Survival Modeling for Offshore Wind Prognostics

Phase II Final Written Documentation

Prepared for:

National Offshore Wind Research and Development Consortium (NOWRDC),
New York State Energy Research and Development Authority (NYSERDA),
Massachusetts Clean Energy Center (MassCEC)

Julian Fraize
Program Manager, NOWRDC

Prepared by:

Tagup Inc.
361 Newbury Street, Boston, MA 02115

Rory Polera
Director, Projects

Sam McCormick
Lead Data Scientist

Kyle Richards
Data Scientist

Jon Garrity
Founder and CEO

Notice

Abstract and Keywords

Wind turbine (WT) failure is expensive and hard to predict. Offshore WTs are decidedly less accessible than onshore WTs, with inspection and maintenance significantly more expensive. Wind curtailment is a reduction in power generation (measured as active power) due to alarmed operating states and/or component failure (i.e., mechanical, or electrical). Curtailed power generation results in a loss in operating revenue when compared to modeled/forecasted revenue, thus increasing the cost of ownership, and reducing the economic efficiency of wind power (i.e., LCOE). Using a combination of static, time series and labeled event data together with machine learning (ML), we deployed a combination of supervised time-to-event (TTE) models and unsupervised drift and changepoint detection algorithms that identify/detect anomalous operating states leading to curtailment and/or downtime events for six operational wind farms (WFs) in North America (approximately 1 GW of rated capacity across 447 WTs). The models and algorithms incorporated over 3,000 turbine-years of historical operating data from SCADA data systems and included over 40 sensor measurements recorded every 10 minutes. Outputs update daily and are available via secure user application, application programming interface (API) and email notifications (as a daily digest and weekly summary). A combination of data scientists, performance and reliability engineers and operators/maintenance technicians can receive the outputs on a recurring basis to inform maintenance planning and inventory management/procurement decision making. Like Phase I, we expect the outputs to improve offshore wind maintenance planning and reduce annual O&M spend by up to 10%.[1]

Machine learning, artificial intelligence, anomaly detection, drift, changepoint, failure prediction, predictive maintenance

[1] We need an additional three to six months to run more thorough forward testing (which unfortunately extends beyond the funding allocation of this agreement).

Acknowledgments

We are grateful for the National Offshore Wind Research and Development Consortium (NOWRDC), the Massachusetts Clean Energy Center (MassCEC), the New York State Energy Research and Development Authority (NYSERDA) and the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office for funding this important work.

We are also grateful for the in-kind support of Pattern Energy as they provided the data and O&M expertise, without which this study would not be possible.

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

NOWRDC	National Offshore Wind Research and Development Consortium
NYSERDA	New York State Energy Research and Development Authority
MASSCEC	Massachusetts Clean Energy Center
WT	Wind Turbine
LCOE	Levelized Cost of Energy
ML	Machine Learning
TTE	Time-to-event
WF	Wind Farm
MW / GW	Megawatt / Gigawatt
SCADA	Supervisory Control and Data Acquisition
API	Application Programming Interface
O&M	Operations and Maintenance
GE	General Electric
OEM	Original Equipment Manufacturer
WMS	Work order management system
CLV	Convex Latent Variable model
CIGRE	International Council on Large Electric Systems
MFCC	Mel-Frequency Cepstral Coefficients
CUSUM	Cumulative Sum
ROC-AUC	Receiver Operating Characteristic-Area Under the Curve
RUL	Remaining Useful Life
KPI	Key Performance Indicators

Executive Summary

The Massachusetts Clean Energy Center (MassCEC) and National Offshore Wind Research and Development Consortium (NOWRDC) provided funding to develop, deploy, and demonstrate novel machine learning (ML) methods as decision support tools for improving offshore wind maintenance planning.[2] Funding was divided over two phases, Phase I and Phase II, with a critical go/no-go decision at the end of Phase I. Phase I focused on the development and generalization of ML-based time-to-event (TTE) and economic optimization models, demonstrating sufficient signal in the data to predict critical events and generate value.[3] Phase II focused on deploying the best-performing models and demonstrating the efficacy of value creation for the deployed outputs in an operational environment.[4] This report summarizes a non-confidential version of all work completed under Phase II and demonstrates the technical feasibility of deploying novel ML methods as decision support tools for improving wind maintenance planning.

Key observations and findings from Phase II include the following:

1. We successfully engaged two offshore wind development partners to test/evaluate the ML analytics on operating WTs in both Europe (Phase I) and North America (Phase II).[5]
2. We successfully generalized the deployment of our technology from offshore to onshore wind farms (WFs)—Teesside → Spring Valley/Broadview/etc. and vice versa—albeit differences in wind turbine (WT) design, installation, event frequency and use.[6]

---

[2] As measured by a reduction in equipment failures, increase in capacity factor and reduction in LCOE.

[3] In Phase I, we trained over 2,100 ML-based TTE models that indicated a 5-10% reduction in annual operation and maintenance (O&M) spend for two offshore wind farms operated by EDF Renewables in the United Kingdom (32 WTs; 102 MW capacity).

[4] In Phase II, we deployed unsupervised ML methods to identify anomalous operating states at six of Pattern Energy's onshore wind farms (WFs) in North America (Pattern Energy is a new development partner primarily in North America with ~8.6 GW onshore and offshore wind projects including 5.6 GW currently in operation). The outputs update daily for the 447 Siemens turbines with 1 GW generating capacity.

[5] In Phase I we worked with EDF Renewables at two offshore WFs in the United Kingdom (with limited historical data, 45 turbine-years) to demonstrate sufficient signal in the data for predicting critical events using ML. In Phase II, we worked with Pattern Energy at six onshore WFs in North America (with nearly 100x the amount of historical data used in Phase I; specifically, 3,000 turbine-years) and deployed the analytics on updating data in an operational environment.

[6] The underlying microservices were used to deliver the capabilities for both offshore and onshore WFs: Ivaldi, Saga, Ran, Vidar, Baldr and Midgard for data collection and ingestion (Ivaldi), data indexing and management (Saga), model development and management (Ran & Vidar), model deployment and visualization (Vidar, Baldr, Midgard). The modeling techniques (available in Vidar) and associated results/value stories varied in Phase I and II depending on

3. We successfully deployed unsupervised ML methods (namely, drift and changepoint detection) on six onshore WFs in North America (447 Siemens turbines equivalent to 1 GW installed capacity) with a possible extension to 2,500 WTs (in total), pending satisfactory value creation via forward testing.[7]
4. We anticipate the value proposition identified in Phase I (i.e., up to 10% annualized O&M savings) will generalize from offshore wind farms in Europe to onshore (and eventually offshore) WFs in North America; however, we need to evaluate the value proposition via forward testing over longer time horizons (during an additional three to six months), which unfortunately extends beyond the funding allocation of this agreement. As we are already deployed, more information will be available in the future.

As we look beyond Phase II, we plan to continue supporting Pattern Energy by demonstrating the expected value on a going forward basis with the intention of expanding to additional WFs (up to 25 WFs in total; 2,500 WTs from Siemens, GE, Vestas and Mitsubishi; ~6 GW of generating capacity).[8]

---

the target user stories, historical data volumes and overarching technical approaches (all intended to achieve the same analytic objectives).

[7] We incorporated over 3,000 turbine-years of historical operating data to train/tune/configure the algorithms. Outputs update daily and are available via secure user application, API, and email notifications (as a daily digest and weekly summary).

[8] Pattern Energy subscribed to a prorated annual license in December 2022 for our suite of software tools which is set for renewal in December 2023.

1. Project Objectives

WT failure is expensive and hard to predict. Offshore WTs are decidedly less accessible than onshore WTs, with inspection and maintenance significantly more expensive. Wind curtailment is a reduction in power generation (measured as active power) due to alarmed operating states and/or component failure (i.e., mechanical, or electrical). Curtailed power generation results in a loss in operating revenue when compared to modeled/forecasted, thus increasing the cost of ownership, and reducing the economic efficiency of wind (i.e., LCOE).

Using a combination of static, time series and event data together with ML, we worked with Pattern Energy to deploy both supervised TTE models and unsupervised drift and changepoint detection algorithms that identify/detect anomalous operating states leading to curtailment and/or downtime for six operational WFs in North America (447 Siemens WTs; ~1 GW installed capacity). The models and algorithms aim to link failure risk for key events (e.g., capacitor failure) to anomalous operating states at the component (e.g., converter, accumulator, main bearing), asset (WT1), and fleet level (Spring Valley). The key motivating questions we set out to address, include the following:

1. Can we accurately predict curtailment/downtime events such that maintenance actions can be planned just-in-time to minimize downtime and reduce the LCOE of wind power?
2. Of the events that lead to downtime, which "actionable" events/alarms can we accurately predict (e.g., delta module failure)?
3. How does model accuracy vary with prediction horizon (i.e., 1 day, week, 1 month)?
4. How much data is required to accurately predict downtime events (e.g., 5 years, 1 year)?
5. What data types are required to accurately predict downtime events (e.g., 10-minute Supervisory Control and Data Acquisition (SCADA) averages)?
6. Can we generalize downtime event predictions to other wind farms (e.g., North America) of different equipment, wind profiles and design?

Funding was divided over two phases, with a critical go/no-go decision at the end of Phase I. This report summarizes a confidential version of all project work completed under Phase II.

2. The Data

2.1 Demonstration Sites

Pattern Energy ("Pattern") develops, constructs, owns, and operates high-quality wind, solar, transmission, and energy storage projects worldwide. Pattern owns and/or operates over 1,500 Siemens WTs which accounts for more than 3.5 GW generating capacity.[9] Pattern has operationalized a centralized data lake for operating data (in Azure), such that model outputs can easily be deployed for evaluation at an individual or across multiple WFs in North America and/or around the world.

Since kicking off Phase II, we incorporated operating data from six WFs in North America. Figure 1 shows the location of each wind farm. The oldest WF was commissioned in 2013 (Spring Valley) and the newest in 2019 (Broadview Grady). The six WFs are made up of 447 WTs (equivalent to 1 GW of installed capacity), all of which were manufactured by Siemens. A breakdown of technical specifications for the six WFs is in Table 1.

Figure 1: Map of the six onshore wind farm locations in North America: Spring Valley, Amazon, Armow and Broadview (which is made up of three installations: Grady, KW and JN). Five (of the six) WFs are made up of Siemens 2.3 MW WTs whereas Broadview Grady is comprised of Siemens 2.625 MW WTs. In total, the six WFs account for 447 WTs and approximately 1 GW of installed capacity (which is nearly 8% of Pattern's installed WTs and over 15% of their operating capacity).

[9] Pattern has ~8 GW of onshore and offshore wind projects, with 5.6 GW currently in operation. Pattern owns/operates equipment from Siemens (3.5 GW), GE (2 GW), Vestas and Mitsubishi.

Table 1: A summary of the data integrated for the six WFs used for field testing. We incorporated metadata, time series data and event data from all WFs based on what was made available. Various outputs (depending on the data integrated) are available via API for users to access on an ongoing basis. The red "x" over Amazon indicates a data processing issue resulting in unavailable data flows/streams.

Description / Data Source	Data Type	Spring Valley Available (Y/N)	Broadview Grady Available (Y/N)	Broadview JN Available (Y/N)	Broadview KW Available (Y/N)	Armow Available (Y/N)	Amazon Available (Y/N)
Manufacturer		Siemens	Siemens	Siemens	Siemens	Siemens	Siemens
Commissioning Year		2013	2019	2017	2017	2015	2015
Number of Turbines		66	84	79	62	91	65
Rated power		2.3 MW	2.625 MW	2.3 MW	2.3 MW	2.3 MW	2.3 MW
Rated Capacity		152 MW	221 MW	182 MW	143 MW	209 MW	150 MW
SCADA & Weather	Time Series	Y	Y	Y	Y	Y	Y
Maintenance Logs	Event Data	Y	Accumulator Replacements Only			Y	N
Alarms	Event Data	Y	Y	Y	Y	Y	Y

The six WFs were chosen as they are all self-performed (and/or planning to be self-performed in the near future).[10] As each WF ages it will continue to challenge Pattern with both expected and unexpected maintenance issues (motivating the use of ML-based analytics).

2.2 Data Integration

Pattern provided direct access to their SQL database via Azure Synapse in December 2022. This database consolidates operating data across their global install base (including the six self-performed WFs), appending new data as it is received. Operating data is available dating back to January 2014 and includes more than 3,000 turbine years of historical operating data for the six self-performed WFs. The data available maps to the following 3 groups:

1. **Metadata**, which includes static attributes like WT manufacturer, location and rating.
2. **Time series data**, such as sensor data from the SCADA system or weather data from the metrology system.
3. **Event data**, such as work orders associated with planned and unplanned maintenance events (as found in the work order management system, or WMS), alarm data and labeled downtime as captured in the SCADA system.

[10] "Self-performed" is a term Pattern uses to label WFs that are operating outside the warranty period, such that Pattern operators, technicians and engineers are responsible for maintaining the equipment (i.e., Pattern and their investors own the failure liability/risk outright, as it is no longer covered by a complex Master Service Agreement, or MSA with the original equipment manufacturer, or OEM).

The data volumes for all six WFs (ranging from 12 to 111 continuous measurements) flow into our data pipeline for preprocessing, ingestion, inference, and visualization using a combination of our core microservices.[11] Figure 2 depicts how these microservices interact from raw data ingestion (Ivaldi → Saga) through model visualization in the application (Midgard).

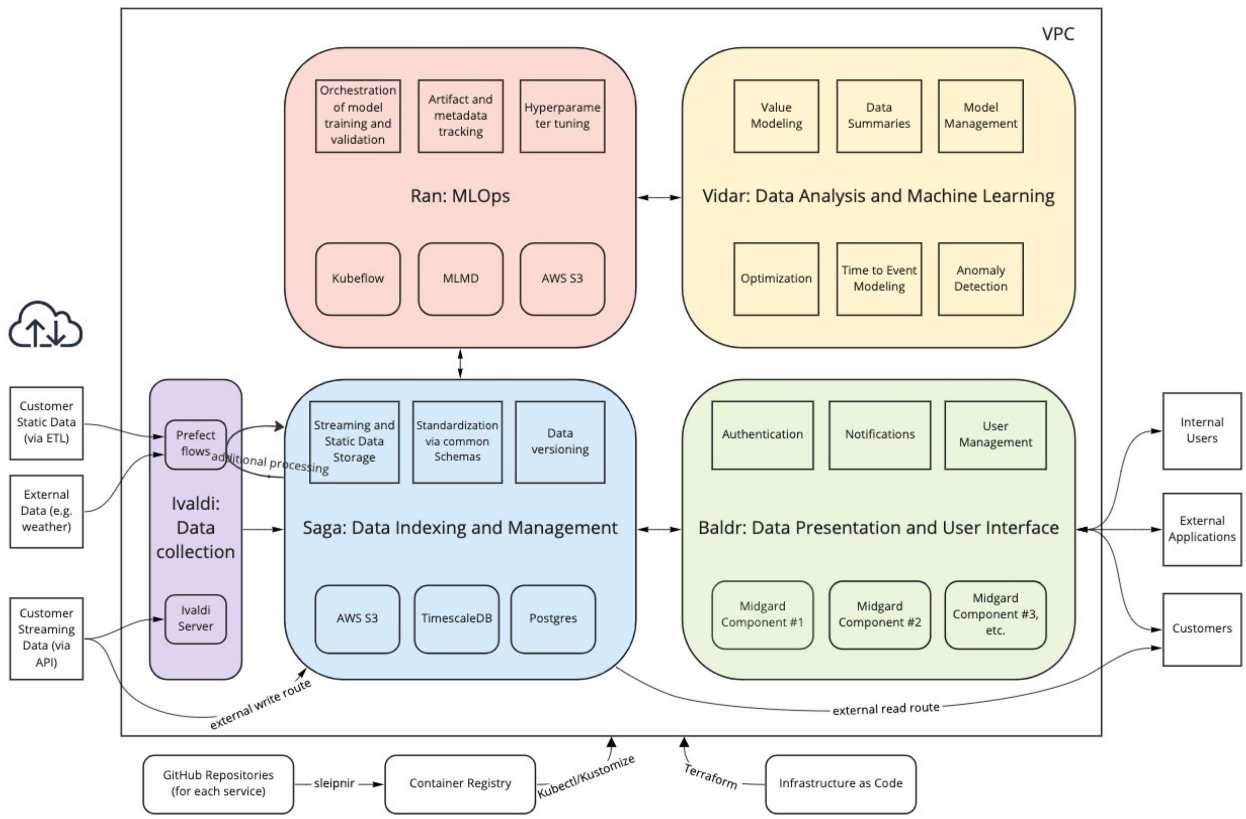


Figure 2: System architecture depicting how data flows between the respective microservices. The microservices deployed include Ivaldi, Saga, Ran, Vidar, Baldr and Midgard for data collection and ingestion (Ivaldi), data indexing and management (Saga), model development and management (Ran & Vidar), model deployment and visualization (Vidar, Baldr, Midgard). Additional details for each of the microservices can be provided upon request.

[11] Additional details on the feature set available and integrated in Saga (12 to 111 continuous time series measurements) is available upon request.

Live data was first configured for Spring Valley on January 13, 2023. After a few additional weeks of testing, debugging and user onboarding, time series data with initial model outputs are live, updating every day starting at 0900 UTC (4:00 am EST). New time series data triggers the following sequence of events daily with no more than 2 hours latency for all six WFs (from start to finish; moving raw data from Azure to *Midgard* with model outputs available):

1. 0900 UTC: The last day (i.e., prior 24 hours) of time series data is uploaded to Saga. This takes about 15 minutes.
2. Immediately following the conclusion of step 1, any new events from the last day of data are detected and uploaded to Saga. In parallel, custom engineered features are calculated and uploaded to Saga. This step takes around 10 minutes.
3. Immediately following the conclusion of step 2, model inference runs, and results are uploaded to Saga. This takes about 4 minutes.
4. 1100 UTC: The front-end application refreshes its cache key so that new data will be loaded from Saga as seamlessly as possible.

This process was generalized and deployed across the five additional WFs on February 13, 2023.[12] As new models were trained, validated, and deployed, the inference pipeline was updated accordingly.[13] Besides configuring updating data (via Saga and Midgard), all historical time series measurements were backfilled for data exploration and model validation purposes (depending on the model that was deployed).[14]

[12] The initial data integration took five weeks from start to finish (i.e., for Spring Valley we integrated data in Saga, trained, validated, and deployed the first model using Vidar and Ran and made all cleaned data and initial model outputs visible in Midgard from December 22, 2022, to January 20, 2023).

[13] This was (and still is) an ongoing process as we modify our technical focus (based on what we learn) throughout the analytics evaluation period.

[14] Some outputs don't require backtesting, and therefore are only evaluated on a forward testing basis (e.g., Pulse outputs for drift detection, changepoint detection, anomaly detection, etc.). Refer to 4. Methods & Results for more details.

3. Technical Approach

Our technical approach is centered on two core capabilities: supervised ML (commercially known as Foresight) and unsupervised/semi-supervised ML (commercially known as Pulse). Technical objectives for each capability include:

- **Foresight**: Estimate failure risk of critical events (i.e., converter failures) by modeling time to event (TTE) at the component, asset, and fleet level. Aggregate failure risk to estimate parts demands for capacitors and/or accumulators.
- **Pulse**: Classify, detect and/or predict normal and anomalous operating states, starting with temperature drift for one WT relative to the install base, normalized for wind speed, load, etc. Generalize the approach to measurements other than temperature (e.g., pressure, RPM, electrical) and identify times with a step change in operating state and/or performance (relative to the other WTs).

3.1 Foresight

Foresight is centered on the convex latent variable (CLV) survival model.[15] Derived from standard survival modeling techniques, the CLV model is designed to make uncertainty-aware predictions of asset failure probabilities as a function of historical data. The CLV model uses a unified history of static attributes, time series data and labeled event data to estimate both instantaneous and cumulative contributions to a particular asset's probability of failure over time.

We use distributed computing to perform feature extraction and model training/validation at scale.[16] The CLV model uses a convex loss function which allows for highly efficient and scalable model training. This function is weighted to eliminate bias that results from unbalanced classes such as more observations of non-failure periods vs. failure examples. The CLV model learns to predict a hazard function based on a *conditional* likelihood formulation for each asset given its operating history. This formulation encodes the predicted probability of failure at each

[15] Garrity, J, et al. "Latent variable survival models for network transformer prognostics." CIGRE Grid of the Future, 2018.

[16] Feature extraction refers to techniques that transform data via mathematical methods such as Mel Frequency Cepstral Coefficients (MFCC) and continuous wavelets. These methods borrow concepts from signal processing and audio recognition to identify and categorize different frequency ranges observed in the input time series data. Intuitively, abnormalities in data are assumed to be correlated with phenomena that can lead to undesirable events. Data that is not transformed may be referred to as the *identity*.

timestep given survival until the previous timestep.[17] Because the CLV model is specified as a convex optimization problem, the training procedure is guaranteed to find the globally optimal parameters in a relatively short amount of time compared to non-convex formulations (e.g., deep learning-based approaches).

While the hazard function is not directly useful, statistics of interest are derived from it, including event probability over any time horizon (a tunable parameter required by Pattern). Figure 3 shows the predicted curtailment event probability curves in the next 1 week for two offshore WTs. The top WT did not fail within the 3-month window, whereas the bottom WT did. This model uses 18 time series measurements, averaged every 10-minutes to construct a hazard function for each asset (not pictured). The variability in the hazard function is a result of the temporal changes in the underlying SCADA data. From the underlying hazard functions, curtailment event probabilities can be derived which allows us to rank order assets by failure risk and optimize preventative maintenance planning.

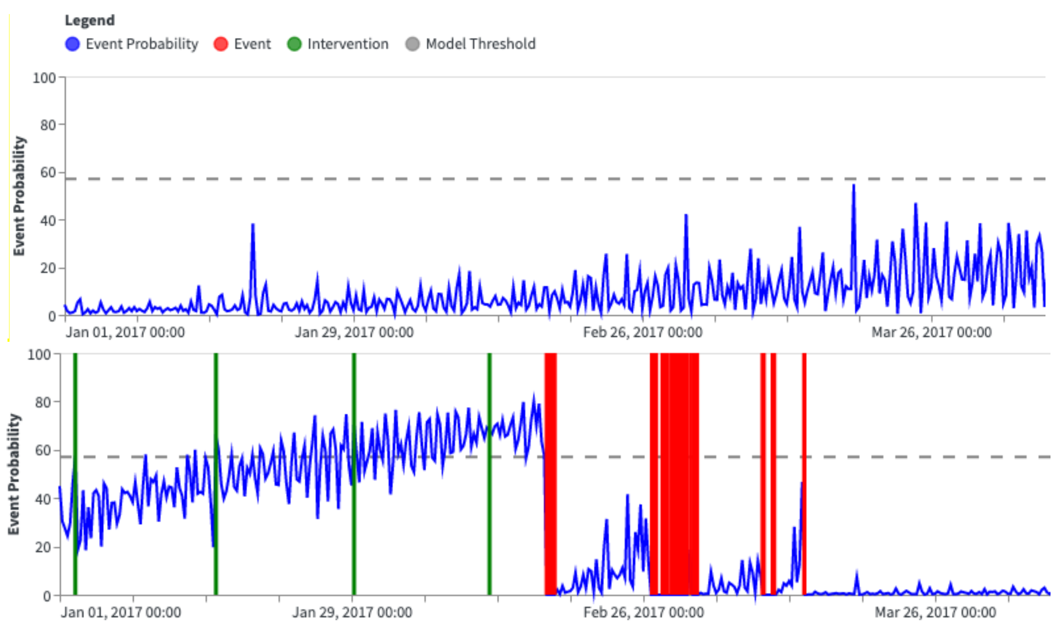


Figure 3: Curtailment event probabilities for a non-failure and failure. The model threshold is based on average event and intervention costs captured in the value model (from Phase I).

[17] Traditional likelihood formulations use all information observed up until the moment of prediction to predict a *hazard function*, which is then compared to the pattern of events to compute likelihood (a good model should maximize hazard at event times and minimize it elsewhere). However, this formulation doesn't accommodate advance notice of an impending event which is strictly required for prognostics to be actionable by an O&M team. Thus, we have reconfigured the likelihood formulation to be conditioned on past data rather than referencing all data up to time-to-event. This modification significantly improved the model's predictive power.

3.2 Pulse

Pulse, on the other hand, is a group of algorithms that are used to detect active issues in real time (i.e., temperature drift, changepoint, outliers/anomalies which if left untreated, can result in an event, like downtime). For example, we can detect drift of one WT measurement relative to the rest of the WF using a simple and interpretable algorithm like CUSUM. CUSUM operates on Z scores (signals normalized to have 0 mean and unit variance) and uses two parameters, Omega and Threshold, to vary sensitivity.[18] When we take the Z score across WTs, it can tell us when a single WT has unusually high values with respect to the rest of the WF for a sustained period. To be practically useful, we condition CUSUM on wind speed so that we're computing outputs only for periods when there is sufficient wind to generate power, and the WT is operational.[19]

Relatedly, we can train a model using moving averaging methods to detect a changepoint; that is, the algorithm detects when a SCADA measurement for a WT has a step change (either, up or down) in value.[20] Pulse outputs are useful in helping operators and engineers synthesize large volumes of data (i.e., hundreds of SCADA measurements, thousands of alarms and 3rd party outputs) to pinpoint active issues at a specific WT, relative to the others across the WF. Refer to Figure 4 for what this looks like in practice.

---

[18] Omega is a cutoff, below which the signal detracts from, rather than contributes to, evidence for upward drift. Since the input signal is a Z score, this would correspond to sections that are less than 2 standard deviations from the mean. Threshold determines how long the signal must exceed omega for an event to be triggered. Think of the threshold like a bucket that fills up over time as the signal exceeds omega. Basically, it tells us when a signal is significantly above its mean for a sustained period, i.e., the mean is "drifting" away from where it should be.

[19] We exclude periods when the wind speed ≤ cut-in wind speed and/or ≥ cut-out wind speed.

[20] We started by using a simple averaging method by computing the ratio of moving average over a small window to moving average over a much larger window and selecting a threshold to identify changepoints. A model is trained on multiple years of historical operating data to capture seasonality and other temporal variance that informs changepoint detection.

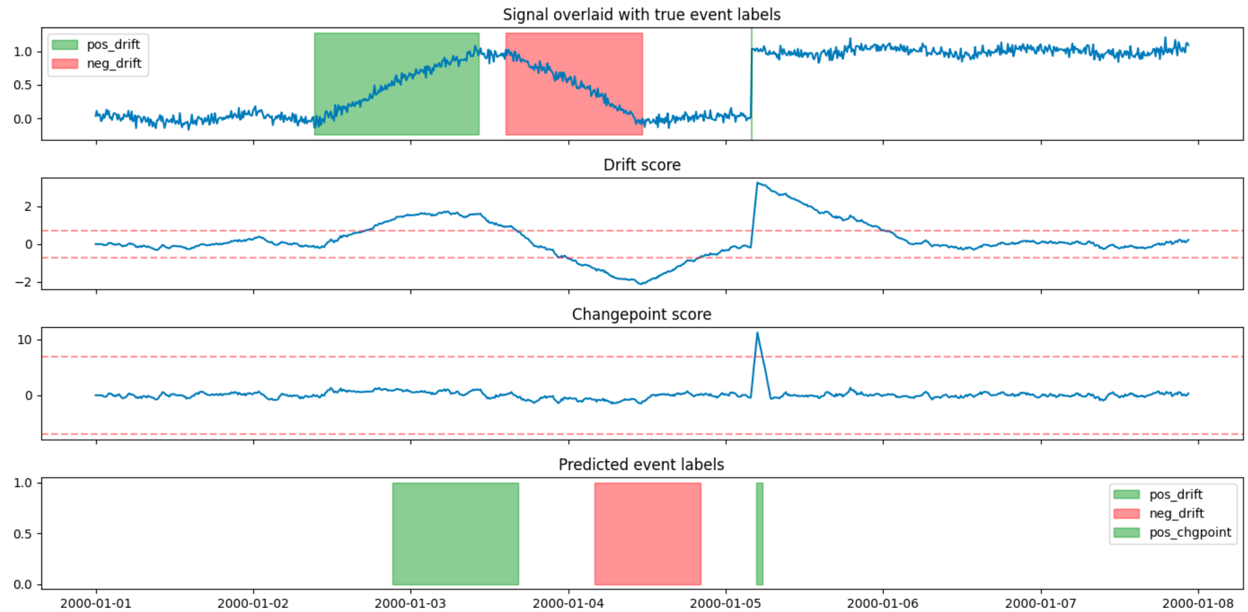


Figure 4: Drift and changepoint events for a mock time series measurement. Drift and changepoint events can be labeled positive or negative, as indicated by the legend. Tunable parameters like Omega and Threshold (red dashed lines) can be modified to adjust the sensitivity of the algorithm.

4. Methods and Results

4.1 Data Ingress and Wrangling

After data ingress (*Ivaldi*) but before injection (*Saga*), data is cleaned and wrangled for schema conformance. These actions are scripted, and may include a combination of the following:

- **Combine disparate data**, such as data extracts stored in separate files or tables.
- **Remove erroneous data**, which may include values labeled as "BAD".
- **Impute missing data**, such as NaNs or "-", and replace with zeroes.
- **Synchronize timescales** between features where required for modeling.

Once completed, cleaned data (whether its historical or updating daily) is injected into *Saga* where historical events may be labeled for model training and validation (using *Ran* and *Vidar*), and inference is run on a recurring basis as new data is received/processed.

4.2 Model Inference Pipeline

The best performing model generates predictions on the updating data via *model inference*. Data is continuously ingested (*Ivaldi*) into the data pipeline via live integration with the source data. Inbound data is cleaned and wrangled to conform with a predefined schema. The schema may tolerate certain data deficiencies such as missing timestamps. However, the model inference pipeline may fail if a feature used in the model is missing. In these cases, an error will automatically notify developers via Slack. Cleaned data is stored in *Saga* and passed to *Ran* and *Vidar* which collectively orchestrate model inference. These results are written back to *Saga* and passed to *Midgard* for display in the application. This flow is depicted in Figure 5.

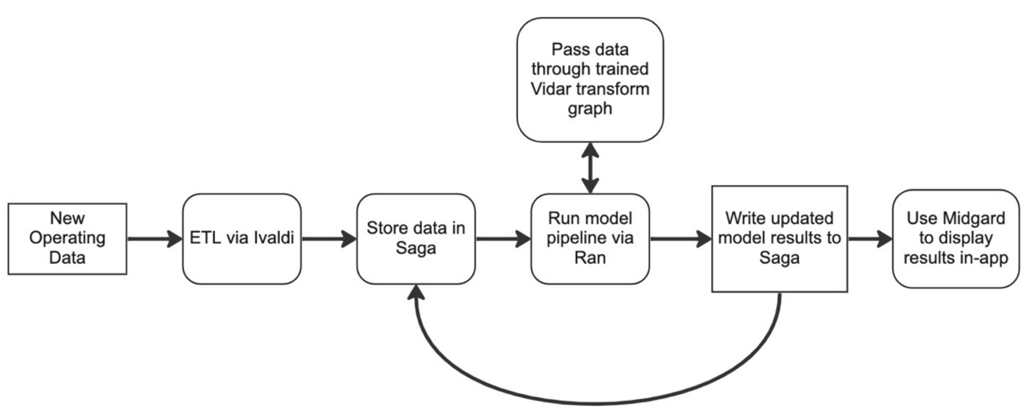


Figure 5: A process flow diagram of the model inference pipeline from receipt of new data through display of updated outputs in application.

4.3 Analytic Methods

4.3.1 Foresight Methods

During training/validation of a supervised ML model (i.e., Foresight), we use backtesting, a process in which we train a model using only information available (training dataset) before a fixed date in the past (conditioning date). We then make predictions about the time between the conditioning date and the present and use the withheld data (validation dataset) to evaluate those predictions. Foresight utilizes Katib (orchestrated by Ran) to efficiently tune hyperparameters to enhance model performance.[21] Performance may be quantified with a combination of discriminative and absolute validation metrics. We use the 1-week ROC-AUC as the primary validation metric to quantify and compare model performance for each experiment.[22]

4.3.2 Pulse Methods

Because Pulse outputs (for both drift and changepoint) are unsupervised in nature, it is very difficult to "validate" the predictive capacity of an algorithm that doesn't use labeled historical events in the training/fit process. As such, we evaluate by hand (or inspection) how well the Pulse output aligns with performance issues (i.e., precedes larger issues like SCADA alarms and/or downtime events). For that reason, we incorporated synthetic drift to verify the algorithm picks up on useful signals.[23]

4.4 Analytic Results

We trained/validated and/or deployed (iteratively) over 200 models and/or algorithms depending on the target technical approach listed in Table 2. Outputs are available on an ongoing basis for Pattern operators and engineers to evaluate and recognize value accordingly.

---

[21] Katib is an open source automated hyperparameter tuning library. Katib allows a user to define a "feasible space" for model hyperparameters, and then runs a series of trials, each with a single fixed hyperparameter configuration. Katib uses a Bayesian optimization algorithm to efficiently and intelligently search the hyperparameter feasible space in an iterative manner, using the results of previous trials to guide the search.

[22] ROC-AUC refers to "receiver operating characteristic-area under the curve". ROC-AUC measures a model's ability to correctly classify an asset's status (event or no event, in our case). Like the concordance index, scores range from 0 to 1. A ROC-AUC of 1 represents a perfect classification model, whereas a score of 0.5 indicates the model's predictive capacity is equivalent to random guessing.

[23] More details regarding how we injected synthetic drift can be made available upon request.

Table 2: A summary of the models/algorithms tested to date. This summary includes the data labeling methods (indicated by Primary Data), technical approach applied (albeit Foresight or Pulse), status of model (e.g., Complete and Deployed, vs not), scope of the approach/outputs including best performing model results (i.e., model objective based on the event of interest), and the event count used to evaluate model performance/efficacy.

Primary Data	Technical Approach	Status	Scope/Method	Event Count[24]
Maintenance records & Alarm records	Foresight	Complete	Estimate p(fail, 7 days) for converter events • Maintenance events on the converter (from WMS) • All downtime events with any alarm active (from SCADA data, rules) ○ Active power ≤ 0 MW FOR ≥ 1 hour WITH ○ Wind speed ≥ 3m/s AND ≤ 25 m/s (> 50% of the time)	Tens to tens of thousands
Alarm records	Foresight	Complete	Estimate p(fail, 7 days) for converter-related downtime events (≥ 1 hour).[25] The best performing model had an AUC-ROC = 0.6.[26]	Hundreds to thousands
Maintenance records	Foresight	Incomplete; Insufficient Data	Estimate number of capacitor failures/replacements expected in 90 and 180 days at Armow (Spring Valley dataset shared but incomplete; missing Broadview)	Tens
Maintenance records	Foresight	Complete; Insufficient Data	Estimate number of accumulator failures/replacements expected in 90 and 180 days at Broadview wind farms. The initial model outperformed the two baselines by 18% on the 3-month horizon and 14% on the 6-month horizon.	Hundreds
Maintenance records	Foresight & Pulse	Received data 8/8/2023	Estimate main bearing Remaining Useful Life (RUL) for WTs with main bearing vibration alarms active.[27]	Tens
SCADA data	Pulse	Deployed (Accessible via API for 5 WFs, missing Amazon) Backfilled as of 3/1/2023 Available in application for SVW	Identify drift for 1 WT relative to others (both dense measurement and list of events) • Started with 5 temperature measurements • Expanded to 41 primary measurements (10-minute averages) w/ the ability to add 33 more measurements (primarily diagnostics)	Depends on tunable parameters (**Ω** and threshold). Range: Tens to hundreds in 1 year for 1 WF.
SCADA data	Pulse	Deployed (Accessible via API for 5 WFs, missing Amazon). Backfilled as of 3/5/2023	Changepoint detection on all measurements • Started with ambient temperature—units switched from °C to °F • Expanded to 3 specific measurements for the Gearbox and Converter. • Can generalize to 41 primary measurements	Depends on tunable parameters (Window Sizes, Threshold). Tens in 1 week for all 5 WFs

[24] Specific event counts were redacted for confidentiality purposes.

[25] As defined by ≥ one 13xxx alarm w/in 1 hour before the start of the event.

[26] This result suggests there is signal in the data; however, we need more failure examples to drive up model accuracy and ultimately operationalize the value in the predictions.

[27] This is an evolving event scope that will continue beyond the term of this grant agreement. We will likely use both Foresight and Pulse to identify when a main bearing issue is present (from the Pulse drift algorithm) followed by an estimate of main bearing RUL (from a Foresight model trained/validated on historical examples). Operators and engineers can use these outputs (if accurate enough) to effectively plan when maintenance should optimally occur (every main bearing alarm has a different duration until the main bearing eventually fails and/or is replaced), and therefore mitigate WT derating (which is also known as intentional curtailment). There is a lot of pressure to effectively plan main bearing replacements in periods of low wind (i.e., especially during the summer).

Initial Foresight models were trained and validated to predict converter-related TTE (using our CLV model), however the value proposition was unclear how Pattern operators and engineers would preemptively intervene to mitigate events/issues thus realizing the value of accurately predicting converter-related downtime events, let alone general downtime events writ large. Since the initial models were deployed (earlier in Phase II, Task 9), we modified the model formulation to aggregate counts of parts required at varying times in the future (namely, 90 and 180 days), as global supply chain issues challenge Pattern's sourcing team to get the right parts at the right time on the shelves (with minimum inventory) while meeting O&M needs.[28] The best performing Foresight model (under the original converter-related event definition) was initially deployed for Spring Valley (in Task 9 to demonstrate the capability/pipeline functionality) but is currently unavailable in the application until we receive more data to estimate parts requirements for accumulators at 90 and 180 days in the future.

For Pulse, we started with a very basic and more interpretable approach, using CUSUM and moving averages to detect drift and changepoint, respectively.[29] For drift, we tuned omega and threshold to find the best fit before deploying in production. We deployed both a dense output (CUSUM score as function of time) and a list of events detected by the algorithm (Omega = 2 and Threshold = 200), on an ongoing basis (highlighted in green, Table 3).[30] An aggregation of the drift event counts by WT for all WFs (for a fixed omega and threshold) is available for six temperature measurements in the Fleet Explorer (to demonstrate capability). The dense underlying signal (for a fixed omega = 2) is available in the Asset Explorer for Spring Valley and via API for all 41 measurements across all six WFs. Furthermore, a table of all drift and changepoint events (for fixed omega and threshold values) for the downselected 41 SCADA measurements is available via API as March 1, 2023.

---

[28] Both the accumulator replacements at Broadview and capacitor datasets were incomplete for training/validating Foresight models at scale.

[29] We also tested more sophisticated techniques like a modified version of Kats (modified to be parallelizable for model training and deployment at scale—across thousands of WTs), but deployed CUSUM first (with input from Pattern) to simply demonstrate value in the approach before adding more complexity with less interpretability.

[30] We backfilled CUSUM outputs as of March 1, 2023, so we have some historical data to evaluate algorithm performance. The list of events can be queried to rank order by count of events by WT.

Initial feedback (from forward testing) suggests that Pulse drift events show promise in accurately identifying when a main bearing failure is starting to occur (when compared to a 3rd party vibration dataset). Pattern engineers indicated that if we can accurately estimate when the main bearing will fail from the point of drift/degradation detected (i.e., RUL of the main bearing from vibration/drift to failure), there will be material value to realize.[31] Given these initial results, we have a business case and opportunity to combine the Foresight and Pulse models to provide value to Pattern operators and engineers on an ongoing basis. As we continue (beyond the scope of this agreement), we will continue to evaluate the efficacy of the Foresight and Pulse outputs for main bearing, capacitors, accumulators, and/or other subsystems of interest.

[31] Initial analysis shows that main bearings are replaced anywhere from 0 to 800 days after significant vibration is initially detected, with two clusters of failure/replacements around 4 months and 9 months. To maximize the operating window of the WT, and effectively plan the replacement, operators will derate the WT and run it at an intentionally curtailed output (up to 50% of its rated operating capacity). Over time and across a large install base this can account for material lost production. Initial estimates indicate that tens of WTs were derated during a 27-month operating window resulting in hundreds of thousands in lost revenue (based on power curve modeling for the expected power output at the recorded wind speed).

5. Value Model

5.1 Initial Business Opportunity

Wind service contracts are typically established with WT OEMs for at least the first 5 years under a warranty period. During the warranty period, the developer/owner relies on the OEM for standard asset O&M with the OEM accountable for several terms and KPIs, including asset uptime and operational availability (i.e., time-based, or production-based availability).

During the warranty period, the developer/owner has access to operational data and OEM technical expertise but typically has limited resourcing and in-depth expertise to monitor and evaluate OEM O&M decision-making. Third party monitoring services are available to provide an independent view on asset operating health, however there is no simple way to benchmark/evaluate the overall value delivered by the OEM during the warranty/contracted period. Relatedly, the developer/owner has limited access to transfer O&M expertise from the OEM to engineers/technicians/operators should they decide to bring O&M in-house upon expiration of the warranty period. Typically, in this scenario an expensive service level agreement is maintained with the WT OEM to provide technical support on an ad-hoc basis.

5.2 Expected Business Value

Building analytic models that accurately forecast asset downtime enables wind developers like EDFR and Pattern to effectively build in-house O&M expertise upon commissioning new assets. Analytic outputs, like the probability of main bearing failure by WT (conditioned on drift detection), can be used to compare WT uptime and availability vs. standard OEM O&M and better plan maintenance on low-wind days. These model outputs (among others) can ultimately be used to evaluate the value delivered by the OEM within the warranty/contracted period.

Given the complex, contractual relationships that can exist between the developer/owner and OEM/service provider, we will continue to work with Pattern Energy to better understand the value story as a function of implementation and sustainment costs, especially as it relates to the main bearing failure mode indicated in Section 4.4 Analytic Results (combining the capabilities developed throughout this grant agreement: Foresight and Pulse). We anticipate the value identified in Phase I (i.e., up to 10% annualized O&M savings) will generalize from offshore wind farms in Europe to onshore (and eventually offshore) WFs in North America; however, we need to evaluate the value proposition via forward testing both Foresight and Pulse over longer time horizons (during an additional three to six months), which unfortunately extends beyond the funding allocation of this agreement.

5.3 Operational Value Measurement & Verification

As we move beyond this grant agreement, we will continue to "evaluate " the accuracy of the models and/or algorithms deployed and identify the expected utility on an ongoing basis (to compare with the ongoing sustainment costs). Measurement and verification, in partnership with Pattern operators and engineers, will include the following:

- Identifying WTs at highest risk of failure
- Assessing key data to confirm accurate identification
- Operationalizing the outputs and key decision criteria for maintenance and repair with WT operators and maintenance technicians, and
- Quantifying the associated value proposition on an ongoing basis (to compare to the cost of deployment/sustainment)

Although Pulse outputs (both drift and changepoint) are unsupervised in nature, we will continue to investigate historical operating data to identify how well Pulse outputs align with performance-related issues (i.e., how often do Pulse events precede larger issues, like SCADA alarms and/or downtime events), and specifically how we can use Foresight in combination with Pulse to better identify and estimate TTE.

6. Product Deployment

Foresight and Pulse outputs were packaged and deployed iteratively using our core microservices. We summarize the deployment process into two categories of service:

1. **Application**: The application uses a series of our microservices to serve users updated outputs referencing real-time operating data and the best performing model outputs.
2. **Email Notifications**: In cases where another portal/user application is not preferred, users can access the information from the underlying microservices (namely, the model inference pipeline) in a consolidated email that summarizes on a daily and/or weekly basis the events detected by the models and/or algorithms.

6.1 Application

The application (represented as *Midgard* components) is connected to the model inference pipeline (via *Saga*) to serve users updated model outputs referencing real-time operating data. Visualizations were updated for users to assess both Foresight and Pulse outputs in real time. The application supports fleet level views in the *Fleet Explorer* as well as tailorable comprehensive operational histories for a single WT via the *Asset Explorer*.

The Fleet Explorer may be used to examine asset metadata and aggregated model outputs in a dynamic table for the entire install base (across one WF, or all six). Users may sort, filter, and pivot raw data and model outputs, which automatically update the geospatial visualization and fleet KPIs, accordingly (top of Figure 6). Notably, the dynamic table may be used to generate a rank-ordered list of WTs by Pulse events detected (bottom of Figure 6). Alternatively, event counts and/or durations may be visualized as a choropleth by color (Figure 6). This bird's eye view may facilitate daily or weekly maintenance planning or reporting to management.

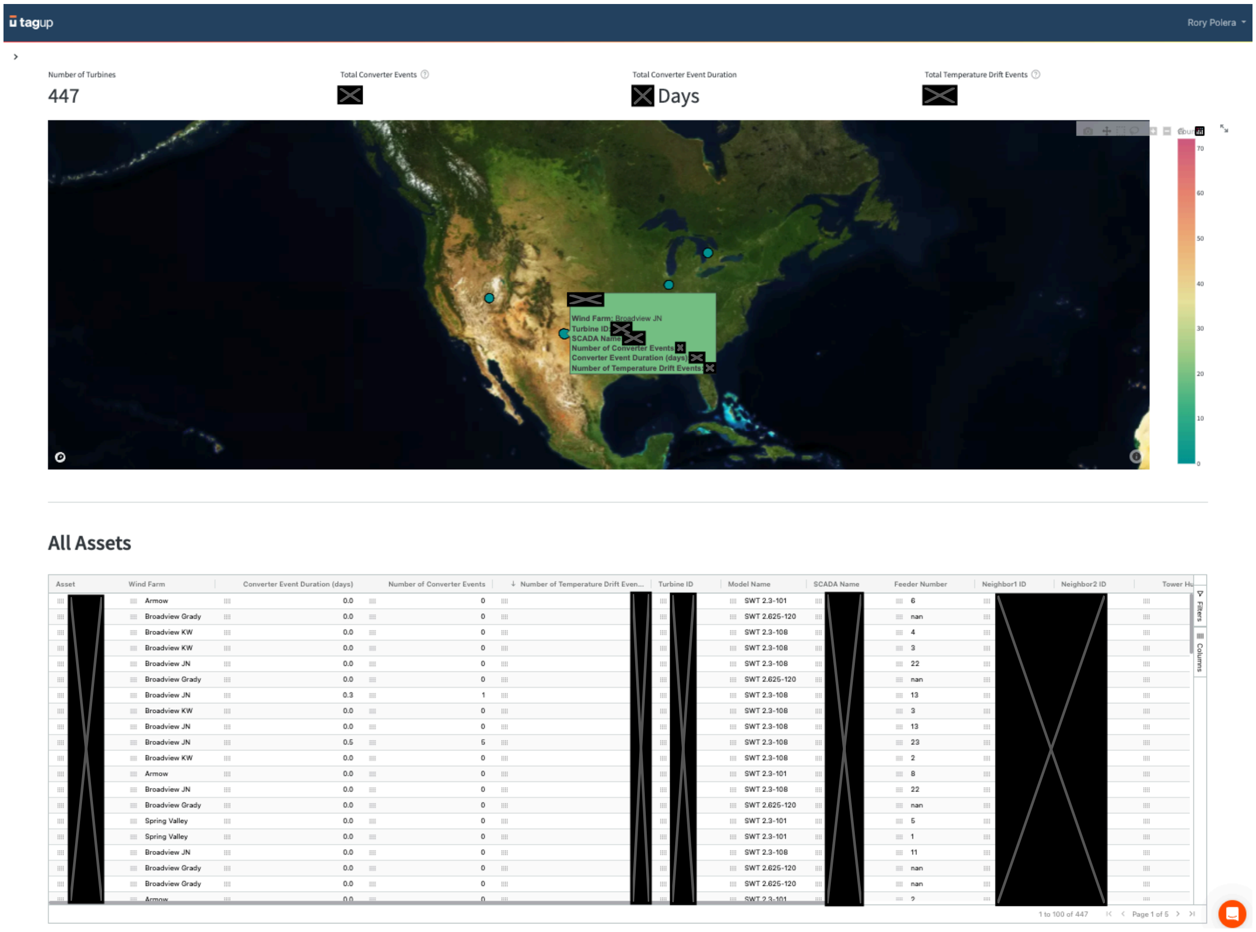


Figure 6: Fleet Explorer with Pulse events detected in aggregate as a KPI and individually in the choropleth map and tooltip. The dynamic table (at the bottom of the screen) may be sorted to generate a rank ordered list of WTs by output selected. Data can be exported (as csv, for example) by right clicking over the table.

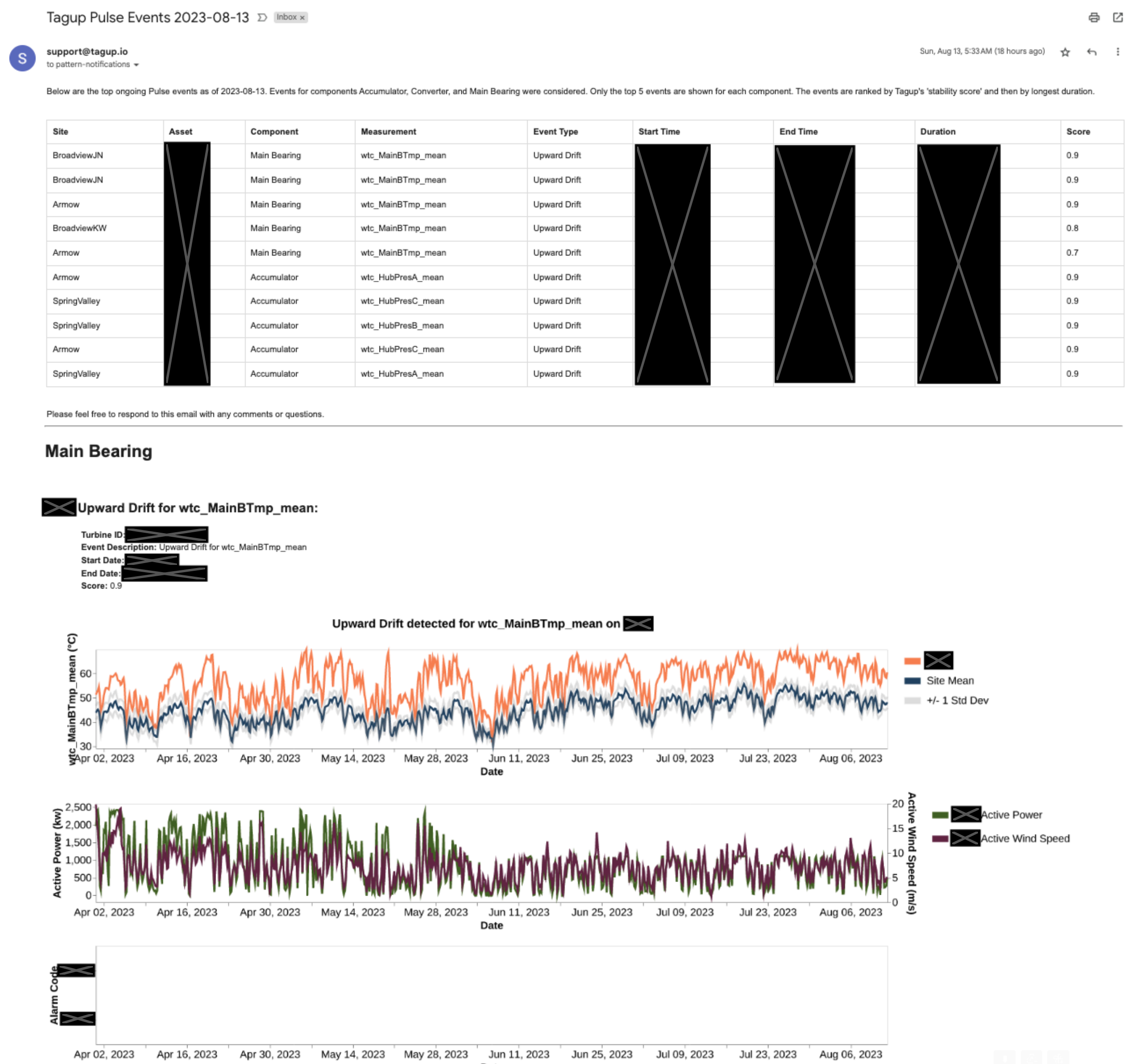
The Asset Explorer may be used to overlay model outputs with historical downtime events and a custom selection of time series data over a user-defined date range for a single WT (Figure 7). This consolidated record may assist in root cause analysis (RCA) or support one off evaluations.

Figure 7: Asset Explorer with Pulse (drift) events labeled as they're detected. Events are backfilled as of 3/1/2023. Visualizations are limited as the user cannot compare Pulse events for a specific WT to averages across the WF (the basis for the analysis for both drift and changepoint detection).

Although the application (both the Fleet and Asset Explorers) is available with Pulse outputs on an ongoing basis, Pattern plans to primarily consume model outputs via API, particularly so they can build additional front-end visualizations that meet the needs/requirements of their users (engineers, operators, and technicians). We will continue to use the application to inspect model outputs and suggest visualizations for Pattern users.

6.2 Email Notifications

Using the inference pipeline (Figure 5), Pulse outputs are available on an ongoing basis (with Foresight currently on hold). At Pattern's request, Pulse outputs are also aggregated via email in a Daily Digest and Weekly Summary (as of late June 2023). The Daily Digest summarizes the top five ongoing Pulse events as of the date specified (for Accumulator, Converter, and Main Bearing) and the Weekly Summary shows a list of all new and/or old events ongoing and/or completed. A screenshot of the Daily Digest is captured in Figure 8 with the Weekly Summary available upon request. Pattern engineers will continue to use beyond the grant agreement to identify/quantify value in the deployed outputs.


Tagup Pulse Events 2023-08-13 Inbox ×

support@tagup.io
to pattern-notifications

Sun, Aug 13, 5:33 AM (18 hours ago)

Below are the top ongoing Pulse events as of 2023-08-13. Events for components Accumulator, Converter, and Main Bearing were considered. Only the top 5 events are shown for each component. The events are ranked by Tagup's 'stability score' and then by longest duration.

Site	Asset	Component	Measurement	Event Type	Start Time	End Time	Duration	Score
BroadviewJN	[redacted]	Main Bearing	wtc_MainBTmp_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
BroadviewJN	[redacted]	Main Bearing	wtc_MainBTmp_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
Armow	[redacted]	Main Bearing	wtc_MainBTmp_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
BroadviewKW	[redacted]	Main Bearing	wtc_MainBTmp_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.8
Armow	[redacted]	Main Bearing	wtc_MainBTmp_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.7
Armow	[redacted]	Accumulator	wtc_HubPresA_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
SpringValley	[redacted]	Accumulator	wtc_HubPresC_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
SpringValley	[redacted]	Accumulator	wtc_HubPresB_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
Armow	[redacted]	Accumulator	wtc_HubPresC_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9
SpringValley	[redacted]	Accumulator	wtc_HubPresA_mean	Upward Drift	[redacted]	[redacted]	[redacted]	0.9

Please feel free to respond to this email with any comments or questions.

Main Bearing

[redacted] **Upward Drift for wtc_MainBTmp_mean:**

Turbine ID: [redacted]
Event Description: Upward Drift for wtc_MainBTmp_mean
Start Date: [redacted]
End Date: [redacted]
Score: 0.9

Figure 8: Daily Digest email with a list of the top 5 drift events by component of interest. Events are identified for the Accumulator, Converter and Main Bearing. The top five events are shown for each component, rank ordered by "Score" which is a metric that determines how confident the Pulse model is that the event in question is a true positive. The events with the highest scores correspond to the WTs with the highest need of attention. Key SCADA data (including relevant alarms) is presented in the email for every event detected (we truncated the output to show the first main bearing event identified above).

7. Conclusion & Recommendations

In summary, we successfully configured, deployed, and tested both supervised ML (Foresight) and unsupervised ML outputs (Pulse) on a subset of historical operating data for up to six self-performed WFs owned/operated by Pattern Operators LP.[32] A variety of event-based Pulse outputs were made available to Pattern users via user application, API and email notifications for further processing, evaluation, and operationalization/value recognition.[33]

As we look ahead, we will continue to work with Pattern engineers and data scientists to enhance the analytic capabilities (specifically predicting RUL of the main bearing after detecting drift and/or vibration), measure their performance and quantify the expected value of the analytics deployed (to justify a license renewal in December 2023). Value will be measured by the model correctly predicting an impending main bearing event so that repair and/or replacement can reduce the cost of lost production associated with derating a WT. Analytics will be benchmarked against other failure prediction tools, or status quo procedures (we are being considered head-to-head with other third-party providers).

As we look beyond Phase II, we plan to continue supporting Pattern Energy by demonstrating the expected value on a going forward basis with the intention of expanding to additional WFs (up to 25 WFs in total; 2,500 WTs from Siemens, GE, Vestas, and Mitsubishi; ~6 GW of generating capacity).

[32] Our integrated microservices autonomously inject (via Ivaldi) the updating data into the analytics engine (Ran and Vidar), which runs the inference pipeline to compute model outputs (and write to Saga) daily (around 9am). The updated model outputs may be viewed across the fleet and on a per-asset basis via two views (via Midgard): the Fleet Explorer and Asset Explorer. The Fleet Explorer displays model outputs in aggregate across the selected WFs whereas users can identify the most at-risk assets across the connected install base. If the operator chooses to zoom in on specific events for a given WT, the Asset Explorer displays those views (Spring Valley only).

[33] Event-based Pulse outputs are centered around a stability score that evaluates for multiple algorithm configurations, what is the stability score of the output, ranging from 0 to 1. This allows users to rank order the confidence of the algorithms in identifying anomalous operating conditions, albeit drift of one WT measurement (relative to the other WTs at the same WF), or a change in operating conditions for one WT relative to how it was "historically" operating.