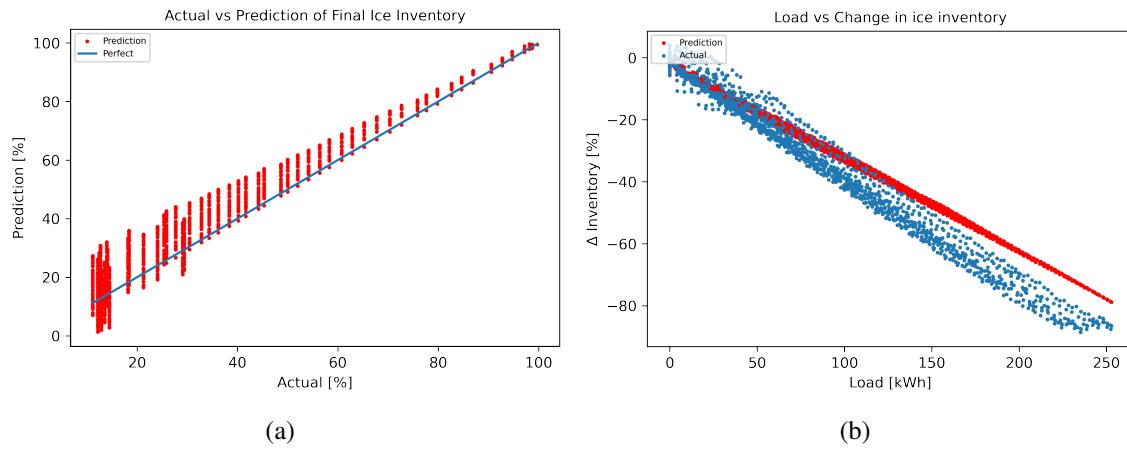


Figure 21. (a) Neural network model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) Neural network model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.

5.5. Gaussian Process Model Results

Figure 22 shows the results for the GP model for the charging mode. The prediction follows the actual inventory well, though it tends to under-predict at higher ice inventories. The change in the ice inventory tends to be under-predicted, though it does overlap with the actual change in inventory. These plots show the mean prediction for each point, but the GP model also returns the variance of the prediction. For these data sets, the model uncertainty is on the order of tenths of a percent (e.g., $60\% \pm 0.2\%$), which is much less than the measurement uncertainty, which is around an absolute $\pm 3\%$. This means that the predicted value, even with the variance, falls within the uncertainty of the measurement.

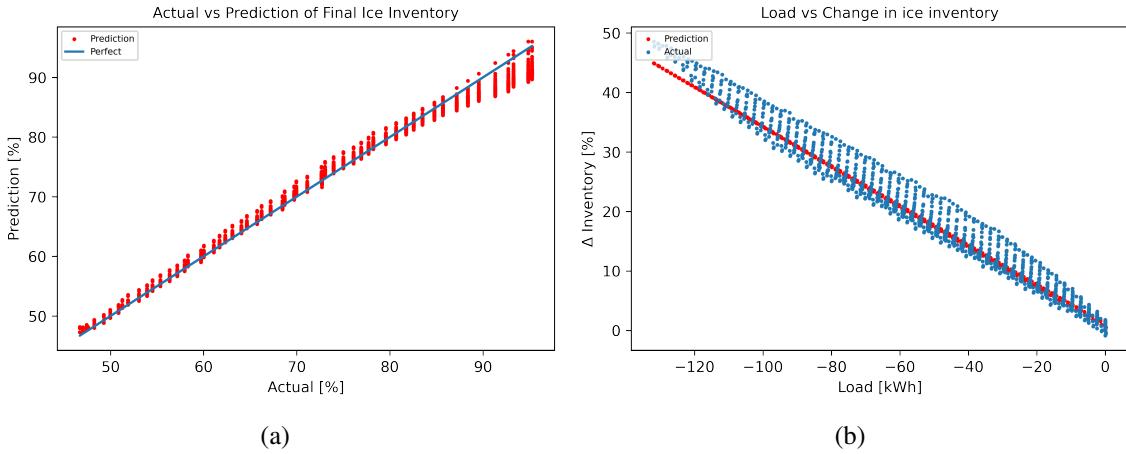


Figure 22. (a) GP model prediction of the ice inventory vs the actual ice inventory for the charging mode. (b) GP model prediction of the change in ice inventory as a function of integrated load for the charging mode.

Figure 23 shows results of the GP model for the discharging data. The model tends to over-predict the final inventory. Figure 24 shows the results for the smoothed discharge data.

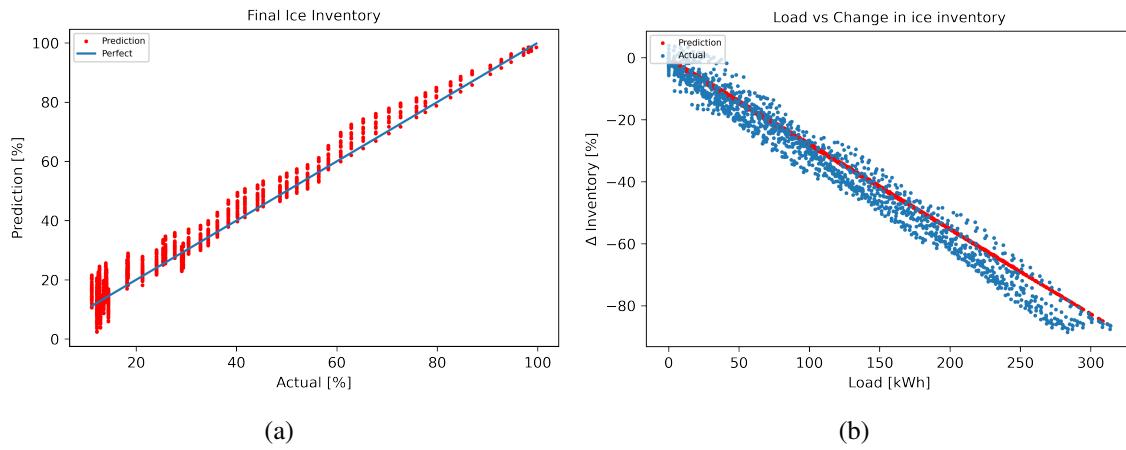


Figure 23. (a) GP model prediction of the ice inventory vs the actual ice inventory for the discharging mode. (b) GP model prediction of the change in ice inventory as a function of integrated load for the discharging mode.

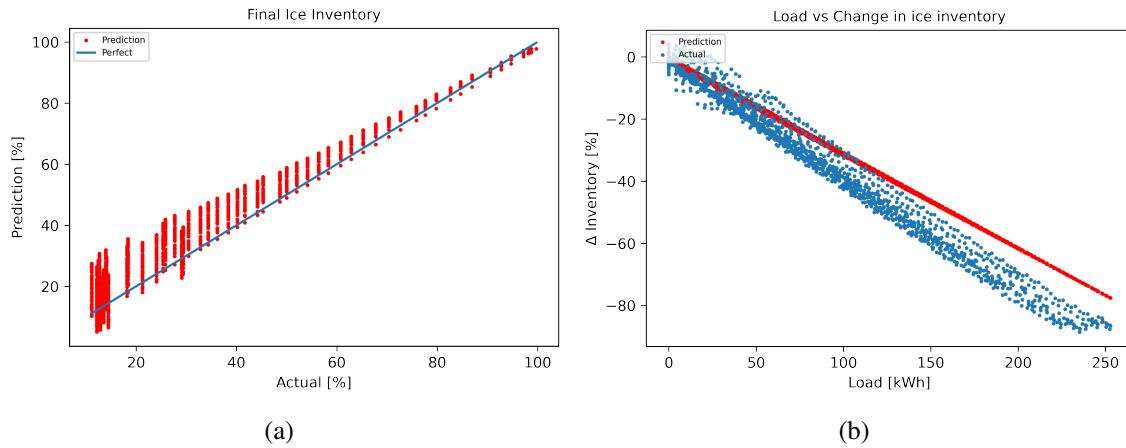


Figure 24. (a) GP model prediction of the ice inventory vs the actual ice inventory for the discharging mode - smoothed data. (b) GP model prediction of the change in ice inventory as a function of integrated load for the discharging mode - smoothed data.

6. Combined Model Results

After completing the preliminary models using a single charge or discharge data set for training, we decided to see if a single model could be used for both modes. Table 7 shows the training and testing data sets used for the models. In all cases the training data was a combination of the the discharge and charge data in Table 1, but the testing data was either the combination or the individual charge or discharge data set.

Table 7. Data sets used for the combined models.

Training	Testing
Filtered	Filtered
Smoothed and Filtered	Smoothed and Filtered
Filtered	Discharge Filtered
Smoothed and Filtered	Discharge Smoothed and Filtered
Filtered	Charge Filtered
Smoothed and Filtered	Charge Filtered

Table 8 shows the RMSE results for models trained on a combination of both charge and discharge data. In all cases the data have been filtered, but this table shows the difference when the data are also smoothed. The “Combined” column is the RMSE for a test data set that is the combination of the charge and discharge data sets used previously; the “Discharge” column is the result for the discharge test data set; the “Charge” column is the result for the charge test data set.

Table 8. RMSE of the models developed on a combination of charge and discharge data.

Model	Raw			Smoothed		
	Combined	Discharge	Charge	Combined	Discharge	Charge
Interpolation	1.76	2.42	1.54	2.4	4.16	1.48
LR	1.92	3.08	1.46	2.5	4.6	1.47
NN	2.48	2.7	2.4	2.2	4.3	1.04
GP	1.78	2.98	1.28	2.41	4.72	1.2

Recall that the smoothing process only impacted the discharge data set since the charge data set did not have the erroneous flow meter readings. However, the RMSE on the charge test data set does change with smoothing because now the training data set does contain the erroneous readings. This impact is minimal for interpolation, LR, and GP, but the RMSE for the NN applied to the charge test set is much lower when the data are smoothed. However, in all of the models, the RMSE increases for the discharge test set when the data are smoothed. This suggests that even when the data are filtered to remove nonlinearities, there is still a difference between the training and test sets for the discharge case. This

difference may be pointing to an issue with using a relatively small amount of data to train the models. A single data set, or a simple combination of one discharge and one charge data set, may not adequately capture the uncertainties, including measurement uncertainty, in the real system. This may result in a model that is not generalizable and is overfitting the data it is trained on, so when it is applied to other data sets it performs below expectations.

The overall conclusion, however, is that it is possible to create a single model that adequately captures the linear behavior in both charging and discharging modes. Having a single model simplifies the integration of the model into other calculations.

7. Discussion

In this study we built four different machine learning/statistical models to predict the final ice inventory level of an ice-on-coil TES given an applied load and initial ice inventory. The integrated load was calculated over all possible time durations for the entire period of the data set, thereby removing any temporal dependency in the data. Each sample set (initial ice inventory, integrated load, final ice inventory) is independent of every other sample set and they are identically distributed. This makes the data modelling easier because there are fewer correlations to account for in modelling. This section is a discussion of the various models, pointing out key details of the development procedure, advantages and disadvantages of each approach, and situations where each approach may be the most appropriate. The following sections address some of the key criteria for evaluating these models, including:

- convergence properties - how fast is the optimization and is it a global optimum,
- time complexity - number of function evaluations required to create/use the model,
- accuracy (RMSE) - see Eq. 10,
- ease of implementation,
- data set requirements - how much data and of what type,
- size of model class - flexibility to approximate different functions, and
- interpretability - is it apparent how the model makes a prediction.

7.1. Convergence

Models developed in this study include black-box models such as neural networks and a clear-box physics-based model. The typical process of building a mathematical/statistical model can be broken down into two stages, training and inference. In the training stage, a standard optimization procedure such as stochastic gradient descent is used to update the model parameters to best fit the training data. In most cases, mini-batch gradient descent is used instead of full-batch gradient descent because with a large data set, calculating the

gradients at each point is computationally expensive. In addition, since these gradients are unbiased estimators, evaluating them on a random sample of points in expectation is the same as evaluating gradient information for the entire data set. The loss function for linear regression is convex, so the convergence is not affected by the initialization of the parameters, but the initialization can cause problems for NNs and GPs because their loss functions are typically non-convex, and therefore they are prone to the pitfalls of iterative gradient methods. In the inference stage, the trained model is applied to a test input - i.e., prediction.

LR, with its convex loss function, has the advantage over NN and GP in terms of fast convergence to the global optimum, whereas NN and GP can take longer to converge and they may arrive at a local optimum. As applied in this study, interpolation and the physics-based model do not use optimization at all, so the convergence rate is not a factor.

7.2. Complexity

The processes of training and inference have different levels of complexity for each model. Linear regression requires approximately $O(n_{features})$ [14] function evaluations during training, whereas GP requires approximately $O(n_{samples}^3)$ for both training and inference separately. NN complexity is in between, as it takes about $O(n_{freeparameters} * n_{samples}) + O(n_{sgc})$ during training and $O(n_{freeparameters} * n_{samples})$ in inference. Interpolation and the physics-based model are both $O(n_{samples})$ and are therefore the least complex models.

7.3. Accuracy

Tables 5 and 6 summarize the accuracy results for all models for the discharge and charge modes. In general, the charge models are much more accurate than the discharge models. Interestingly, almost every model, with the exclusion of the physics-based model, has worse or the same accuracy when using the smoothed discharge data versus the original data with the incorrect flow rate measurements. For example, after smoothing the discharge mode data to replace the erroneous flow rate, the accuracy of LR was worse because the test data no longer looked like the training data. Refer back to Fig. 8; after smoothing the data sets, the slopes of the training and test data are different in part because of the inclusion of nonlinear data, resulting in a worse prediction. Something similar happens for all the other models except the physics-based model, which does not have a training process, and the NN, where the result was basically the same. However, in most cases the accuracy on the training data set improved, so cleaning the data did allow for a better model, but in this case the test data set no longer looked like the training data set. Accuracy improved in all cases when the nonlinear data were removed.

7.4. Ease of Implementation

The simplest models to implement are the physics-based model and linear interpolation. The main disadvantage of the physics-based model is that it requires some knowledge of

the thermal properties of water and the physical geometry of the TES. On the face of it neither of these requirements is difficult, however, the assumptions include that the thermal properties are constant and that there are no deviations from the manufacturer specifications of the TES due to installation issues. Linear interpolation, on the other hand, requires no assumptions about the system, but it does require sufficient data, which in this case are available. However, this is one area where the physics-base model has the advantage; it is not limited to data within the range of the training data set. LR is also relatively simple to implement using pre-existing library calls. The NN requires some trial and error to select the architecture, but there are pre-existing library functions that make it otherwise straightforward for the user. The same can be said of the GP model.

7.5. Summary of Each Model

Table 9 lists the advantages and disadvantages of each model. The following sections go into more detail for each model.

Table 9. Model Comparison Table

Model	Advantages	Disadvantages
Physics-based	- Uses theoretical knowledge - Requires a single model for charge and discharge modes	- Requires knowledge of the total mass of water - Assumes thermal properties are well known
Linear Interpolation	- Simple - Interpretable	- Cannot extrapolate beyond the range of data it is trained on
Linear Regression	- Interpretable - Convex loss function	- Sensitive to noisy data
Neural Network	- Rich model class - Wide array of regularization options - Non-parametric approach - Useful when prior knowledge of data is not known	- Not interpretable - Can easily overfit to training data - No statistical guarantees on test performance
Gaussian Process	- Probability distribution over parameters instead of point estimates - Non-parametric approach - Easy to incorporate prior knowledge about model parameters	- More complicated - Requires more computational resources

7.5.1. Physics-Based Model

The physics-based model has mixed results in terms of accuracy. It does as well as the other models for the charging mode, but it is worse than all the other models on the original

discharge data, though it does better on the smoothed and filtered discharge data set. This latter result is likely because unlike the other models, which rely on the training data to either learn the parameters of a function or to create a lookup table, the physics-based model treats all data independently. In this implementation, it does not fit the test data based on the training data. In the charge mode case, the models all perform about the same, probably because the data are clean to begin with and the underlying physics are linear. The main disadvantage of the physics-based model is that it requires knowledge of the physical properties of the system - the melting/freezing properties of water and the total mass of water in the TES, which can be taken from manufacturer specifications. However, the actual installation of the system can differ from the off-the-shelf system. In the IBAL, for example, the TES was incorrectly installed, so the total volume of the water in the TES is less than the manufacturer specified volume, but the magnitude of that difference is unknown. Slight differences between the manufacturer specifications and the assumed properties of water at a given temperature and pressure could cause some error in the physics-based model. However, a training procedure could be used to learn the optimal values for the properties rather than relying on specifications.

7.5.2. Linear Interpolation

Linear interpolation is an interesting case because the key factor is not the size of the data set, but rather the range of data. If the training data covers a wide enough range of possible values, then, for modelling an inherently linear system, linear interpolation is the best choice. It is highly interpretable and it requires no training since it is just a lookup table. As you can see in the error values for the training data in both the charging (Table 6) and discharging cases (Table 5), the interpolation model has the lowest error relative to the other models. However, it can not extrapolate beyond what the user chooses to allow. For example, the user could set the values beyond the limits in the table to the mean, zero, or some user-defined function. Using this model requires knowledge of the data this model might see in the future, any outlier data and the functional form of the underlying physics (e.g. linear interpolation vs polynomial interpolation). If the training data cover the range of all possible values and the functional form of the relationship matches the type of interpolation model (e.g. linear), interpolation works well. This is evident from the fact that the linear interpolation actually does better than the physics-based model using the original data. However, it does worse when using smoothed data. One reason might be that by smoothing the data the maximum range of the data is reduced, which limits the model's interpolating capability.

7.5.3. Linear Regression

The relationship between the ice inventory level and the integrated load for an ice-on-coil TES is linear, so linear regression is an obvious approach for modelling. In addition to modelling linear systems well, it has been widely studied and has very good statistical properties. As previously mentioned, one such property is that the loss function is convex

and so optimization results in the global optimum. However, unlike NN, LR is very sensitive to distributional shifts in the training and test data. LR came out of the statistical community, so it has stronger assumptions regarding the data distributions it encounters in both training and testing. This is evident from the results in Table 5 - although the RMSE increases for both NN and LR, it increases more for LR when the data are smoothed and the test data no longer tracks the training data well.

7.5.4. Neural Network

The NN is a step up the ladder of complexity from LR. This is a rich class of models; a NN with the right architecture and parameter values [15] can be used to model any function. NNs are often used in situations where there is a lack of domain knowledge about the data and the goal is to fit a function to a set of input-output pairs without knowing the relationship between them. NN is a black-box model and there is no framework to help select the right model architecture. It is often a process of trial and error, although this can be done more cleverly using techniques such as grid search and Bayesian optimization. In addition, being a black-box model, it can be difficult if not impossible to interpret how it makes predictions, so NN scores low on interpretability. For example, the NN performs worse for the charge mode than the other models, but there is no tangible way to understand why. Recent work in ML has focused on improving this flaw, so future work may include using an explainable NN approach. Unlike in LR, the loss function for a NN is typically non-convex, so it may optimize to a local optimum (though this can be mitigated with a sufficiently large NN [16]) or, even worse, a saddle point. However, unlike LR and interpolation, NNs can be applied to a wide array of equipment where the underlying physics are not linear. They also have more potential for modeling transient behaviors. Finally, although NNs can be sensitive to distribution shifts and noise in the training and test data, they are less sensitive than LR because of their more complex architecture.

7.5.5. Gaussian Process

Of the models we studied, GP is at the top of the ladder of complexity. GPs have the ability to express model uncertainty explicitly. One advantage of using a GP is that it can be non-parametric; the architecture can be changed during the training process. For example, layers can be continuously added on top of one another until the loss function is minimized. Each additional layer of GP means passing the output of previous layers through another function. For example, if layers 1, 2, and 3 model functions f_1 , f_2 , f_3 , respectively, then adding layer 2 to layer 1 would be $f_2 \circ f_1$ (composition of functions). Unlike in LR, where the parameter values are adjusted to minimize the loss function, here layers can be added or removed to reduce the loss function. As the GP is entirely defined by its mean and kernel, changing the kernel means an entirely different set of functions can be modelled, which makes GPs very versatile. In addition, kernels can be combined to model more complex sets of functions, including periodic functions. As with LR, GPs came out of the statistical community and therefore have stronger assumptions on the data distributions of

the training and testing data, which is why the RMSE for the smoothed discharge data was so much higher than on the original data (see Table 5).

8. Conclusions and Future Work

We developed five different ice-on-coil thermal energy storage tank models in this study. One of those models was a simple physics-based model, and the other four were data driven models built using real laboratory data. These models all have advantages and disadvantages that make them more or less appropriate for different applications. For this TES, in which the change in ice inventory is a linear function of integrated load and sufficient data are available, the best overall option is linear interpolation. This approach does not require assumptions about the thermal properties of water and the physical geometry of the TES, so in many ways it is simpler than the physics-based model. Interpolation was also almost always more accurate than the physics-based model. The NN and GP approaches are more complicated to implement than the other options, so they are not as appropriate for this TES, but they are likely to be more appropriate for more complex systems, especially if there is a need to model transient behavior. For the TES, although most of its operation is in the linear region, if there is a need to model the nonlinear data these more complex approaches are necessary. They can each model complex functions. GP has the advantage if there is a desire to have some understanding of the model uncertainty. In all cases, it is important to properly clean the data to achieve the best results.

This study raised several questions that we did not have time to fully investigate, but might be worth future study. We downsampled the data from every 10 s to every 10 min. This time frame was selected based on observations - the ice inventory does not change beyond the measurement uncertainty of the measurement in 10 s, but over 10 min it can change more than the uncertainty if the integrated load is sufficient. The effect on model accuracy of different downsampling periods and techniques (i.e., selecting every nth point vs using the mean over the downsampling period) should be further investigated. There is also a question about the effect on the model of any imbalance in the data set. For example, in the data used here, the integrated load has a value of zero more often than any other value; what is the impact of that imbalance?

A detailed study should be conducted to understand how much data is necessary to create a truly generalizable model that captures the uncertainties of the real system. Removing erroneous data through smoothing should have improved the model results, but in most cases it did not. Some of the issue was due to the inclusion of nonlinear data in a linear model, but that does not seem to fully explain the issue. We would also like to focus on the nonlinear regions to see how well the existing models capture that behavior and if it is necessary to build a separate model for those regions. It should be possible to build a NN or nonlinear GP to model both the linear and nonlinear behaviors. A better understanding of the data distributions of the charge and discharge data sets can also help determine how much data might be required to produce generalizable models. It is important to

remember that any approach that increases the number of samples will require changing the architecture of the NN, and using a variational approximation for the GP since the costs of training and inference for the GP are $O(n_{samples}^3)$.

References

- [1] NIST (2021) Intelligent building agents project. March 29, 2023 Available at <https://www.nist.gov/el/energy-and-environment-division-73200/intelligent-buildings-agents-project>.
- [2] Pertzborn A, Veronica D (2018) Intelligent building agents laboratory: Air system design. [https://doi.org/https://doi.org/10.6028/NIST.TN.2025](https://doi.org/10.6028/NIST.TN.2025).
- [3] Pertzborn A (2016) Intelligent building agents laboratory: Hydronic system design. <https://doi.org/https://doi.org/10.6028/NIST.TN.1933>.
- [4] Lee AH, Jones JW (1996) Modeling of an ice-on-coil thermal energy storage system. *Energy Conversion and Management* 37(10). Available at <https://www.osti.gov/biblio/653177>.
- [5] King DJ, Potter RA Jr (1998) Description of a steady-state cooling plant model developed for use in evaluating optimal control of ice thermal energy storage systems. *ASHRAE Transactions* 104(1A). Available at <https://www.osti.gov/biblio/653177>.
- [6] Henze GP, Krarti M (2005) Predictive optimal control of active and passive building thermal storage inventory (US DOE, Germantown, MD), DE-FC26-01NT41255. <https://doi.org/10.2172/894509>. Available at <https://www.osti.gov/servlets/purl/894509>
- [7] Ooka R, Ikeda S (2015) A review on optimization techniques for active thermal energy storage control. *Energy and Buildings* 106:225–233. <https://doi.org/https://doi.org/10.1016/j.enbuild.2015.07.031>. SI: IEA-ECES Annex 31 Special Issue on Thermal Energy Storage Available at <https://www.sciencedirect.com/science/article/pii/S0378778815301420>
- [8] Tarragona J, de Gracia A, Cabeza LF (2020) Bibliometric analysis of smart control applications in thermal energy storage systems. a model predictive control approach. *Journal of Energy Storage* 32:101704. <https://doi.org/https://doi.org/10.1016/j.est.2020.101704>. Available at <https://www.sciencedirect.com/science/article/pii/S2352152X20315413>
- [9] Henze GP, Kalz DE, Liu S, Felsmann C (2005) Experimental analysis of model-based predictive optimal control for active and passive building thermal storage inventory. *HVAC&R Research* 11(2):189–213. <https://doi.org/10.1080/10789669.2005.10391134>. <https://www.tandfonline.com/doi/pdf/10.1080/10789669.2005.10391134> Available at <https://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391134>
- [10] Henze GP, Felsmann C, Knabe G (2004) Evaluation of optimal control for active and passive building thermal storage. *International Journal of Thermal Sciences*

- 43(2):173–183. <https://doi.org/https://doi.org/10.1016/j.ijthermalsci.2003.06.001>. Available at <https://www.sciencedirect.com/science/article/pii/S1290072903001327>
- [11] Touretzky CR, Baldea M (2014) Integrating scheduling and control for economic mpc of buildings with energy storage. *Journal of Process Control* 24(8):1292–1300. <https://doi.org/https://doi.org/10.1016/j.jprocont.2014.04.015>. Economic nonlinear model predictive control Available at <https://www.sciencedirect.com/science/article/pii/S0959152414001164>
- [12] NIST (2023) Ibal database. August 30, 2023 Available at <https://ibal.nist.gov/>.
- [13] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis (3rd ed.)* (Chapman and Hall/CRC).
- [14] scikit-learn developers (2023) Scikit-learn linear models. April 26, 2023 Available at https://scikit-learn.org/stable/modules/linear_model.html#ordinary-least-squares.
- [15] Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5):359–366. [https://doi.org/https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/https://doi.org/10.1016/0893-6080(89)90020-8). Available at <https://www.sciencedirect.com/science/article/pii/0893608089900208>
- [16] Choromanska A, Henaff M, Mathieu M, Arous GB, LeCun Y (2014) The loss surface of multilayer networks. *CoRR* abs/1412.0233. [1412.0233](https://arxiv.org/abs/1412.0233) Available at <http://arxiv.org/abs/1412.0233>.