



Machine learning-based prediction of critical load events in airborne wind energy systems

Monika Pasquali^{1,2}, Oriol Cayon¹, Spilios Theodoulis¹, Christophe Farges², Tudor-Bogdan Airimitoai², Patrick Lanusse², and Roland Schmehl¹

¹Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands

²Univ. Bordeaux, CNRS, Bordeaux INP, IMS, 33405 Talence, France

Correspondence: Monika Pasquali (monika.pasquali@bordeaux-inp.fr) and Roland Schmehl (r.schmehl@tudelft.nl)

Abstract. Predicting and avoiding critical load events is crucial for the reliable autonomous operation of airborne wind energy (AWE) systems. Such events can lead to actuator saturation, loss of steering authority, and increased structural stresses, eventually resulting in long-term degradation of the kite, tether, and ground-station components. The present study explores a data-driven approach to predicting tether force peaks, defined as instances in which the force exceeds a threshold set by the winch controller. The method is developed using a large-scale SCADA dataset from Kitepower’s leading-edge inflatable V9 kite, including more than 1000 pumping cycles spanning 26 h of kite operation and covering a wide range of wind and operating conditions. Tether force peaks are detected and used to extract pre-peak time windows that capture the system behavior before each threshold exceedance. From these windows, statistical, temporal, and spectral features are computed from multiple kite, tether, and ground-station measurements. A supervised classification model is trained to distinguish normal operation from pre-peak dynamics. The proposed approach predicts critical load events 1 s in advance and shows strong generalization capability on flight data that was not used for training the model, correctly identifying 85.4 % of the critical load events (recall) and achieving a 90.9 % F1-score, a metric that indicates a strong balance between correctly detected events and false alarms. Moreover, the extracted features provide valuable insight into the system dynamics preceding these events. The obtained results enable preventive control actions, thereby improving the reliability of AWE systems.

1 Introduction

Airborne wind energy (AWE) is a technology that converts the kinetic energy of wind into electricity using tethered aerial devices, such as power kites. The attribute “airborne” indicates that these systems do not rely on static structures, such as wind turbine towers, to support the energy-harvesting system component. Instead, the flying kite and its ground station are controlled automatically, converting aerodynamic forces into electricity.

This conversion concept bears several advantages. First, fewer energy-intensive materials, such as concrete and steel, are used, thereby reducing the environmental footprint and simplifying construction, transportation, and installation (Fagiano et al., 2022). Second, the absence of a static structure allows reaching higher altitudes, typically between 200 and 500 m above



ground. At these heights, wind speeds are generally higher and more persistent. Since the available power scales with the cube of the wind speed, even small increases can lead to significant gains in energy production (Bechtle et al., 2019).

25 However, these benefits come at the cost of increased system complexity. Safe and efficient operation requires a high level of automation to control the kite trajectory and the power conversion process. For AWE to become a viable renewable energy technology, systems must operate reliably and autonomously for long periods, while remaining cost-effective.

The present work focuses on ground-based power conversion, specifically pumping power generation. In this configuration, the kite generates aerodynamic forces that pull on a tether deployed from a ground-based winch. The winch torque drives an electric generator, converting mechanical power into electrical power. Most existing research on ground-based AWE systems focuses on maximizing power output through trajectory optimization (Horn et al., 2013; Licitra et al., 2019; Trevisi et al., 2022). In contrast, only a few studies address their reliability and safety. For instance, Salma et al. (2020); Salma and Schmehl (2020) present a failure mode and effects analysis (FMEA) combined with a fault tree analysis (FTA) for a flexible-wing kite power system. These studies confirm the tether force as a key variable affecting system safety. While a higher force generally increases power production, force peaks may exceed structural limits and damage the kite, the tether, or ground-station components. Consequently, accurate monitoring and control of the tether force are essential for both performance and safety.

Particularly challenging is the strong coupling between the ambient wind speed, the flight speed and the aerodynamic loading. Because of their low mass-to-surface ratio and inertia, soft-kites react comparatively quickly to changes in the wind speed, such that a wind gust or turbulent fluctuation can result in sudden variations of the tether force. Several control-oriented studies have addressed tether force regulation and power fluctuation mitigation, often through model-based control design and simulation analysis. For example, Hummel et al. (2024) analyze tether force control for the MegAWES system (Eijkelhof and Schmehl, 2022) and propose model-based strategies to reduce power fluctuations by acting on winch dynamics and controller tuning. Instead, Vertovec et al. (2024) introduce an avoidance maneuver for the path-following controller of a tethered aircraft through Hamilton-Jacobi reachability analysis, which is initiated when a tether rupture is potentially imminent.

Data-driven methods have recently been introduced to complement model-based approaches in AWE. To the best of the authors' knowledge, such techniques have so far been applied mainly to tether force or power estimation (Rushdi et al., 2020; Gupta et al., 2025). Only Rapp and Schmehl (2021) explicitly consider the prediction of critical load events. In particular, the proposed approach learns from simulations of upset conditions to anticipate when the tether force is likely to exceed safe limits, enabling the controller to perform a mitigation maneuver. While promising, this work remains entirely simulation-based and may not capture unmodeled dynamics, sensor noise, or environmental uncertainties. Moreover, it relies on wind data, which are often difficult to measure reliably at the kite's altitude and may not be consistently available in commercial AWE implementations, thus limiting its practical applicability. In addition, tether force peaks are detected with a pre-peak horizon of only 0.2 s, which may be insufficient for executing effective avoidance manoeuvres given the kite's inertia and the response time of the control actuators.

From a control perspective, the effectiveness of any avoidance strategy is limited by the different time scales of wind gusts, actuator dynamics, and the kite's response. Wind gusts can modify the resultant aerodynamic force within the order of the



aerodynamic timescale, which, for the considered kite size and flight speeds, is of the order of 0.1 s (Leuthold, 2015), causing rapid variations in tether tension. In contrast, the onboard depower actuation system, which adjusts both steering lines symmetrically to reduce the aerodynamic loading, in the kite control unit (KCU) operates on time scales of the order of seconds due to its high gear ratio. While the ground-based winch has a very fast inner torque-control loop (≈ 0.02 s), the overall tether-force regulation remains slower, with a response time of ≈ 0.3 s. Because KCU actuation is comparatively slow, fast disturbance mitigation must primarily rely on the ground station winch.

However, the kite dynamics introduce an additional limitation. While lightweight kites react quickly to control inputs, heavier or larger kites exhibit slower dynamic responses due to inertia, which can further delay the effect of mitigation actions. Moreover, during turns in crosswind flight or reel-in phases, steering and depower actuators may approach or reach saturation, reducing the available control authority. Consequently, mitigation maneuvers must be initiated sufficiently in advance to account for actuator limitations, potential saturation effects, and the kite's dynamic response before tether loads reach critical levels.

To address these limitations, the present work proposes a data-driven approach to early predict high-tether-force events in ground-based pumping AWE systems. In contrast to existing simulation-based studies, the proposed approach is trained and validated on a large-scale SCADA dataset collected from the Kitepower V9 kite system, including 26 h of flight conducted between 2024 and 2025 and covering more than 1000 pumping cycles across diverse operating conditions. Building on these real-flight measurements, the predictor is designed to anticipate tether-force peaks with a pre-peak horizon of 1 s, providing sufficient time for the control system to perform mitigation maneuvers. Moreover, the model outputs an explicit measure of prediction confidence, enabling risk-aware corrective actions. Finally, the generalization capability of the proposed method is demonstrated through evaluation on an out-of-domain flight, spanning 6.2 h and 145 pumping cycles, highlighting its ability to operate reliably under previously unseen conditions and offering additional insight into the dynamic patterns that precede critical load events.

The remainder of this paper is organized as follows. Section 2 introduces the Kitepower V9 kite system with its main components, the sensor setup, and available measurements, outlining the operating conditions covered by the flight campaigns. Section 3 presents the proposed learning-based tether-force prediction framework, including data segmentation, feature extraction, and classification methodology. The prediction performance is assessed on both in-domain and out-of-domain flight data in Sect. 4. Finally, conclusions and future research are in Sect. 5.

2 Experimental setup and dataset

This section introduces Kitepower's soft-kite-based AWE system, referred to as the Falcon system, equipped with the V9 kite, as well as the SCADA dataset used in this study. We first describe the pumping-cycle operation and main components, sensor setup, dataset characterization, and operating envelope covered by the measurements.



2.1 System overview

- 90 Kitepower's AWE system harvests wind energy in pumping cycles, generating electricity on the ground using a generator driven by the tether. The main components are the ground station, the tether, and the kite, which includes the flexible wing, the bridle line system, and the suspended kite control unit (KCU), with attached ram-air turbine. The KCU connects the tether to the bridle lines and provides control actuation. Figure 1 shows the system in operation at Kitepower's test site in Bangor Erris, Ireland. The main specifications are reported in Table 1. We note that the tether of this prototype is considerably over-



Figure 1. V9 kite flying at the AWE test center in Bangor Erris, Ireland. Photo courtesy of Kitepower B.V.

- 95 dimensioned for the power rating of the system, with the main reason to cover all possible overload situations with a sufficiently wide safety margin. On the other hand, this means that we are able to analyze critical overload situations that would lead to failure of a properly dimensioned commercial AWE system.

Table 1. Main physical parameters of the soft-kite-based AWE system used in the experiments. The shape of the KCU can be approximated by a cylinder with its axis aligned with the kite.

Wing		KCU		Tether	
Flat wing area	60 m ²	Diameter	0.62 m	Density	724 kg m ⁻³
Projected wing area	46.85 m ²	Length	1.2 m	Diameter	14 mm
Mass (incl. bridle line system)	62 kg	Mass (incl. ram-air turbine)	31.6 kg		



The kite operates in consecutive pumping cycles with two alternating phases: traction and retraction, also denoted as reel-out and reel-in. In the traction phases, the kite flies crosswind trajectories, producing a high tether force and reeling the tether from a drum connected to the generator. When the maximum tether length is reached, the crosswind maneuvers are stopped, and the system switches to retraction. Here, the generator acts as a motor, reeling the tether back onto the drum, while the kite flies in an aerodynamically efficient configuration as it prepares for the next cycle. The front bridle lines carry most of the aerodynamic load, while the rear lines are used for steering and depowering. Steering is achieved by shortening one steering line relative to the other, creating an asymmetric deformation of the wing. Depower adjusts both steering lines symmetrically, changing the pitch of the wing relative to the KCU. This pitch strongly affects the kite's aerodynamic properties and is primarily used to further reduce the required force and power to pull the kite back towards the ground station. The aerodynamic loading generated by the kite results from a coupled interaction between the KCU and the ground-station winch controller. Because the winch controller regulates the reel-out dynamics, it has a central role in limiting tether force excursions during power production. Since the present work focuses on predicting critical load events, the winch control strategy is introduced in the next section.

2.2 Ground-station winch control

For optimal performance of a pumping AWE system, the energy generated during reel-out has to be maximized, and the energy consumed during reel-in has to be minimized. From a control perspective, this is achieved through a cascaded control architecture that separates ground station winch control from the kite flight controller (Rapp et al., 2019; Hummel et al., 2024). The ground station regulates the reeling speed v_r of the tether by commanding the generator torque τ , thus controlling also the tether force F_t with the additional input of the ambient wind speed v_w , as illustrated in Fig. 2. The force tracking error $F_{t,e}$ denotes the difference between the reference $F_{t,ref}$ and measured tether force F_t .

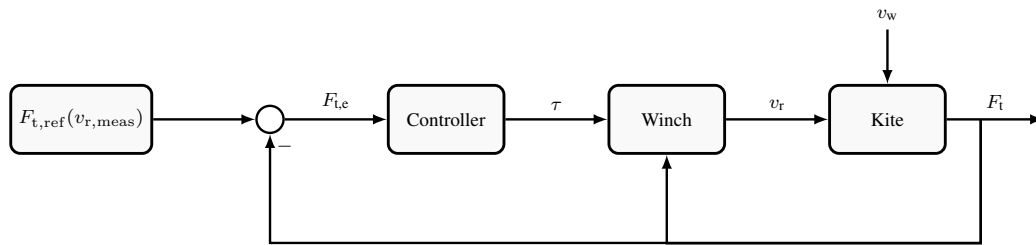


Figure 2. Winch control architecture of the Kitepower Falcon system.

The mechanical power transferred to the ground station is defined as $P = F_t v_r$. To optimize this product, the winch control law prescribes how the reference tether force $F_{t,ref}$ has to vary as a function of the measured reel-out speed $v_{r,meas}$, while staying within safe operating limits. In fact, the tether force F_t should not exceed a certain threshold F_t^* to avoid an impaired steering capability of the KCU. However, despite the implemented feedback control strategy, the experimental campaigns reveal frequent short-duration force spikes exceeding the threshold. Two principal limitations of the existing control architecture may contribute to these overshoots:



1. **Winch dynamic limitations:** The winch must be able to react sufficiently fast to sudden increases of the wind speed. If the compensating increases of the reel-out speed are too slow, large tether force overshoots can occur. On the other hand, if the winch reacts too fast or is insufficiently damped, the dynamic system becomes overly sensitive to disturbances. Therefore, an appropriate trade-off between responsiveness and inertia is required. Hummel et al. (2024) propose a frequency domain analysis of winch dynamics to properly size the winch of the AWE system.

2. **Lack of anticipatory actions:** The current winch controller architecture is predominantly reactive, since the generator torque is adjusted only after an increase in tether force has already been measured. Consequently, high-frequency aerodynamic disturbances may produce force transients that cannot be compensated for sufficiently fast by the feedback loop alone. To address this, Hummel et al. (2024) investigated a torque feedforward strategy. However, if reel-out acceleration is constrained by power limits, a torque feedforward alone cannot guarantee avoidance of force spikes. Instead, the controller could command kite depower or trajectory change to reduce the aerodynamic loading.

These limitations of the existing control architecture motivate the strategy proposed in the present work: a data-driven approach that augments the existing winch feedback architecture. As explained in Sect. 3, a machine learning model is trained on numerous experimental datasets to predict with a horizon T_r when the tether force is about to exceed a predefined threshold F_t^* . This technique will provide the means to take countermeasures, such as triggering suitable avoidance maneuvers to avoid the imminent overload scenarios. The development of these active countermeasures is left for future work.

2.3 Sensor setup and measured signals

The relevant sensors in the investigated AWE system and the measured signals used to construct the SCADA dataset for training and evaluation of the proposed machine learning algorithm are described here. Although the system is equipped with additional instrumentation on the kite, including Pitot tubes with flow vanes to measure the relative flow velocity, and a vertical wind profiler lidar positioned on the ground, upwind of the AWE system, these measurements were not consistently available across all experimental campaigns. Since this additional instrumentation is not part of the standard commercial system configuration, and to ensure a homogeneous and reliable set of inputs across the full SCADA dataset, the corresponding measurements were not considered in this study. The relevant sensors are:

1. **Load cell:** The tether force is measured using a load cell installed at the tether exit point of the drum.
2. **Tether length encoder:** The deployed length of the tether is measured using the kite's reeling system. An incremental angular encoder on the drum records its rotational speed, from which the tether length is computed (Van der Vlugt et al., 2013).
3. **Tether angle sensors:** The elevation and azimuth angles of the tether are measured at the ground station using magnetic angular encoders. Together with the tether length, these measurements allow estimation of the kite position (Peschel, 2013).



4. **Pixhawk IMU/GPS unit:** Mounted on the kite’s central structure, providing position, velocity and orientation measurements.

The measured signals available for each flight are summarized in Table 2. The measurements are sampled with a constant timestep $\Delta t = 0.1$ s, corresponding to a sampling frequency $f_s = 10$ Hz. More details about the sensor setup, the various measurement techniques, the use and quality improvement of the acquired data can be found in Oehler and Schmehl (2019); Schelbergen and Schmehl (2020, 2024); Cayon et al. (2025).

Table 2. Measured variables available for the Kitepower Falcon system.

Category	Variable	Description	Unit
Time	Date, time, time of day	Calendar date and UTC timestamp	–
Tether / ground	Tether force	Measured tether force	N
	Tether speed	Reel-out speed of the tether	m s^{-1}
	Tether length	Deployed tether length	m
Kite IMU/GPS	$V_{\{x,y,z\}\text{kite}}$	Velocity components in the NED earth frame	m s^{-1}
	V_{kite}	Kite’s speed in the NED earth frame	m s^{-1}
	Kite latitude, longitude, altitude	Geodetic position coordinates	rad, m
	Kite roll, pitch, yaw	Orientation angles of the kite	rad
Relative position	Position east, north, height	Kite position relative to the ground station	m
	Kite azimuth, elevation	Spherical coordinates of the kite	rad
	Kite distance	Distance between kite and ground station	m
	Kite heading	Heading angle of the kite	rad
	Kite course	Course angle of the kite	rad
Commands	Kite steering, depower	Steering/depower commands	%
Energy / dynamics	Ground mechanical power	Mechanical power measured at the winch	W
	Kite turn rate, course rate	Rate of change of heading and course	rad s^{-1}
Flight phase	Flight phase, flight phase index	Active flight phase label and corresponding index	–

2.4 Dataset characterization and operating envelope

Data from 55 individual test flights conducted between 2024 and 2025 at Kitepower’s test site in Bangor Erris, Ireland, were available for the present study. Although the region is generally governed by consistently strong winds, several flights contained almost no critical load situations, and including those would have mainly added redundant non-critical data, increasing the computational cost for training without providing significant additional information. The selected subset, therefore, focuses on flights with a higher frequency of critical load situations while still covering a representative range of wind conditions and

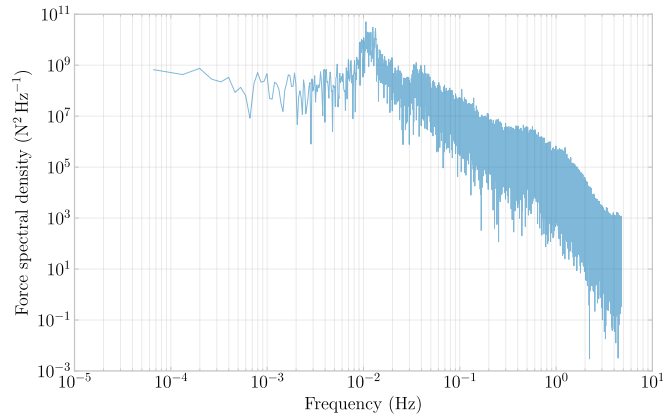


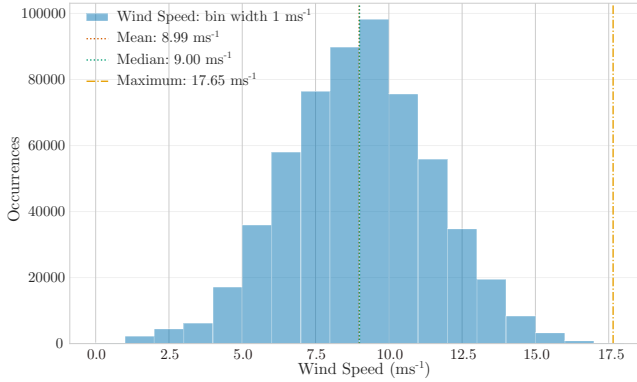
Figure 3. Tether force spectral density for the test flight of 5 June 2024. The observable frequency range is limited to $f < 5$ Hz by the Nyquist frequency corresponding to the sampling rate.

operating regimes. This ensures sufficient variability to learn meaningful patterns, preceding tether force peaks, while keeping the training process computationally efficient.

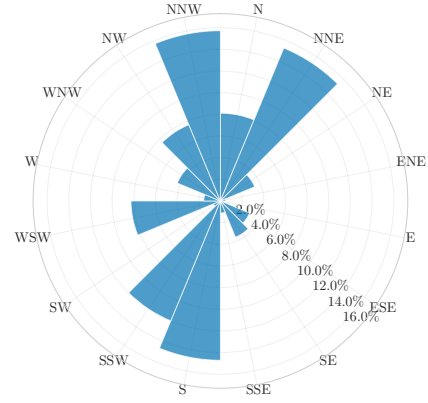
The training dataset comprises 12 flights with more than 1000 pumping cycles, corresponding to 26 h of autonomous kite operation. The data captures the dynamic behavior of the kite, tether, and ground station during both normal operation and critical load events, providing a realistic basis for data-driven prediction. Next to this, an additional test flight spanning 145 pumping cycles was determined for out-of-domain evaluation of the machine learning model. The data of this specific flight is available in open access (see data availability section). Figure 3 shows the spectral density of the measured tether force to illustrate the dynamic behavior of the system over the observable frequency range. The diagram shows the base frequency of the pumping cycle operation (time period of 90 to 120 s) at around 10^{-2} Hz and the base frequency of the individual figure-of-eight patterns (time period 25 to 30 seconds) at around 3.5×10^{-2} Hz.

The following analysis summarizes the statistical distributions of key measured quantities for the 12 training flights, including ambient wind conditions, tether force, and power production. Figure 4 shows the distribution of wind speed and upwind direction (relative to north, clockwise positive) measured at ground level for the flights where wind data were available. The ground wind speed measurements range up to 17.65 m s^{-1} , with a mean value of 8.99 m s^{-1} .

The distribution of the tether force provides insight into the operational loading regime and the severity of extreme events. Figure 5(a) shows that most measurements remain below the winch controller threshold, $F_t^* = 24.5 \text{ kN}$. The values above this threshold can be used for fault labeling. The statistics of the corresponding instantaneous mechanical power is illustrated in Fig. 5(b), distinguishing between reel in (negative power) and reel out (positive power). Since the sampling interval is equidistant, the average net cycle power can be calculated from Fig. 5(b) as the difference between the positive (blue) and negative power occurrences (red).

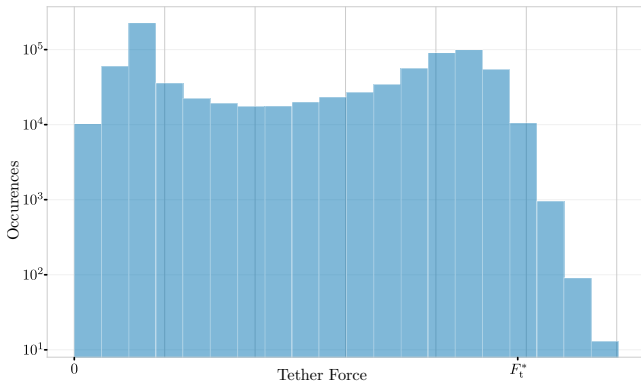


(a) Ground wind speed

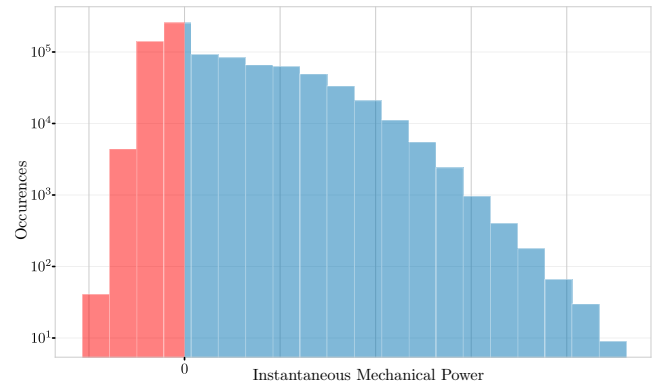


(b) Ground upwind direction

Figure 4. Statistical distributions of ground wind speed and upwind direction for the 12 training flights.



(a) Tether force



(b) Instantaneous mechanical power

Figure 5. Statistical distributions of tether force and power production.

Figure 6 shows the relation between the measured tether force and reel-out speed, covering both the traction and the retraction phases. For positive reel-out speeds, the controller increases the commanded speed based on the measured tether force, producing the steep, approximately linear branch in the center of the diagram. Once the force reaches the threshold value F_t^* , the winch continues increasing the reeling speed to maintain the force at that value. This is the plateau in the upper right of the diagram, where the force stays close to F_t^* . The scatter is caused by the finite response time of the winch to the force fluctuations along the figure-eight trajectory, which also explains the samples slightly above F_t^* . Note that F_t^* is a control threshold and not a structural limit: the maximum force that the system can withstand is always higher, leaving margin for these fluctuations.

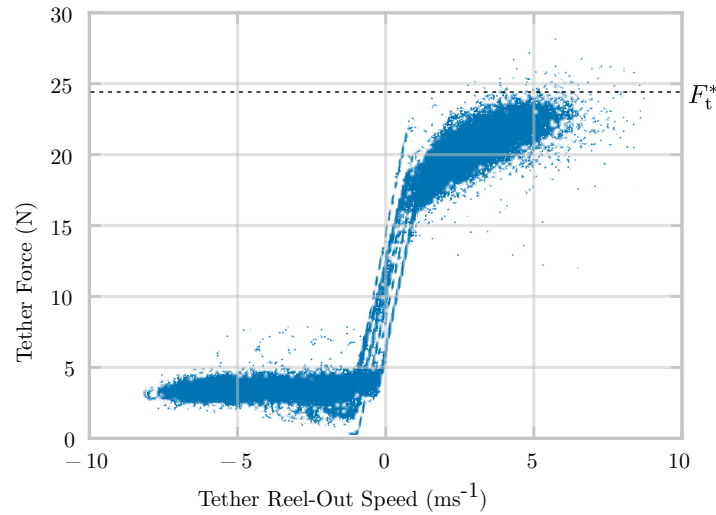


Figure 6. Tether force versus reel-out speed for the test flight of 5 June 2024.

For negative reel-out speeds, the controller follows a second set-point curve, offset from the one used during traction. When the force reaches its minimum value, the force setpoint is held constant to ensure a minimum tension in the tether. The samples therefore form a nearly horizontal band of low force spanning a wide range of reel-in speeds, with no $F_t(v_r)$ relation visible. The scatter around this band is of similar magnitude as during reel-out, so the minimum force also has to be set with enough margin to avoid too much slack in the tether.

3 Prediction of critical load events

This section presents the proposed data-driven approach for predicting tether force peaks. The objective was to detect, in advance, conditions that may have led to critical load events, allowing sufficient reaction time for avoidance control strategies. Tether force peaks were detected and used to extract pre-peak time windows that capture the system behavior leading up to each event. From these windows, statistical, temporal, and spectral features were computed from multiple kite, tether, and ground-station measurements listed in Table 2, excluding kite latitude, longitude, orientation angles, and flight phase indices, which were found to be uninformative during model training and therefore discarded for computational efficiency. A total of 21 measured signals were retained. A supervised classification model was then trained to distinguish normal operation from pre-peak dynamics. For evaluation, two of the most common machine learning approaches were tested: support vector machines (Noble, 2006; Suthaharan, 2016) and logistic regression (Hosmer Jr et al., 2013; Sperandei, 2014). The proposed approach predicts critical load events up to 1 s in advance and demonstrates generalization capability on out-of-domain flight data. The complete pipeline is outlined in Fig. 7 and each block is detailed in the following subsections.

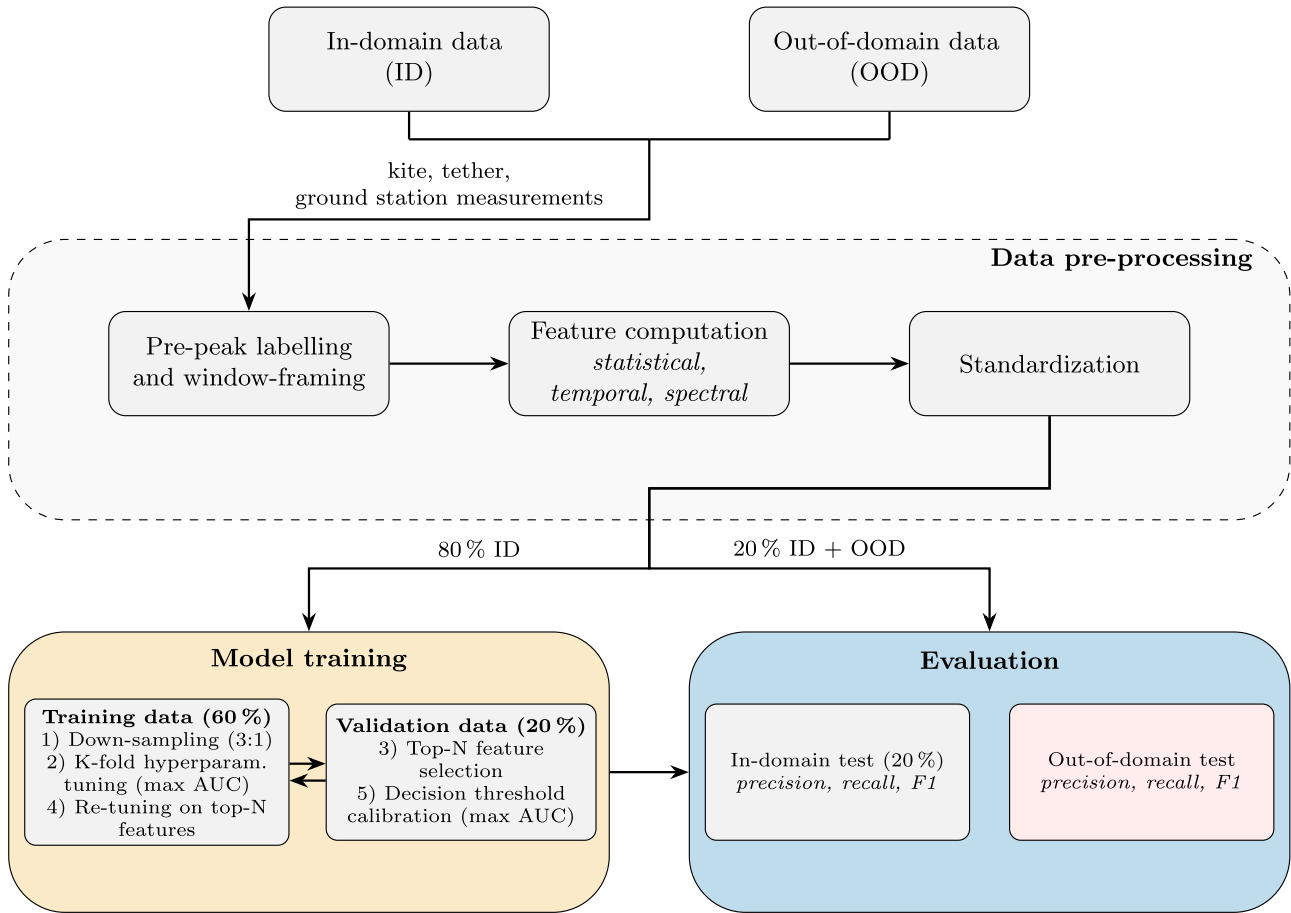


Figure 7. Machine Learning pipeline.

3.1 Data pre-processing

This section describes the data pre-processing pipeline, from fault event labeling and frame construction to feature computation, dataset splitting, and standardization.

3.1.1 Pre-peak labeling and window-framing

215 To capture the system dynamics that lead to a tether force overshoot, fault events during reel-out phases are identified by detecting the first instance in each phase when the tether force exceeds the threshold F_t^* . Only the first peaks are considered because these represent the actual fault conditions, whereas later peaks may be due to the dynamic system response, thereby providing less independent information for early fault prediction.



Table 3. Segmentation parameters used in the pipeline.

Parameter	Value	Description
t_{frame}	5 s	Frame length used for feature extraction
t_{pre}	(0.2, 0.5 and 1) s	Pre-peak horizon
F_t^*	24 525 N	Tether force threshold

For each detected peak, the segmentation procedure moves backward in time by a fixed pre-peak horizon t_{pre} , which represents the available reaction time for the controller to prevent the overshoot. A temporal frame of duration t_{frame} is then extracted immediately before this shifted time and labeled as faulty (label 1). This guarantees that the frame contains only the system behavior leading up to, but not including, the force peak itself. Different frame durations were tested, and the one yielding the best fault prediction performance was retained. The remaining tether force signal is segmented into frames of equal duration t_{frame} , starting from the beginning of the signal and without any overlap. These frames are labeled as non-faulty (label 0). All additional measured signals listed in Sect. 3 are segmented correspondingly, using the same start and end indices as the tether force frames. Each frame of these additional signals inherits the label assigned to the corresponding tether force frame. A binary label vector $\mathbf{Y}_{\text{tot}} \in \{0, 1\}^N$ is obtained, where $N = 14489$ is the total number of frames.

Illustrated in Fig. 8, this procedure guarantees that all frames have the same duration, only pre-fault dynamics are used for faulty samples, and prevents any mixing between faulty and non-faulty data. All numerical values used in this segmentation step are listed in Table 3.

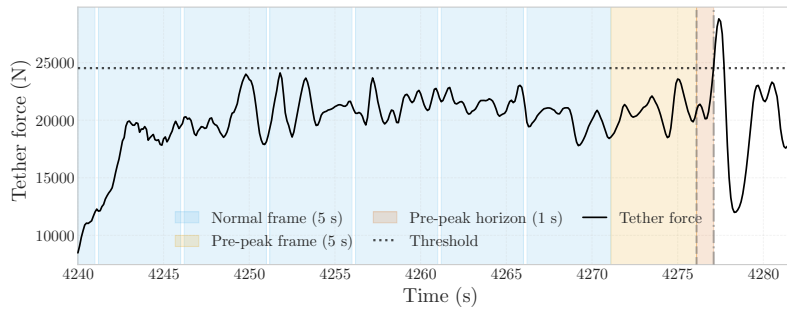


Figure 8. Tether force signal segmented into 5 s frames. Exceedance of the force threshold F_t^* (dashed line) defines a pre-peak horizon t_{pre} and a preceding 5 s pre-peak window. Blue frames correspond to normal operation. Identical framing is applied to the retained measured signals to capture the full system dynamics.



3.1.2 Dataset composition and class distribution

Figure 9(a) presents the dataset composition and class distribution per test flight, while Figure 9(b) shows the class distribution across all flights. The dataset contains 13908 non-faulty and 581 faulty frames, resulting in a class imbalance ratio of approximately 24:1.

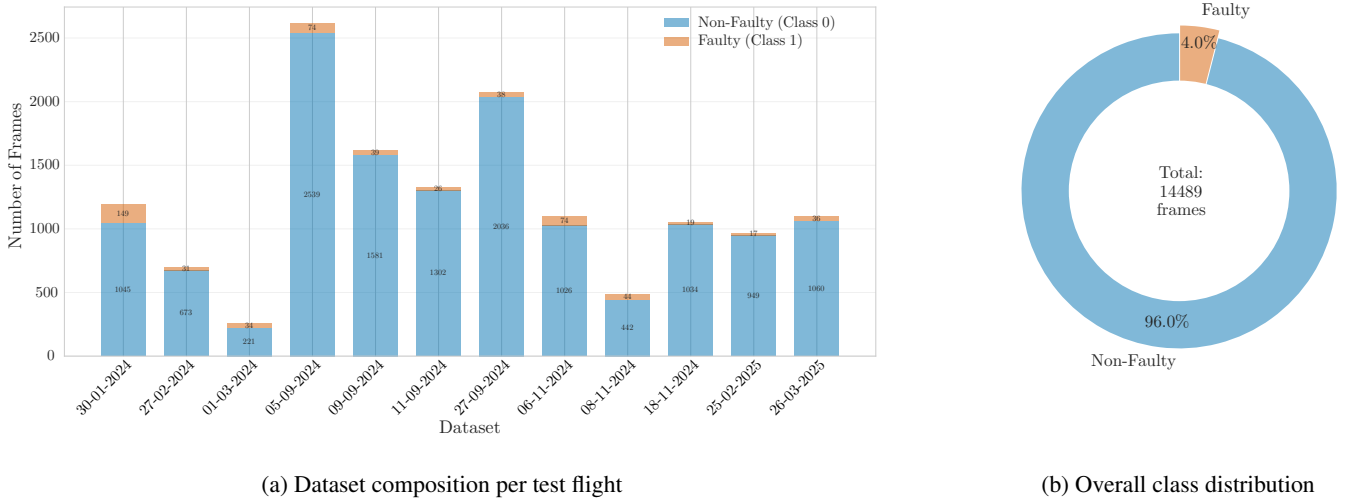


Figure 9. Dataset composition showing the number of faulty and non-faulty frames for each of the 12 test flights and overall class distribution.

235 3.1.3 Feature computation

Once the signals have been segmented and labeled, the system behavior within each frame is characterized through a set of descriptive features. For each frame, statistical, temporal, and spectral features are extracted from the tether force signal and the retained measured variables.

The statistical features include the mean, standard deviation, median, minimum, maximum, peak-to-peak amplitude, inter-quartile range, and the 75th and 90th percentiles. Energy-related characteristics are evaluated using the root-mean-square value and the signal energy. Shape properties are described using skewness, kurtosis, and crest factor.

Temporal dynamics are represented by the zero-crossing rate, time to maximum value, mean absolute deviation (MAD), cumulative sum range (CSR), maximum slope, and time-reversal asymmetry.

Frequency-domain information is obtained from the fast Fourier transform and summarized through the spectral centroid, low-frequency energy ratio, minimum and maximum spectral amplitudes, and the maximum spectral amplitude above 1 Hz. Given the sampling period of 0.1 s ($f_s = 10$ Hz), this feature is computed over the Nyquist-limited frequency range $(1, f_s/2] = (1, 5]$ Hz.

These features are selected because they are computationally efficient and have proven effective in previous time-series analyses and fault detection studies (Rapp and Schmehl, 2021; Fulcher and Jones, 2014). In total, 25 features are extracted per



250 signal. $\mathbf{X}_{\text{tot}} \in \mathbb{R}^{N \times (d \times s)}$ is the resulting feature matrix, where $N = 14489$ is the number of frames, $d = 25$ is the total number of extracted features and $s = 21$ the number of signals. Each frame is associated with a binary label from the vector $\mathbf{Y}_{\text{tot}}$, indicating faulty or non-faulty operation. Together, $(\mathbf{X}_{\text{tot}}, \mathbf{Y}_{\text{tot}})$ define the supervised learning dataset used in the subsequent classification stages.

3.1.4 Splitting into training, validation, and test sets

255 The dataset is first partitioned into train and test sets, preserving the class distribution through stratification. Crucially, the frames comprising the test set are selected to be temporally consecutive with those in the training set, avoiding time leakage. This test set will be used for in-domain evaluation. The training set is further divided into training and validation subsets (also with stratification) for hyperparameter tuning and feature selection, as detailed below. Down-sampling by 3:1 is applied to the train set, but not to the validation and test sets, to preserve class distribution.

260 3.1.5 Standardization

Before training the classification models, the feature matrix $\mathbf{X}_{\text{tot}}$ is processed to ensure numerical consistency. Since the computed features have different physical meanings and numerical ranges, each feature is standardized using z-score normalization (Sujon et al., 2024). Prior to standardization, extreme values are clipped to the 1st and 99th percentiles computed from the training set to reduce the influence of outliers. The standardization is then performed as $x^{\text{norm}} = (x - \mu) / \sigma$, where μ and σ 265 denote the mean and standard deviation of each feature computed on the training set only. The same clipping bounds and standardization parameters are then applied to the test set.

3.2 Model training

This section describes the training methodology, covering the support vector machine (SVM) classifier, feature selection, decision threshold calibration, and a logistic regression baseline.

270 3.2.1 Classification model: support vector machine

The fault detection problem is formulated as a binary classification task, where each frame is classified as faulty or non-faulty, as described in Sect. 3.1.1. SVM is adopted as the primary classifier because it works well on high-dimensional problems and is robust to class imbalance. Given a feature vector $\mathbf{x} \in \mathbb{R}^d$, the SVM decision function is expressed as

$$f(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad (1)$$

275 where $\mathbf{x}_i$ are the support vectors, $y_i \in \{-1, +1\}$ are their labels, α_i are learned coefficients, b is a bias term, $K(\cdot, \cdot)$ is a kernel function and N_s is the number of support vectors. A sample is classified according to the sign of $f(\mathbf{x})$, while the magnitude of $f(\mathbf{x})$ provides the decision margin, which quantifies the confidence of the prediction. More specifically, this decision margin represents the signed distance of a sample from the decision boundary. Large absolute margins correspond to



confident classifications, while margins close to zero indicate uncertain samples located near the boundary between classes.

280 In this work, the SVM's hyperparameters tuning (Schölkopf and Smola, 2001), such as kernel function choice and coefficient learning, using a three-folds cross validation strategy on the training set. This allows us to retain the left-out validation set for feature selection.

3.2.2 SVM-based feature selection strategy

Although the extracted feature representation discussed in Sect. 3.1.3 is rich and informative, it is also highly dimensional and
285 may contain redundant or weakly relevant variables. For this reason, a feature selection strategy (Guyon and Elisseeff, 2003) is introduced to systematically evaluate whether a reduced subset of features can achieve comparable performance to the full feature set. This analysis allows a better understanding of the relevance of each variable within the classification algorithm. The feature selection procedure is included in the training pipeline and is performed using only the training and validation sets. The test set is never involved in this process to avoid any information leakage. The method has three main steps:

- 290
1. Train an initial SVM classifier with all the feature set on the train split (3 folds cross-validation).
 2. Evaluate the importance of each feature using permutation-based criteria on the validation split (Altmann et al., 2010; Fisher et al., 2019).
 3. Select a reduced subset of features and re-train the SVM on the train split using only the selected features.

The idea is to quantify how much each feature influences the SVM decision behavior. Large absolute margins indicate confident
295 predictions, while small margins correspond to uncertain samples. Feature importance is evaluated by randomly permuting one feature at a time in the validation set while keeping all other features unchanged, and by measuring how the SVM decision margin is affected. If permuting a feature significantly alters the margin behavior, that feature is considered informative (Brank et al., 2002; Guyon et al., 2002). Two complementary importance perspectives are adopted:

- 300
- Global margin-based importance: for each feature, its values are randomly permuted across all the validation samples multiple times. The average reduction in the absolute margin values is then computed.
 - Boundary margin-based importance: while global margin-based importance evaluates separability considering all samples, the boundary margin-based importance considers only samples close to the decision boundary. These samples are identified as those whose absolute margin is below an adaptive threshold, selected to ensure a sufficient number of faulty samples are included.

305 Two independent feature rankings are obtained from the global margin-based and boundary margin-based analyses. To avoid excessively large feature subsets, only the top 20 features from each ranking are retained. The final feature subset is then constructed by taking the union of these top-ranked features and it is reported in Sect. 4. This ensures that the selected features include both those that are important for the overall classification and those that are especially useful for samples near the decision boundary.



310 3.2.3 Decision threshold calibration

The SVM produces a continuous decision function value corresponding to the signed distance of a sample from the separating hyperplane, that is the classification margin. The sign of this value determines the predicted class, while its magnitude reflects the confidence of the prediction. For interpretability, this margin can be mapped to the interval $[0, 1]$ to obtain a probabilistic score via Platt scaling (Platt, 1999), which fits a logistic sigmoid to the SVM output using internal cross-validation. Although a
315 default decision threshold of 0.5 (corresponding to a zero-margin decision boundary) is commonly adopted, it is not necessarily optimal for the specific classification task. The decision threshold is therefore calibrated using the validation set. Different threshold values are evaluated, and the one that maximizes the area under the receiver operating characteristic curve (AUC) (Bradley, 1997) is selected, as it provides an appropriate trade-off between precision and recall. Once the optimal threshold is determined, it is fixed and subsequently applied to the test set without further adjustment.

320 3.2.4 Comparison model: logistic regression

In addition to the SVM classifier, a logistic regression (LR) model is also trained as a baseline method. Logistic regression is a linear classifier that provides probabilistic outputs and is widely used due to its simplicity and interpretability. Given a feature vector $\mathbf{x}$, logistic regression estimates the probability of a frame being faulty as

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x} - b)}, \quad (2)$$

325 where $\mathbf{w}$ and b are learned parameters (Hosmer Jr et al., 2013). The LR model is trained using the same training, validation and test splits adopted for the SVM. The same selected feature subset is used in order to ensure a fair comparison between classifiers.

4 Evaluation

The final models are evaluated on the independent test set using the following metrics: precision, recall and F1-score (Goutte
330 and Gaussier, 2005).

4.1 In-domain test

Table 4 summarizes the in-domain classification performance for all models and pre-peak horizons. All models achieve very high performance on the non-faulty class, with precision and recall above 95.0 % across all configurations. This indicates that normal operating conditions are well characterized in the feature space and can be reliably identified. The main challenge lies
335 in predicting the faulty class, which is strongly underrepresented in the dataset. In general, LR performs better than SVM in terms of faulty recall and macro F1-score. The prediction horizon has a clear influence on performance. Increasing the horizon improves performance, particularly for LR. The best performance is obtained for $t_{\text{pre}} = 1$ s using LR, which achieves a faulty recall of 94.0 % and a macro F1-score of 90.0 %. This indicates that relevant dynamics preceding critical events are already



Table 4. In-domain classification performance for different models and pre-peak times.

In-domain dataset										
Model	Pre-peak (s)	Non-faulty (2782 samples, 96 %)			Faulty (116 samples, 4 %)			Macro avg.		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
SVM	0.2	99.1	96.6	97.8	48.9	79.3	60.5	74.0	87.9	79.2
LR	0.2	99.1	96.6	97.8	48.9	79.3	60.5	74.0	87.9	79.2
SVM-FS	0.2	99.4	98.5	98.9	69.7	85.3	76.7	84.6	91.9	87.8
LR-FS	0.2	99.4	98.7	99.1	73.9	85.3	79.2	86.6	92.0	89.1
SVM	0.5	98.9	95.8	97.3	42.4	74.1	53.9	70.6	85.0	75.6
LR	0.5	99.7	98.7	99.2	74.8	92.2	82.6	87.2	95.5	90.9
SVM-FS	0.5	98.9	97.9	98.4	60.0	75.0	66.7	79.5	86.5	82.5
LR-FS	0.5	99.6	97.6	98.5	60.5	89.7	72.2	80.0	93.6	85.4
SVM	1	99.1	96.6	97.8	49.5	79.3	60.9	74.3	88.0	79.4
LR	1	99.7	98.4	99.1	71.2	94.0	81.0	85.5	96.2	90.0
SVM-FS	1	99.4	98.0	98.7	64.7	85.4	73.6	82.0	91.7	86.2
LR-FS	1	99.5	98.7	99.1	73.6	88.8	80.5	86.5	93.7	89.8

present one second before the threshold crossing and can be captured by the extracted features. From a control perspective, a pre-peak horizon of 1 s provides sufficient time for the controller to perform a mitigation maneuver without compromising detection accuracy. Whether longer horizons could further enhance predictive performance remains an open research question. Extending the horizon may capture additional precursor dynamics, although this could also introduce increased uncertainty in the predictions. The feature selection (FS) strategy does not consistently improve classification performance in-domain. While in some cases it increases faulty precision (e.g., SVM at $t_{\text{pre}} = 0.2\text{ s}$), it may also reduce faulty recall (e.g., LR at $t_{\text{pre}} = 0.5\text{ s}$ and 1 s). Overall, the full feature set achieves equal or higher macro F1-scores in most configurations. Removing features may discard information that contributes to class separability. The subsets of features identified by the FS procedure are reported in Table 5 for completeness. But, for the above reasons, the complete feature representation is retained in the final model for subsequent analyses.

4.2 Out-of-domain test

The out-of-domain test uses flight data from the Kitepower V9 kite during a flight conducted on 5 June 2024, described in Sect. 2.4 and illustrated in the frequency domain in Fig. 3. The recorded flight spans 6.2 h from 11:19 until 17:31 UTC,



Table 5. Features selected by the SVM-based feature selection procedure with $t_{\text{pre}} = 1$ s.

Features 1-20				Features 21-39			
n	Signal	Feature	Criterion	n	Signal	Feature	Criterion
1	V_{kite}	root mean square	boundary	21	Tether length	time to maximum value	boundary
2	V_{kite}	energy	margin	22	Tether length	median	boundary
3	V_{kite}	MAD	boundary	23	Tether length	time reversal asymmetry	margin
4	V_{kite}	kurtosis	margin	24	Tether length	max. signal slope	boundary
5	V_{kite}	CSR	margin	25	Tether force	zero-crossing rate	margin+boundary
6	V_{kite}	FFT min. amplitude	margin	26	Tether force	low-freq. energy ratio	margin
7	$V_{x,\text{kite}}$	standard deviation	boundary	27	Position east	minimum	boundary
8	$V_{x,\text{kite}}$	maximum	margin	28	Position north	zero-crossing rate	margin
9	$V_{x,\text{kite}}$	90th percentile	margin	29	Kite elevation	root mean square	boundary
10	$V_{y,\text{kite}}$	mean	boundary	30	Kite elevation	interquartile range	boundary
11	$V_{y,\text{kite}}$	standard deviation	margin	31	Kite distance	mean	boundary
12	$V_{y,\text{kite}}$	minimum	margin	32	Kite distance	energy	boundary
13	$V_{y,\text{kite}}$	kurtosis	margin	33	Kite distance	low-freq. energy ratio	boundary
14	$V_{y,\text{kite}}$	FFT max. ampl. above 1 Hz	margin	34	Kite course	time to maximum value	margin
15	$V_{z,\text{kite}}$	mean	margin	35	Kite steering	time reversal asymmetry	boundary
16	$V_{z,\text{kite}}$	median	margin	36	Kite steering	FFT max. ampl. above 1 Hz	boundary
17	$V_{z,\text{kite}}$	kurtosis	margin	37	Kite depower	root mean square	boundary
18	Kite altitude	MAD	boundary	38	Kite depower	crest	margin
19	Kite altitude	minimum	margin	39	Ground m. power	maximum	boundary
20	Kite altitude	kurtosis	boundary				

during which the kite completed a total of 145 pumping cycles. Wind conditions were variable throughout the flight, with a sudden wind gust occurring around 13:40 UTC. The wind direction was also variable, predominantly coming from the east. Table 6 summarizes the out-of-domain classification performance. All models maintain very high precision and recall for the non-faulty class ($> 94\%$) indicating that normal flight behavior remains well characterized despite the distribution shift. The main difficulty lies in detecting the faulty class, which is even more imbalanced in the out-of-domain data (1.5%). At the shortest prediction horizon ($t_{\text{pre}} = 0.2$ s), both SVM and LR without feature selection achieve high recall but suffer from very low precision, indicating a large number of false alarms. Feature selection improves the balance between the two classes at this horizon. As the prediction horizon increases, LR achieves higher performance than SVM. The best overall performance is obtained with LR at $t_{\text{pre}} = 1$ s without feature selection.

In the following, we provide a more detailed analysis on this best configuration, including an analysis of the confidence associated with the predicted probabilities and the kite dynamics preceding critical load events. Figure 10(a) presents the



Table 6. Out-of-domain classification performance for different models and pre-peak times.

Out-of-domain dataset										
Model	Pre-peak (s)	Non-faulty (2705 samples, 98.5 %)			Faulty (41 samples, 1.5 %)			Macro avg.		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
SVM	0.2	99.8	94.2	96.9	19.1	90.2	31.5	59.5	92.2	64.2
LR	0.2	99.8	95.5	97.6	22.8	87.8	36.2	61.3	91.6	66.9
SVM-FS	0.2	99.5	99.5	99.5	66.7	68.3	67.5	83.1	83.9	83.5
LR-FS	0.2	99.6	99.4	99.5	65.9	70.7	68.2	82.7	85.1	83.9
SVM	0.5	99.3	97.2	98.2	22.4	53.7	31.7	60.9	75.4	64.9
LR	0.5	99.6	99.9	99.7	90.9	73.2	81.1	95.3	86.5	90.4
SVM-FS	0.5	99.1	99.0	99.1	40.9	43.9	42.4	70.0	71.5	70.7
LR-FS	0.5	99.7	99.2	99.4	60.0	80.5	68.8	79.9	89.8	84.1
SVM	1	99.4	98.4	98.9	35.8	58.5	44.4	67.6	78.5	71.7
LR	1	99.8	99.9	99.8	97.2	85.4	90.9	98.5	92.7	95.4
SVM-FS	1	99.6	99.2	99.4	59.6	75.6	66.7	79.6	87.4	83.0
LR-FS	1	99.7	99.7	99.7	79.1	83.0	81.0	89.4	91.3	90.3

confusion matrix, which shows 35 true positives (TP), 1 false positive (FP), 6 false negatives (FN), and 2701 true negatives (TN). The performance metrics are computed as

$$365 \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 97.2\%, \quad (3)$$

$$\quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 85.4\%, \quad (4)$$

$$\quad \text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 90.9\%. \quad (5)$$

For the faulty class, using a decision threshold of 0.868, precision of 97.2% and recall of 85.4% indicate strong detection capability while maintaining low false alarm rates. In practical applications, the trade-off between precision and recall can be adjusted based on operational requirements. For low-cost maneuvers where early warning is prioritized over false alarm minimization, a higher recall (at the potential expense of precision) may be preferred to ensure critical events are not missed.

To understand the model's confidence in its predictions, Fig. 10(b) shows the distribution of predicted probabilities for both classes. The probability $p = P(y = 1 | \mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^{1 \times (d \times s)}$ is a feature vector from the feature matrix $\mathbf{X}_{\text{tot}}$, represents the probability of a sample belonging to the faulty class, while $1 - p$ corresponds to the non-faulty class. The predicted probabilities show that non-faulty samples are concentrated near 0 and faulty samples near 1, with almost no overlap. This indicates clear



class separation and high confidence in the model's predictions, despite the strong class imbalance. As previously mentioned, the decision threshold can be tuned to prioritize recall over precision for low-cost maneuvers. By lowering the threshold from its current value, the model would capture more of the faulty samples, reducing false negatives at the expense of increased false positives.

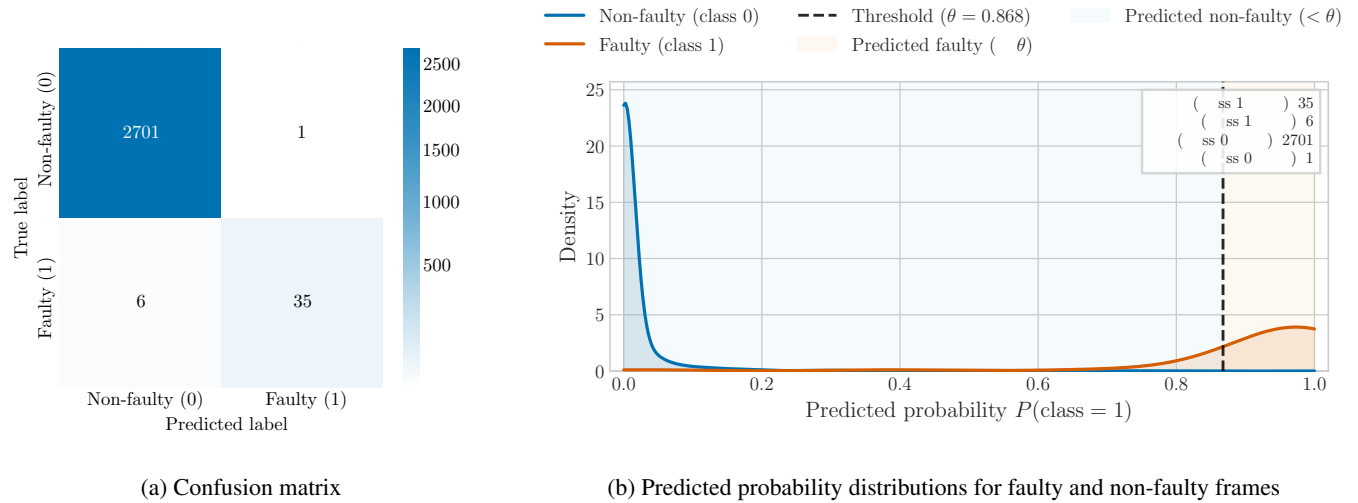


Figure 10. Out-of-domain performance of LR with $t_{\text{pre}} = 1$ s.

380 The signal characteristics preceding detected critical events are illustrated in Fig. 11, which shows overlapped trajectories corresponding to true positive detections. Each subplot shows how the individual trajectories evolve over time, together with a representative median signal. For the sake of clarity, only a subset of the measured variables is presented. The selected signals correspond to those identified as most relevant by the feature selection ranking, along with variables exhibiting consistent and interpretable temporal patterns prior to critical events.

385 The detected events reveal recurrent dynamic behavior preceding critical load conditions. In particular, tether reel-out speed, tether length, and kite speed exhibit a progressive increase as the force peak is approached. A corresponding increase in ground mechanical power is also observed, consistent with the simultaneous rise in tether force and reel-out speed.

Almost all true positive cases occur during the reel-out phase, specifically as the kite exits the downloop turn of the figure-eight trajectory, corresponding to a kite course of approximately 90° or 270° , as illustrated in Fig. 12. At this stage, the kite
 390 flies at low elevation in a nearly horizontal trajectory toward the center of the wind window, reaching high speed and maximum aerodynamic loading. A possible explanation is that, during the aggressive turning maneuver, the kite becomes temporarily over-steered, causing partial flow separation or local stall on parts of the wing and a transient reduction in aerodynamic force. As the kite exits the turn and the steering input decreases, the airflow reattaches to the wing, leading to a rapid recovery of lift and a sharp increase in tether force. This effect is amplified because the steering reduction occurs precisely in the power-
 395 generating region of the trajectory, where apparent wind speed is highest. This behavior is consistent with the studies of the V9 kite flight dynamics by Cayon et al. (2026), which show that the maximum tether force occurs shortly after the points of



highest kite speed. The two visible outliers correspond instead to transition phases between operating modes and therefore do not follow the typical crosswind maneuvering dynamics during reel out.

400 For the out-of-domain test flight, vertical wind speed profile data measured by a LiDAR are available. They are used here to investigate further the relationship between the ambient wind speed and the occurrence of tether force peak events. Figure 13 shows the measured wind speed as a function of altitude, together with the wind measurements associated with detected tether force peaks. The results confirm that tether force peak events are generally associated with higher wind speeds compared to normal operating conditions. In addition, these events tend to occur at relatively lower altitudes.

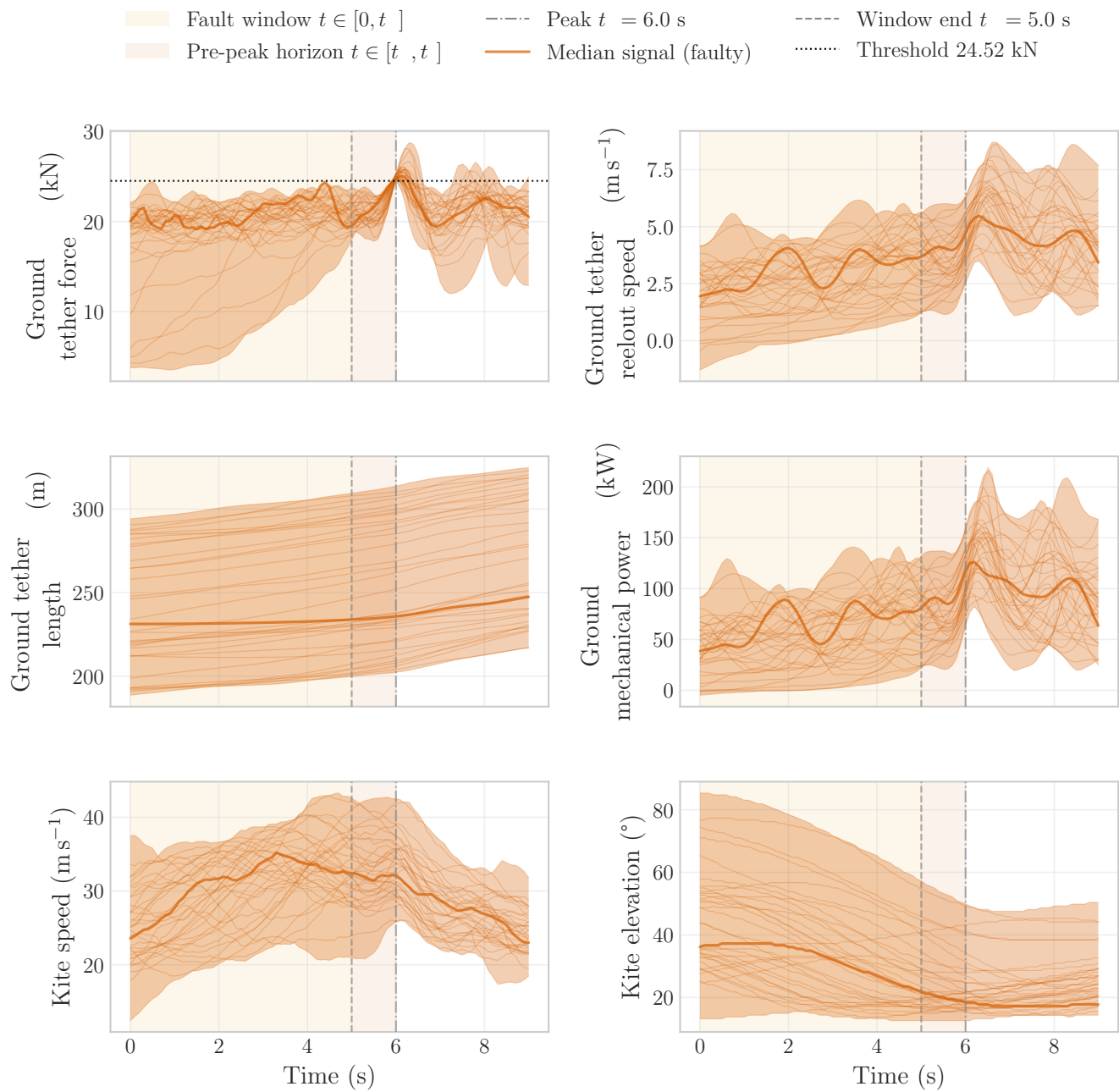


Figure 11. Overlapped true positive (TP) signals from the test dataset. Each subplot shows all TP samples for a given signal, with a shaded envelope (min-max range), a highlighted median-representative signal, and individual trajectories in lighter transparency.

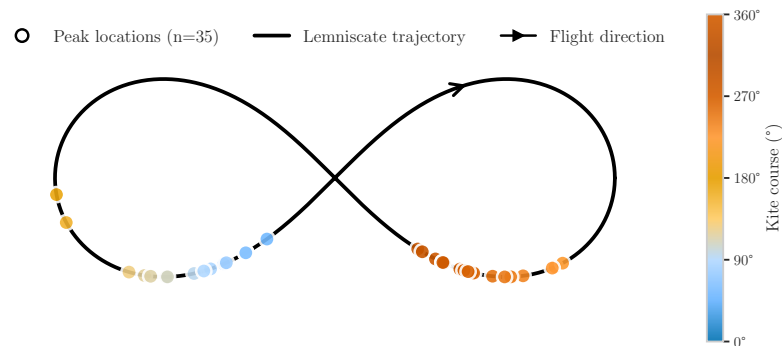


Figure 12. Lemniscate trajectory showing tether force peaks predicted by the model.

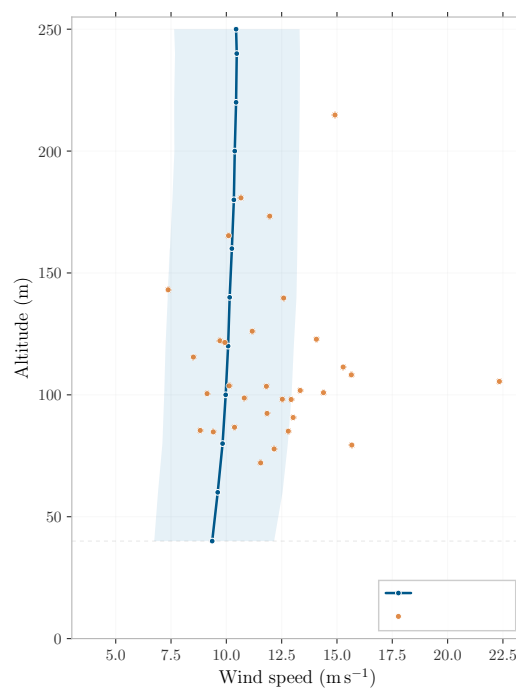


Figure 13. Vertical wind speed profile measured with LiDAR for the out-of-domain test flight on 5 June 2024. Scatter points indicate the altitude and wind speed corresponding to detected tether force peaks.

5 Conclusions

405 This study presented a data-driven approach for the early prediction of critical loads in airborne wind energy systems. The approach was developed using a large-scale SCADA dataset from the Kitepower Falcon system, comprising more than 1000 pumping cycles across different operating conditions. The performance of the machine learning model was demonstrated with



an additional SCADA dataset, covering 145 pumping cycles. From pre-peak time windows, statistical, temporal, and spectral features of multiple kite, tether, and ground-station measurements were extracted and used to train supervised classification models.

The results show that the proposed method can reliably detect the system dynamics that lead to critical load events up to 1 s before the tether force peak occurs. Logistic regression achieved the best overall performance, correctly identifying 85.4 % of the critical load events and achieving an F1-score of 90.9 %, indicating that the model detects most critical events while generating few false alarms. The method demonstrated strong generalization capability on the out-of-domain flight data, collected under different wind and operating conditions and including a stronger imbalanced faulty class of only 1.5 % of the total samples. Robust performance under these conditions is critical, as the final goal of the machine learning predictor is to operate in real-time on the kite system with newly collected flight data, where operating conditions and environmental factors may differ substantially from those represented in the training dataset. The ability of the model to provide reliable predictions despite extreme class imbalance and shifts in data distribution highlights its suitability for deployment in real-world scenarios.

Additionally, the analysis of the detected events revealed consistent signal patterns preceding critical load situations. The predicted force peaks were concentrated in regions of the lemniscate trajectory where the kite flies at low elevation and is getting out of the turns. From a control perspective, anticipating tether force peaks one second in advance provides sufficient reaction time for preventive actions. The model predictions could be integrated directly into the kite's control architecture, enabling feedforward adjustments such as modifying winch torque or the kite trajectory. This integration would allow the system to proactively mitigate critical loads, complementing existing feedback controllers and enhancing overall operational safety and reliability.

Importantly, the model does not rely on wind measurements, which were not available for our study, and that will neither be provided with the commercial Kitepower system. Despite this limitation, the method still achieves strong predictive performance, demonstrating that critical load events can be anticipated using only kite, tether, and ground-station signals. Once available, incorporating reliable wind measurements could further enhance predictive capability and enable more detailed risk characterization, including identifying the wind conditions most likely to trigger critical load events.

In this study, we focused on the early prediction of tether force peaks as a representative critical event to demonstrate the performance of the proposed model and the effectiveness of the overall methodology. While tether force peaks were chosen due to their operational relevance, the approach is generalizable and can be applied to other critical situations, such as preventing stall effect or unsafe kite trajectories.

Data availability

The code used to generate the results presented in this paper is maintained in <https://github.com/awegroup/awe-event-predictor>. The dataset of Kitepower's V9 kite from 5 June 2024 used for the out-of-domain model evaluation is publicly accessible from <https://github.com/awegroup/Flightdata05062024> (also: <https://doi.org/10.5281/zenodo.14238586>). Other datasets of the V9 kite in similar operational settings are accessible from <https://github.com/awegroup/Flightdata12052023> (also: <https://doi.org/10.5281/zenodo.14238586>).



[org/10.5281/zenodo.17246215](https://doi.org/10.5281/zenodo.17246215)) and <https://github.com/awegroup/Flightdata27112023> (also: <https://doi.org/10.5281/zenodo.14237421>).

Author contributions

Conceptualization, M.P., O.C., R.S., S.T., C.F., T.A., P.L.; methodology, M.P.; software, M.P.; investigation, M.P.; writing-
445 original draft preparation, M.P.; writing-review and editing, O.C., R.S., S.T., C.F., T.A., P.L.; supervision, R.S., S.T., C.F., T.A.,
P.L.; funding acquisition, R.S., C.F., T.A., P.L. All authors have read and agreed to the published version of the manuscript.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Wind Energy Science. R.S. is a co-founder of and advisor
for the start-up company Kitepower B.V., which is commercially developing a 100 kW kite power system and provided their
450 test facilities and staff for performing the in situ measurements described in this article.

Acknowledgments

This work has been supported by the French State as part of *France 2030* and the MERIDIONAL project, which receives
funding from the European Union's Horizon Europe Programme under the grant agreement No. 101084216. The authors also
gratefully acknowledge Kitepower B.V. for providing a SCADA dataset covering 26 h of flight with more than 1000 pumping
455 cycles of their V9 kite system. We also acknowledge the use of OpenAI's ChatGPT and DeepL for assistance in refining the
writing style of this manuscript.



References

- Altmann, A., Tološi, L., Sander, O., and Lengauer, T.: Permutation importance: a corrected feature importance measure, *Bioinformatics*, 26, 1340–1347, <https://doi.org/10.1093/bioinformatics/btq134>, 2010.
- 460 Bechtle, P., Schelbergen, M., Schmehl, R., Zillmann, U., and Watson, S.: Airborne wind energy resource analysis, *Renewable Energy*, 141, 1103–1116, <https://doi.org/10.1016/j.renene.2019.03.118>, 2019.
- Bradley, A. P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, 30, 1145–1159, [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2), 1997.
- Brank, J., Grobelnik, M., Milic-Frayling, N., and Mladenic, D.: Feature selection using support vector machines, *WIT Transactions on*
 465 *Information and Communication Technologies*, 28, <https://doi.org/10.2495/DATA020271>, 2002.
- Cayon, O., Watson, S., and Schmehl, R.: Kite as a Sensor: Wind and State Estimation in Tethered Flying Systems, *Wind Energy Science*, 10, 2161–2188, <https://doi.org/10.5194/wes-10-2161-2025>, 2025.
- Cayon, O., van Deursen, V., and Schmehl, R.: Translational Dynamics of Bridled Kites: A Reduced-Order Model in the Course Reference Frame, *Wind Energy Science*, 11, 1097–1121, <https://doi.org/10.5194/wes-11-1097-2026>, 2026.
- 470 Eijkelhof, D. and Schmehl, R.: Six-Degrees-of-Freedom Simulation Model for Future Multi-Megawatt Airborne Wind Energy Systems, *Renewable Energy*, 196, 137–150, <https://doi.org/10.1016/j.renene.2022.06.094>, 2022.
- Fagiano, L., Quack, M., Bauer, F., Carnel, L., and Oland, E.: Autonomous airborne wind energy systems: accomplishments and challenges, *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 603–631, <https://doi.org/10.1146/annurev-control-042820-124658>, 2022.
- 475 Fisher, A., Rudin, C., and Dominici, F.: All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously, *Journal of Machine Learning Research*, 20, 1–81, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8323609/>, (last access: 24 August 2026), 2019.
- Fulcher, B. D. and Jones, N. S.: Highly comparative feature-based time-series classification, *IEEE Transactions on Knowledge and Data Engineering*, 26, 3026–3037, <https://doi.org/10.1109/TKDE.2014.2316504>, 2014.
- 480 Goutte, C. and Gaussier, E.: A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation, in: *Advances in Information Retrieval*, edited by Losada, D. E. and Fernández-Luna, J. M., pp. 345–359, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-31865-1_25, 2005.
- Gupta, A., Kashyap, Y., and Kosmopoulos, P.: Tether Force Estimation Airborne Kite Using Machine Learning Methods, *Wind*, 5, 5, <https://doi.org/10.3390/wind5010005>, 2025.
- 485 Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3, 1157–1182, <https://dl.acm.org/doi/10.5555/944919.944968>, (last access: 24 August 2026), 2003.
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V.: Gene selection for cancer classification using support vector machines, *Machine Learning*, 46, 389–422, <https://doi.org/10.1023/A:1012487302797>, 2002.
- Horn, G., Gros, S., and Diehl, M.: Numerical Trajectory Optimization for Airborne Wind Energy Systems Described by High Fidelity
 490 *Aircraft Models*, in: *Airborne Wind Energy*, edited by Ahrens, U., Diehl, M., and Schmehl, R., *Green Energy and Technology*, chap. 11, pp. 205–218, Springer, Berlin Heidelberg, https://doi.org/10.1007/978-3-642-39965-7_11, 2013.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X.: *Applied logistic regression*, John Wiley & Sons, 3 edn., <https://doi.org/10.1002/9781118548387>, 2013.



- Hummel, J. I., Pollack, T. S., Eijkelhof, D., van Kampen, E.-J., and Schmehl, R.: Winch Sizing for Ground-Generation Airborne Wind Energy Systems, in: 2024 European Control Conference (ECC), pp. 675–680, IEEE, <https://doi.org/10.23919/ECC64448.2024.10590780>, 2024.
- Leuthold, R. C.: Multiple-Wake Vortex Lattice Method for Airborne Wind Energy Membrane-Wing Kites: A Literature Review, Student report, Delft University of Technology, <https://resolver.tudelft.nl/uuid:bad76ade-81d3-4d86-a631-c45319812445>, (last access: 27 August 2026), 2015.
- Licitra, G., Koenemann, J., Bürger, A., Williams, P., Ruiterkamp, R., and Diehl, M.: Performance assessment of a rigid wing Airborne Wind Energy pumping system, *Energy*, 173, 569–585, <https://doi.org/10.1016/j.energy.2019.02.064>, 2019.
- Noble, W. S.: What is a support vector machine?, *Nature Biotechnology*, 24, 1565–1567, <https://doi.org/10.1038/nbt1206-1565>, 2006.
- Oehler, J. and Schmehl, R.: Aerodynamic Characterization of a Soft Kite by in Situ Flow Measurement, *Wind Energy Science*, 4, 1–21, <https://doi.org/10.5194/wes-4-1-2019>, 2019.
- Peschel, J.: A Cost-Effective Kite State Estimator for Reliable Automatic Control of Kites, Master’s thesis, TU Berlin, <https://doi.org/10.5281/zenodo.7864661>, 2013.
- Platt, J. C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Advances in Large Margin Classifiers*, 10, 61–74, 1999.
- Rapp, S. and Schmehl, R.: Enhancing resilience of airborne wind energy systems through upset condition avoidance, *Journal of Guidance, Control, and Dynamics*, 44, 251–265, <https://doi.org/10.2514/1.G005189>, 2021.
- Rapp, S., Schmehl, R., Oland, E., and Haas, T.: Cascaded Pumping Cycle Control for Rigid Wing Airborne Wind Energy Systems, *Journal of Guidance, Control, and Dynamics*, 42, 2456–2473, <https://doi.org/10.2514/1.G004246>, 2019.
- Rushdi, M. A., Rushdi, A. A., Dief, T. N., Halawa, A. M., Yoshida, S., and Schmehl, R.: Power prediction of airborne wind energy systems using multivariate machine learning, *Energies*, 13, 2367, <https://doi.org/10.3390/en13092367>, 2020.
- Salma, V. and Schmehl, R.: Flight Anomaly Detection for Airborne Wind Energy Systems, *Journal of Physics: Conference Series*, 1618, 032 021, <https://doi.org/10.1088/1742-6596/1618/3/032021>, 2020.
- Salma, V., Friedl, F., and Schmehl, R.: Improving reliability and safety of airborne wind energy systems, *Wind Energy*, 23, 340–356, <https://doi.org/10.1002/we.2433>, 2020.
- Schelbergen, M. and Schmehl, R.: Validation of the Quasi-Steady Performance Model for Pumping Airborne Wind Energy Systems, *Journal of Physics: Conference Series*, 1618, 032 003, <https://doi.org/10.1088/1742-6596/1618/3/032003>, 2020.
- Schelbergen, M. and Schmehl, R.: Swinging Motion of a Kite with Suspended Control Unit Flying Turning Manoeuvres, *Wind Energy Science*, 9, 1323–1344, <https://doi.org/10.5194/wes-9-1323-2024>, 2024.
- Schölkopf, B. and Smola, A. J.: *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT Press, <https://doi.org/10.7551/mitpress/4175.001.0001>, 2001.
- Sperandei, S.: Understanding logistic regression analysis, *Biochemia Medica*, 24, 12–18, <https://doi.org/10.11613/BM.2014.003>, 2014.
- Sujon, K. M., Hassan, R. B., Towshi, Z. T., Othman, M. A., Samad, M. A., and Choi, K.: When to use standardization and normalization: Empirical evidence from machine learning models and XAI, *IEEE Access*, 12, 135 300–135 314, <https://doi.org/10.1109/ACCESS.2024.3462434>, 2024.
- Suthaharan, S.: Support vector machine, in: *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, vol. 36 of *Integrated Series in Information Systems*, pp. 207–235, Springer, Boston, MA, https://doi.org/10.1007/978-1-4899-7641-3_9, 2016.



Trevisi, F., Castro-Fernández, I., Pasquinelli, G., Riboldi, C. E. D., and Croce, A.: Flight trajectory optimization of Fly-Gen airborne wind energy systems through a harmonic balance method, *Wind Energy Science*, 7, 2039–2058, <https://doi.org/10.5194/wes-7-2039-2022>, 2022.

535 Van der Vlugt, R., Peschel, J., and Schmehl, R.: Design and Experimental Characterization of a Pumping Kite Power System, in: *Airborne Wind Energy*, edited by Ahrens, U., Diehl, M., and Schmehl, R., *Green Energy and Technology*, chap. 23, pp. 403–425, Springer, Berlin Heidelberg, https://doi.org/10.1007/978-3-642-39965-7_23, 2013.

Vertovec, N., Ober-Blöbaum, S., and Margellos, K.: Safety-Aware Hybrid Control of Airborne Wind Energy Systems, *Journal of Guidance, Control, and Dynamics*, 47, 326–338, <https://doi.org/10.2514/1.G007254>, 2024.