Machine Learning Techniques Applied to Sensor Data Correction in Building Technologies

Matt K. Smith[1], Charles C. Castello[2], and Joshua R. New[2]
[1]Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, Alabama 35487 USA
[2]Energy and Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831 USA

Building Technologies Research at ORNL

- Residential and commercial buildings consume 41% of the primary energy (72% electricity) in the U.S.
- Retrofitting inefficient buildings with new and innovative technologies that help to curb energy consumption will reduce environmental impacts.

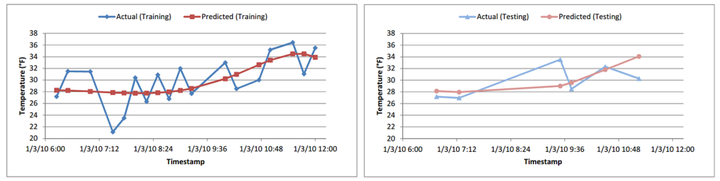
ORNL's Flexible Research Platforms (FRPs)

What is the problem?

- Hundreds of sensors collect data in these buildings to develop and characterize performance.
- ORNL's decades of experience with real-world natural exposure facilities has logged many expected issues as well as unforeseen entropy-related failures.
- These include power backup outages, data acquisition kernel panics, multiplexer timing failures, sensor wire disconnects, failing sensors, and many other issues.
- Decisions based on faulty data could lead to inaccuracies in simulation algorithms or when analyzing components, systems, and whole-buildings for optimal retrofit.

What is the solution?

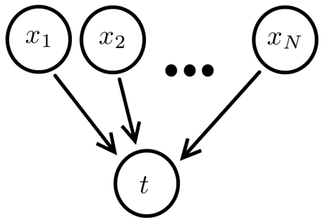
- Modeling techniques are used with validated data to generate models.
- These models can be used to predict data points with missing or inaccurate data.
- Observation windows are used to predict a sensor's data value for each time-step within the observation window.
- The observation window moves forward with no overlap for every possible window within a given set of time-series sensor data.



Example of actual and predicted temperature values of training (left) and testing (right) subsets.

Artificial Neural Networks (ANN) and Bayesian Networks

- ANNs and Bayesian networks are used to predict the data of a target sensor using other sensors' data as inputs.
- ANNs are used to predict data with polynomial regression models.
- Each ANN consists of one hidden layer and ten hidden nodes.
- The Levenberg-Marquardt backpropagation algorithm is used to train the ANN.



$$P(t|x_1, x_2, ...x_N) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The Bayesian network that is used

- The Bayesian network uses each sensor as a node.
- The Bayesian network produces a multivariate Gaussian distribution for a range of likely target values.

ZEBRAlliance Data

- Data from ORNL's ZEBRAlliance project is used to determine the predictive performance of both methods.
- Specifically, temperature, humidity, energy usage, pressure, and airflow sensor data is used from house #2 (of four) during the 2010 calendar year.

Sensor Characteristics

Data Type	Average	STD	Min	Max
Temperature (°F)	62.30	13.86	17.03	91.00
Humidity (%RH)	64.91	19.84	17.00	106.00
Energy Usage (Wh)	12.77	11.32	0.00	125.50
Pressure (psi)	273.09	77.34	0.41	500.90
Airflow (ft^3)	0.46	2.43	0.00	70.72

ZEBRAlliance House #1

How are experiments setup?

- Data from the ZEBRAlliance project are randomly separated into two subsets: training (70%) and testing (30%).
- For each observation window, a model is generated using the training dataset.
- The generated model is then used to predict time-steps where data points are located as if they were missing.
- Root-mean-square error (RMSE) is calculated to determine predictive accuracy to the known data.
- Five to eight other sensors deemed most relevant are used as model inputs to predict the five target sensors.
- Input sensors were chosen using the Joint Mutual Information (JMI) feature selection technique.

What are the results?

List of sensors to model temperature sensor

Priority	Sensor Name	Type	Description
1	H20_PR_LiqAir_Avg	Pressure	IHP Liquid Line
2	Z12_Tm_ERV_OUT_Avg	Temperature	ERV Exhaust
3	Z09_RH_ERVin_Avg	Humidity	ERV Intake
4	Z12_RH_ERVout_Avg	Humidity	ERV Exhaust
5	OD_Air_Avg	Temperature	Outside
6	WindDir_D1_WVT	Direction	Wind Direction
7	E22_SF_Avg	Temperature	Wall
8	E23_SF_Avg	Temperature	Wall

List of sensors to model liquid flow sensor

Priority	Sensor Name	Type	Description
1	A04_WH_ERV_Tot	Energy	ERV
2	NI_DAQ_Tot	Energy	Pulse Counter
3	Z13_WF_Shower_Tot	Water Flow	Shower
4	H32_Wh_IHP_comp_Tot	Energy	IHP Compressor
5	E71_Wh_Housetot_Tot	Energy	Whole House
6	E70_Wh_IHP_tot_Tot	Energy	IHP Total Energy
7	H35_WH_WHelem_Tot	Energy	Heater Elements

List of sensors to model energy usage sensor

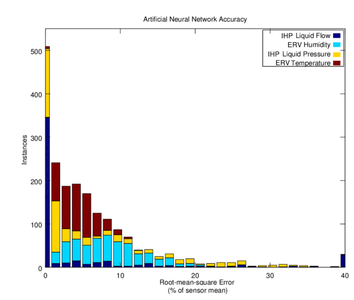
Priority	Sensor Name	Type	Description
1	WindDir_SD1_WVT	Direction	Stan. Dev.
2	OD_AIR_Avg	Temperature	Outside
3	Z11_Tm_ERV_Return_Avg	Temperature	ERV Return
4	E10_SF_Avg	Energy/Area	Whole House
5	A04_WH_ERV_Tot	Energy	ERV
6	WindDir_D1_WVT	Direction	Wind Direction
7	Z10_RH_ERVsup_Avg	Humidity	ERV Supply

List of sensors to model pressure sensor

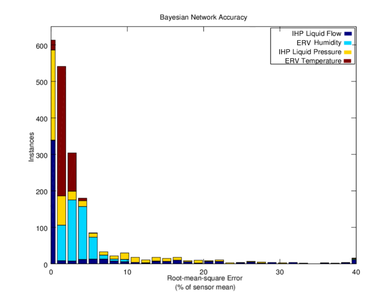
Priority	Sensor Name	Type	Description
1	H32_Wh_IHP	Energy	IHP Total
2	H20_PR_LiqAir_Avg	Pressure	IHP Liquid Line
3	OD_AIR_Avg	Temperature	Outside
4	Z11_Tm_ERV_Return_Avg	Temperature	ERV Return
5	E10_SF_Avg	Energy/Area	Whole House
6	E70_Wh_IHP_tot_Tot	Energy	IHP Total Energy
7	Z12_RH_ERVout_Avg	Humidity	ERV Exhaust

List of sensors to model humidity sensor

Priority	Sensor Name	Type	Description
1	Z10_RH_ERVsup_Avg	Humidity	ERV Supply
2	OD_AIR_Avg	Temperature	Outside
3	H21_PR_LiqBrine_Avg	Pressure	IHP Liquid Line
4	E10_SF_Avg	Energy/Area	Whole House
5	Z09_Tm_ERV_IN_Avg	Temperature	ERV Intake



RMSE of 496 ANNs



RMSE of 496 Bayesian networks

What does this mean?

- Bayesian networks outperformed ANNs for all sensors.
- Both techniques were effective with temperature, humidity, pressure, and liquid flow data.
- Refrigerator energy is harder to predict due to large spikes in energy consumption when the compressor is operating and when the refrigerator doors are opened.

	psi	%RH	°F	gal	Wh
ANN	17.7	5.0	1.9	0.03	0.89
Bayes	13.5	2.1	1.1	0.02	0.85
Difference	4.2	2.9	0.8	0.01	0.04

Average RMSE between ANNs and Bayesian networks

Acknowledgements

- Research sponsored by the Laboratory Directed Research and Development (LDRD) Program (WN12-036) of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy (DE-AC05-00OR22725).
- This work is also supported in part by the U.S. Department of Energy, Office of Science.