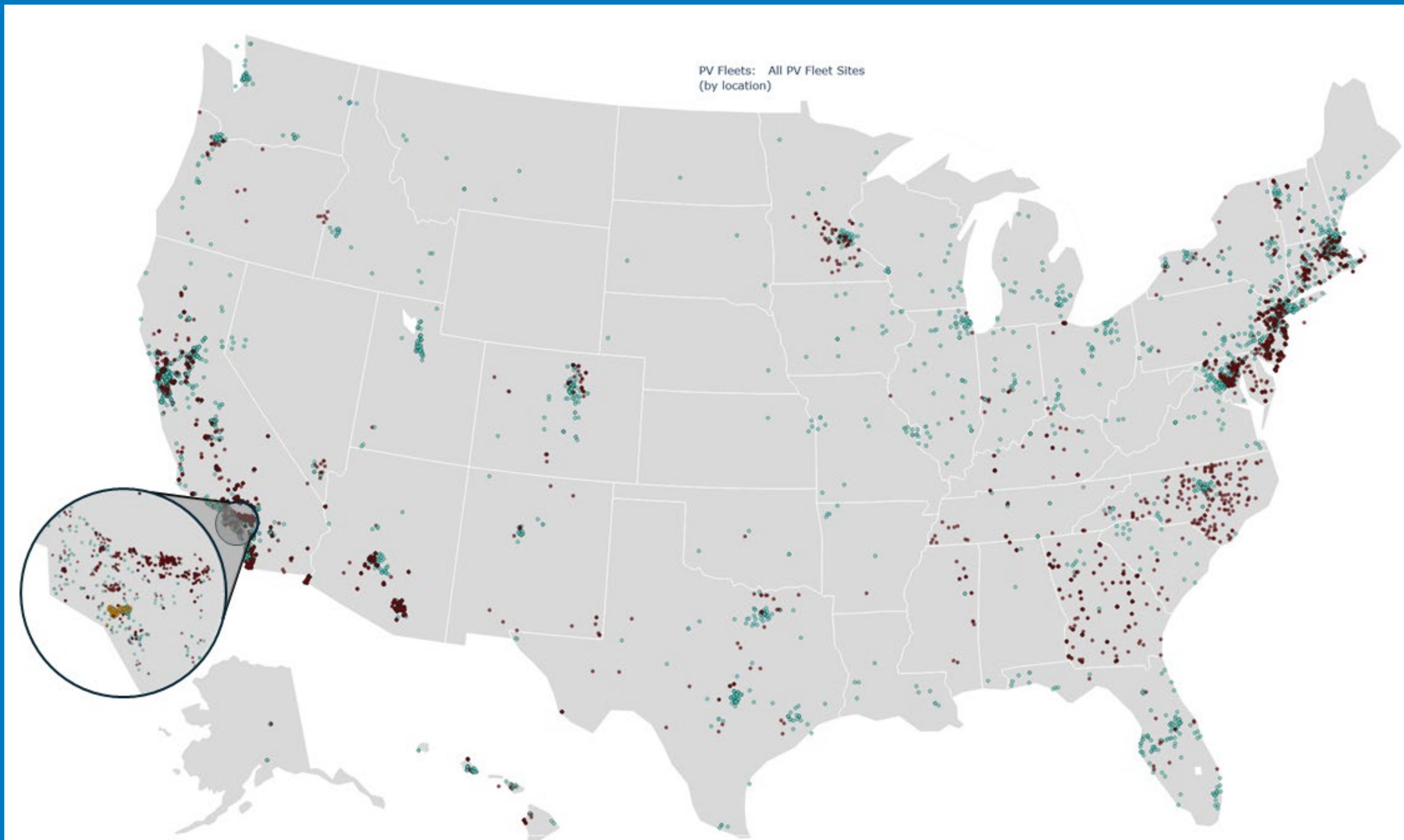


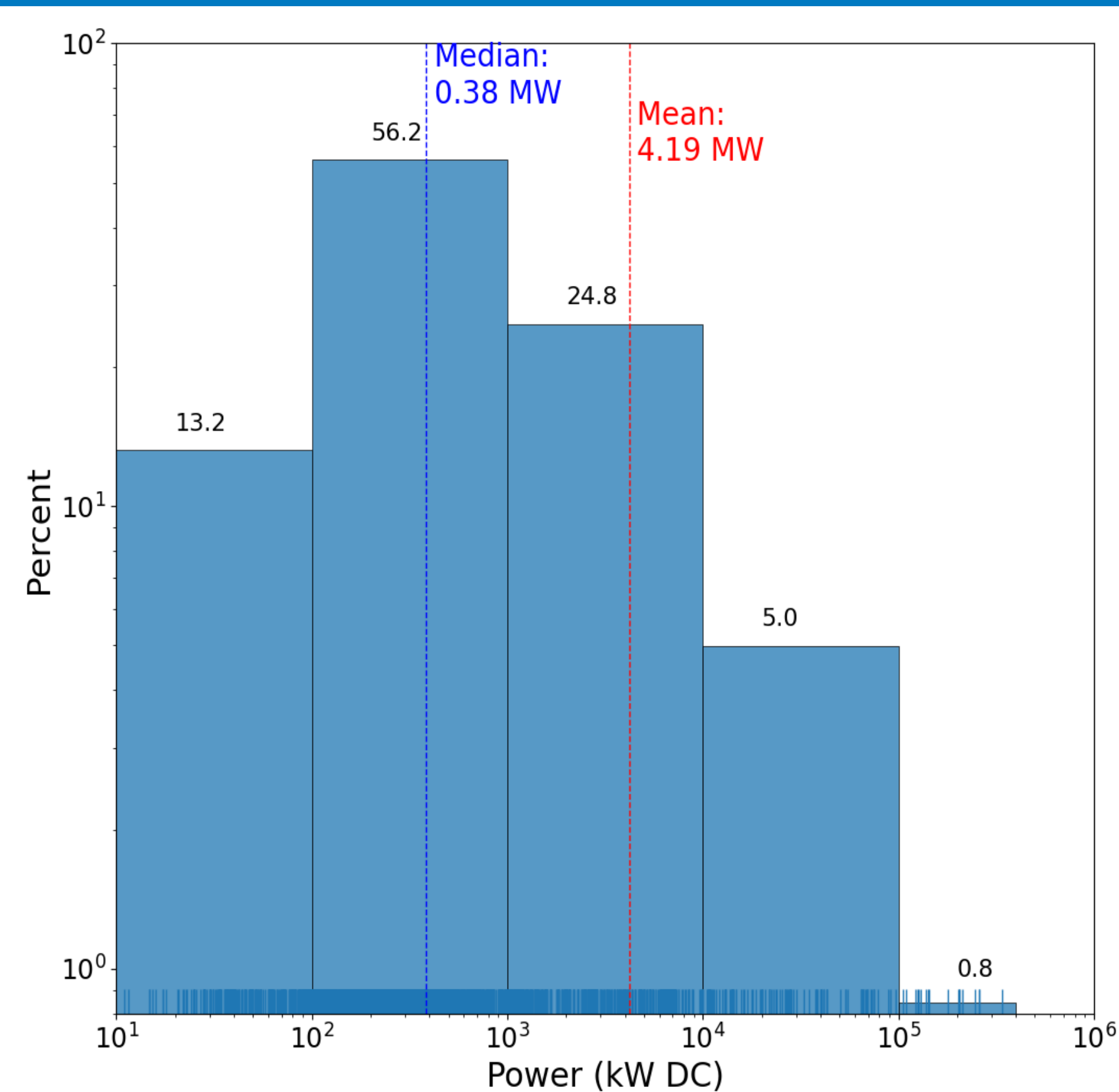
Scaling the PV Fleet Performance Data Initiative

Martin Springer*, Kirsten Perry, Michael Deceglie, Matt Muller, Dirk Jordan, Robert White, and Chris Deline
National Renewable Energy Laboratory, Golden, CO, USA | *martin.springer@nrel.gov

The PV Fleet Performance Data Initiative [1] partners with photovoltaic fleet owners to collect time-series data of PV production and publishes aggregated anonymized results of system performance metrics. We recently incorporated system performance data from pvoutput.org, which increased an already extensive dataset from 2,200 to over 5,000 PV systems across the United States. The initiative aims to ensure that systemic risks in the US PV fleet are detected.



PV Fleet systems distributed across the United States. The median system size is 0.38MW with a mean size of 4.19MW.



The database currently holds over

> 5000 systems

made up of over

> 29,000 inverters,

resulting in more than

>163,000 data channels

with a mean system age of

5.6 years

Parallelization of the analysis pipeline

Leveraging AWS cloud infrastructure and NREL's HPC capabilities lead to a reduction in computation time from more than 1 month to less than 8h for a full fleet analysis!



Speed up was achieved through the parallel and distributed computing library dask

A major focus of recent work was optimizing the analysis pipeline through parallelization and integration of high-performance computing procedures to allow the initiative to scale further while reducing computation time. The performance analysis pipeline was refactored to leverage the Dask distributed computing Python library, enabling parallel analysis of hundreds of systems simultaneously on NREL's high-performance computing cluster Kestrel. Data quality assurance processes were also refactored to take full advantage of the open-source PVAnalytics Python library and the new parallelization pipeline. Both upgrades significantly reduced computational times from weeks to hours for intensive analyses, such as combined soiling and degradation studies (CODS), but also allowed significant speed ups for year-over-year degradation analysis, system availability assessments, and performance index evaluations. We'll provide an update on those analysis results with the new set of system data.

Database



- raw time series data
- meta data
- QA results
- degradation / soiling analysis results
-

Storage



- storage for result files like plots or pickle files
- cache for HPC
-

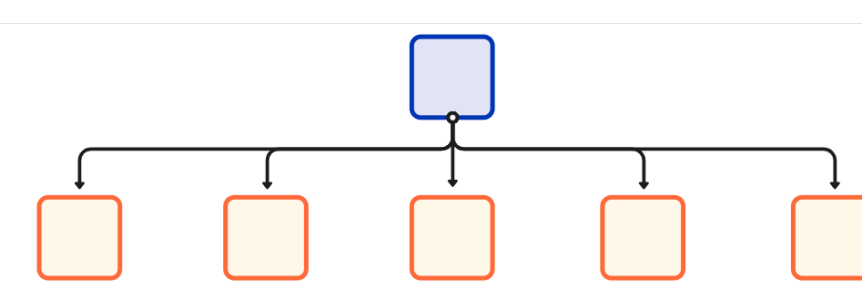
Compute



- parallelization
- processing of data
- running QA / degradation analysis
-

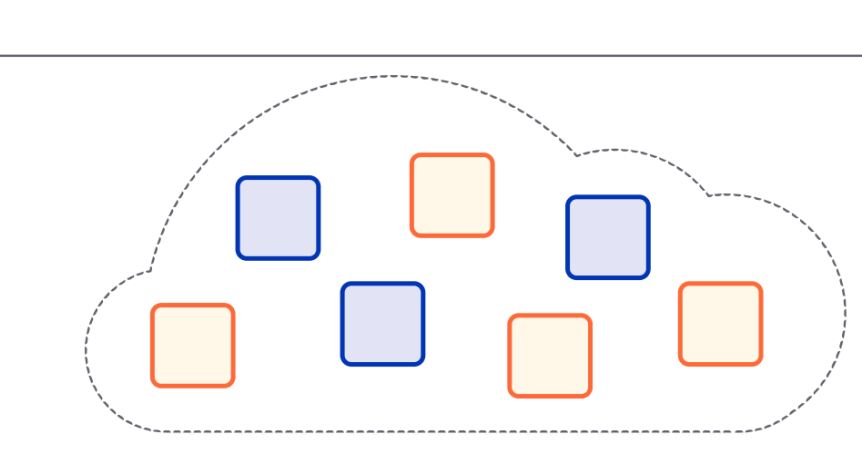
Cluster Manager

Deploys one Scheduler and many Workers by talking to the Resource Manager



Resource Manager

Kubernetes/Yarn/SLURM/PBS/Abstract pods/jobs on top of Physical Hardware



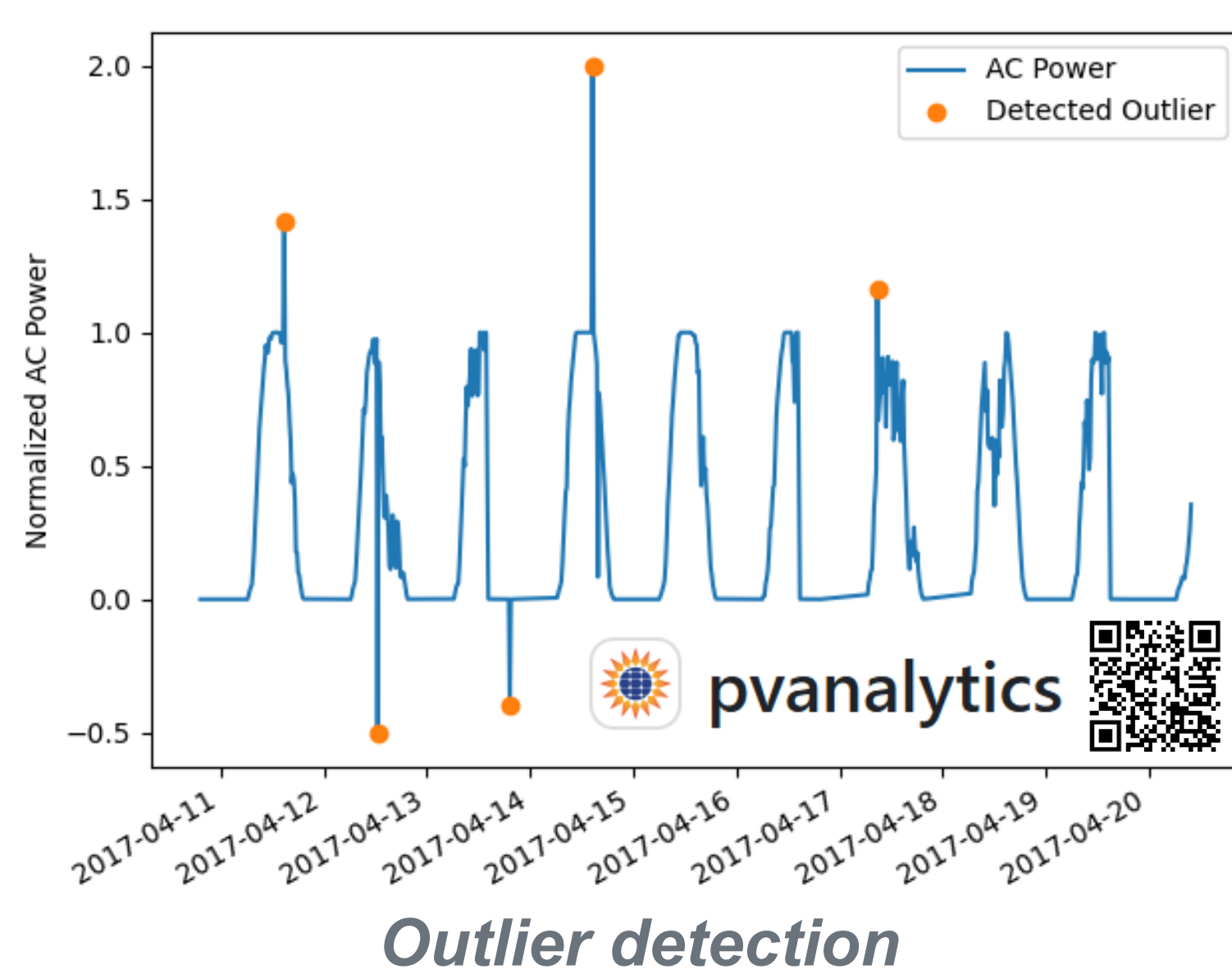
Physical Hardware

Physical CPUs, GPUs, networking and storage; either on-prem or on the cloud

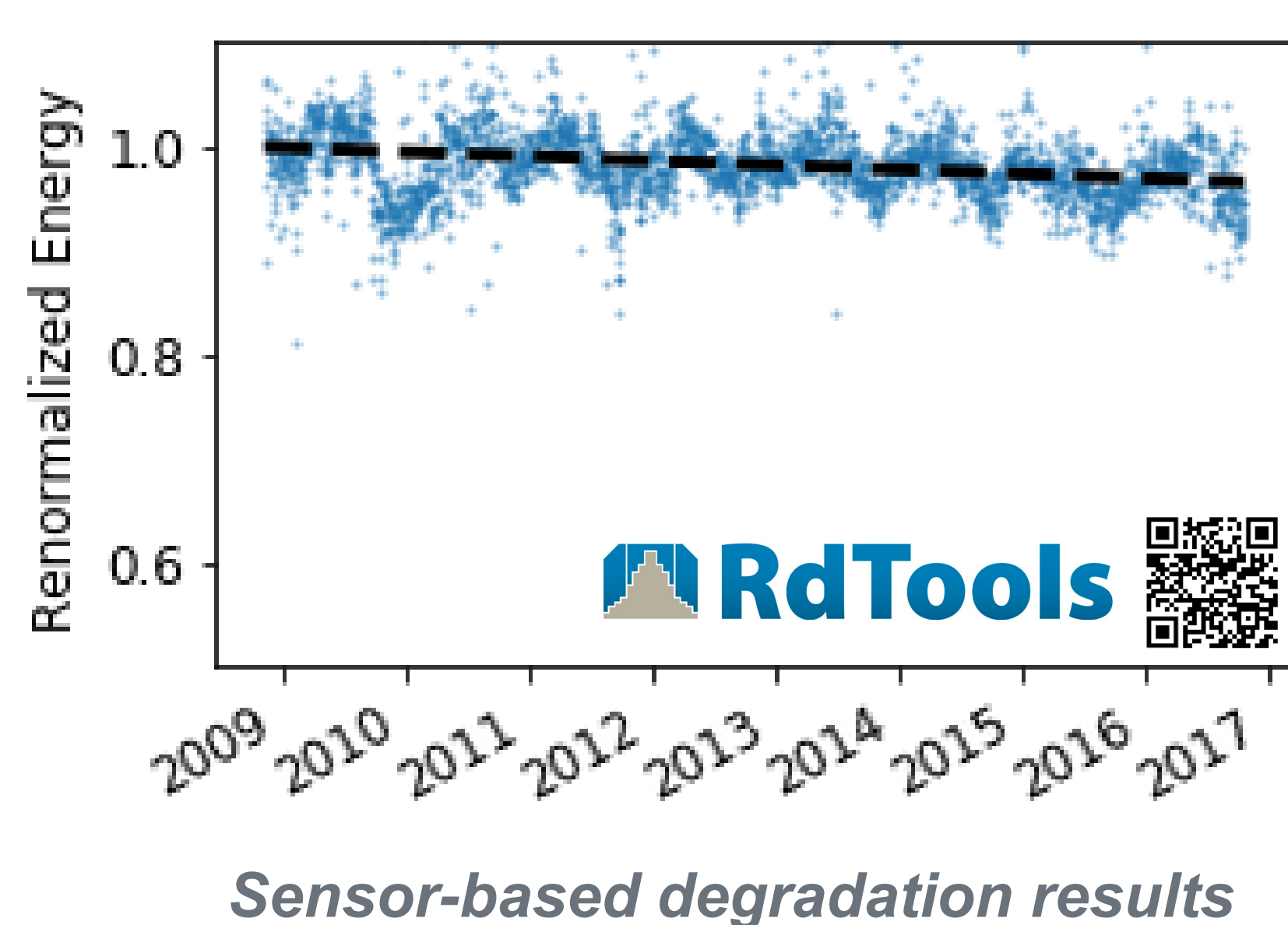


Pipeline workflow and stochastic modeling analysis

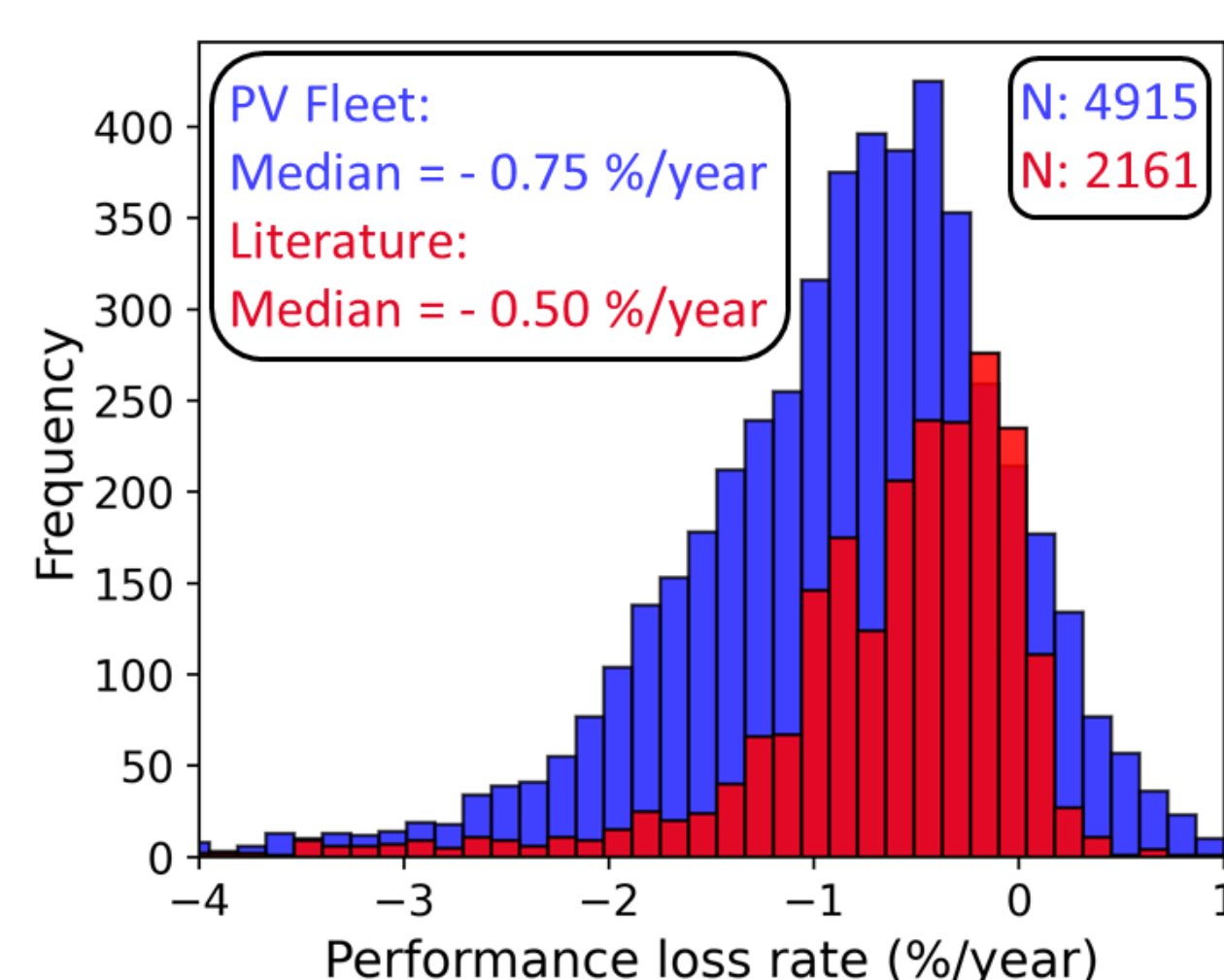
1) Data Quality Assurance (QA)



2) Performance loss rate



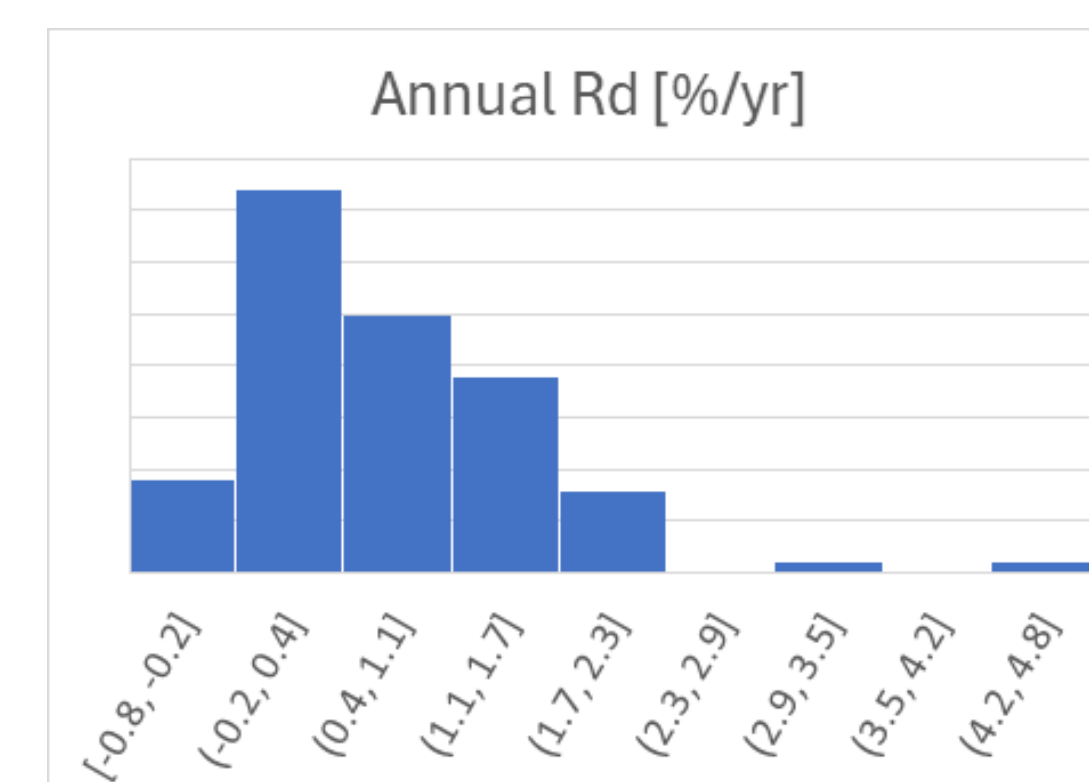
3) Data aggregation



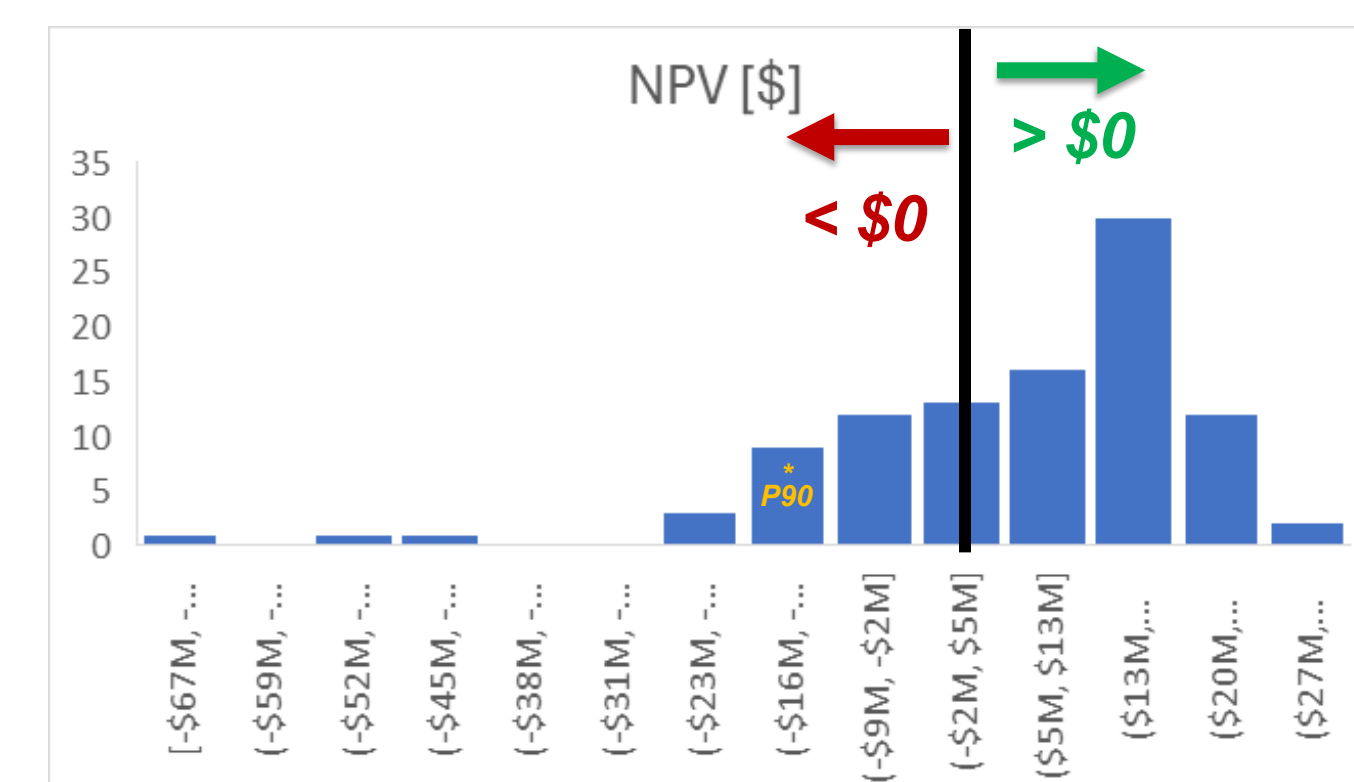
Performance loss rate for the PV Fleet initiative systems compared to literature

4) Stochastic financial modelling

Provides the prospective energy generation and financial viability for a new 100 MW PV system at a specific location.



Distribution of potential performance loss rate



Net present value to indicate potential financial outcomes

These pipeline enhancements enable more comprehensive analyses, such as parametric and stochastic analysis that were previously computationally unfeasible. Leveraging the stochastic modeling capability of SAM, we used 200 individual runs, along with loss factor distributions from the PV Fleet database, resulting in a series of financial output distributions. The main benefit of this type of analysis is our ability to determine alternative statistical values such as P75 or P90 financial outcomes. While the median Net Present Value was positive, risk scenarios such as P90 resulted in significant loss, in the case of, e.g. higher than average degradation rate or soiling loss. This new analysis provides an additional methodology to evaluate different risk scenarios for a given PV project, based on real-life measured loss factors and monthly performance deviations.

[1] C. Deline et al., "PV Fleet Performance Data Initiative (March 2020 Methodology Report)," National Renewable Energy Lab. (NREL), Golden, CO (United States), NREL/TP-5K00-76687, Jun. 2020. doi: 10.2172/1659786.

