Wasserstein distances for wind-turbine power-curve monitoring: a pre-specified simulation comparison

Singularity

November 2026 · preprint

Abstract. Wind-turbine performance monitoring compares recent SCADA records with a healthy reference, usually through a binned power curve or a residual statistic. These tools compare means; many faults instead change the *shape* of the distribution of power at a given wind speed. We compare five detectors on simulated 10-minute SCADA data with a fixed false-alarm rate of 1% per one-week window: the mean residual, a binned power-curve chi-square statistic, the Kolmogorov–Smirnov statistic on residuals, the 1-Wasserstein distance between residual distributions, and a 1-Wasserstein distance conditional on wind speed. Four fault families at four severities were specified before the single run: uniform derating, a shift of the power curve, a power cap and intermittent half-power operation. Averaged over the twelve non-trivial scenarios, detection rates were 73% (conditional Wasserstein), 71% (residual Wasserstein), 67% (binned chi-square), 54% (mean residual) and 28% (Kolmogorov–Smirnov). The Wasserstein detectors were best for derating and intermittent faults — for example 81–83% against 46% at a 2% intermittent share — while the binned chi-square was clearly best for a power cap (94% against 62% at a 5% cap). No detector dominates; the conditional Wasserstein distance is the best single choice in this benchmark and costs no more than a binned power curve. Code and results are released.

1. Introduction

Supervisory control and data acquisition (SCADA) systems record wind speed and active power every ten minutes. The standard way to see underperformance is to compare a recent power curve with a reference, either by bins of wind speed as in IEC 61400-12-1 or through the residuals of a fitted curve. Distance-based variants compare a recent window with a healthy model, for example with a Mahalanobis distance [1], or track power-curve profiles over time [2]. These statistics are built on means or on second moments. Faults such as intermittent curtailment, controller limits or sporadic sensor errors change the *distribution* of power at a given wind speed while moving its mean only a little.

The Wasserstein distance [3] compares whole distributions and is sensitive to where probability mass moves and how far. It is widely used in statistics and machine learning but, to our knowledge, it has not been evaluated as a power-curve monitoring statistic. This study was prompted by an automated forecast that the method would enter energy engineering; we test whether the transfer is useful, on a benchmark fixed in advance.

2. Methods

Simulated turbine. Rated power $P_r = 2$ MW; logistic power curve $P(v) = P_r/(1 + e^{-0.9(v - 9.5)})$ between cut-in 3 m/s and cut-out 25 m/s; Weibull wind speeds (shape 2, scale 8 m/s); multiplicative noise with standard deviation 4% and additive noise of 1% of P_r; power clipped to $[0, P_r]$.

Windows. A healthy reference of four weeks (4032 records) and test windows of one week (1008 records). **Faults.** Derating $P \to (1 - s)P$ with $s \in \{1, 2, 3, 5\}\%$; curve shift $v_{50} \to v_{50} + s$ with

$s \in \{0.1, 0.2, 0.3, 0.5\}$ m/s; power cap $P \to \min(P, (1-s)P_r)$ with $s \in \{2, 5, 8, 12\}\%$; intermittent half power for a share $s \in \{1, 2, 4, 8\}\%$ of records. **Detectors.** Residuals are taken against the reference median power in 0.5 m/s bins. (i) $|\bar{r}|$ over its standard error; (ii) $\sum_b n_b \bar{r}_b^2 / \sigma_b^2$ over bins; (iii) the two-sample Kolmogorov–Smirnov statistic between test and reference residuals; (iv) the 1-Wasserstein distance between the same two residual samples; (v) the conditional distance $\sum_b n_b W_1(P_b^{\text{test}}, P_b^{\text{ref}}) / \sum_b n_b$, where P_b are the powers recorded in bin b. **Protocol.** Thresholds are the 99th percentile of each statistic over 1000 healthy test windows. Detection rates are estimated on 300 faulty windows per scenario. All settings, including the random seed, were fixed before the run, which was executed once.

3. Results

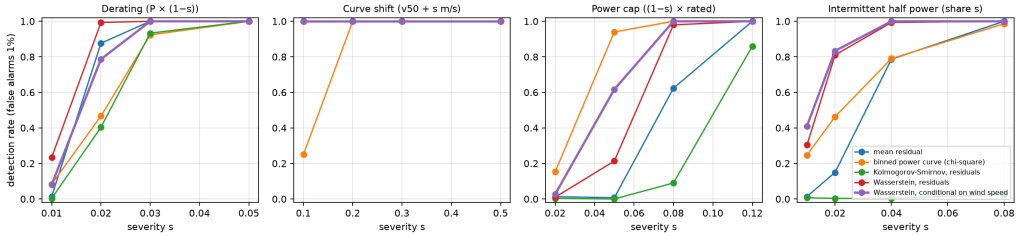


Figure 1. Detection rate at a 1% false-alarm rate, by fault family and severity.

Table 1. Mean detection rate over the four severities of each family. The curve-shift family is detected by every statistic from the smallest severity, except the binned chi-square at 0.1 m/s (25%), and is omitted from the mean.

detector	derating	power cap	intermittent	mean of 12
mean residual	0.72	0.41	0.49	0.54
binned chi-square	0.62	**0.77**	0.62	0.67
Kolmogorov–Smirnov	0.58	0.24	0.01	0.28
Wasserstein, residuals	**0.81**	0.55	0.78	0.71
Wasserstein, conditional	0.72	0.66	**0.81**	**0.73**

Figure 1 and Table 1 summarise the run. Three points stand out.

- **Shape-changing faults.** For intermittent half-power operation the Wasserstein detectors find a 2% share of affected records in 81–83% of windows, against 46% for the binned chi-square, 15% for the mean residual and 0% for Kolmogorov–Smirnov.
- **Small derating.** At 2% derating the residual Wasserstein distance detects 99% of windows, against 88% for the mean residual and 47% for the binned chi-square.
- **Where it loses.** A power cap acts only above rated wind speed, in few bins; the binned chi-square, which weights each bin by its own variance, is clearly better there (94% against 62% at a 5% cap).

The Kolmogorov–Smirnov statistic, which also compares distributions, performs poorly: it is driven by the largest gap between cumulative distributions and ignores how far mass moves, which is exactly what the Wasserstein distance measures.

4. Discussion

The conditional Wasserstein distance is a natural drop-in for a binned power curve: it uses the same bins and the same reference, and replaces a comparison of bin means with a comparison of bin distributions. In this benchmark it is the best single detector on average, and the residual version is the most sensitive to small uniform losses. A practical monitor can combine it with the binned chi-square, which remains best for faults confined to a few bins.

Limitations. The data are simulated, with independent records and a stationary wind climate; real SCADA data have autocorrelation, seasonal air density, wake effects and data gaps, which change the null distribution and call for block-bootstrap thresholds. The fault list is short and the severities were chosen by us. The next step is a test on public SCADA data with logged faults.

Reproducibility. `esperimento.py` regenerates every number and figure; `risultati.json` contains the output of the single run reported here.

Origin of the question. The question came from an automated forecast (Singularity v3, registered on 5 November 2026 with a SHA-256 fingerprint) that the Wasserstein distance would enter the subfield *Energy Engineering and Power Technology*. The forecast concerned the arrival of the method, not the result of this study.

References

1. P. Guo et al., Wind turbine performance degradation monitoring using DPGMM and Mahalanobis distance, *Renewable Energy* (2022).
2. Y. Ding, *Data Science for Wind Energy*, CRC Press (2019).
3. C. Villani, *Optimal Transport: Old and New*, Springer (2009).
4. IEC 61400-12-1, *Power performance measurements of electricity producing wind turbines* (2017).