Contextual Robust State Estimation in Distribution Systems

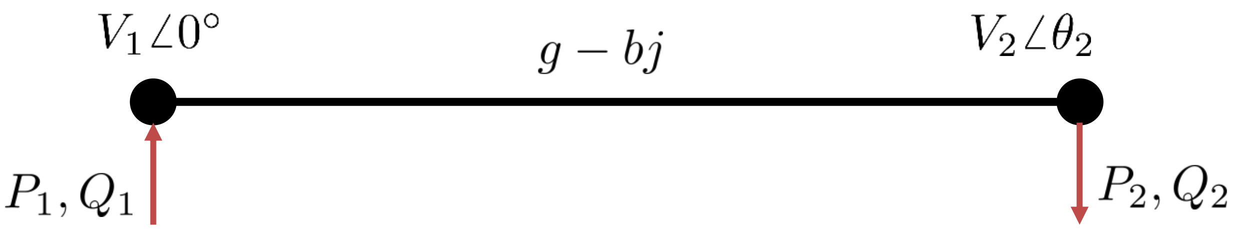
with Scarce Real-Time and Historical Data

Jose Gómez de la Varga, J. M. Morales, S. Pineda

OASYS Research Group · University of Málaga

IFORS 2026 · Vienna · New Trends in Stocastic Optimization applied to Energy Systems

State estimation in a nutshell



$$h_{P_1} = V_1V_2\left(-g\cos(-\theta_2) + b\sin(-\theta_2)\right) + gV_1^2$$
$$h_{Q_1} = V_1V_2\left(-g\sin(-\theta_2) - b\cos(-\theta_2)\right) - bV_1^2$$

$$h_{P_2} = V_2V_1\left(-g\cos\theta_2 + b\sin\theta_2\right) + gV_2^2$$
$$h_{Q_2} = V_2V_1\left(-g\sin\theta_2 - b\cos\theta_2\right) - bV_2^2$$

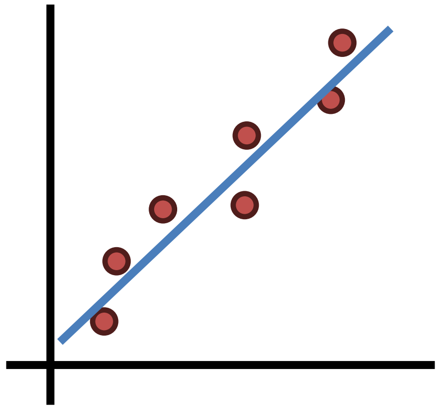
Measurements $\mathbf{z} = [V_1, V_2, \theta_2, P_1, Q_1, P_2, Q_2]^T$

State Vector $\mathbf{x} = [V_1, V_2, \theta_2]^T$

Weighted Least Squares (WLS)

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}} \, (\mathbf{z} - \mathbf{h}(\mathbf{x}))^\top \mathbf{W}(\mathbf{z} - \mathbf{h}(\mathbf{x}))$$

W = diagonal weight matrix (inverse error variances)



With redundancy...

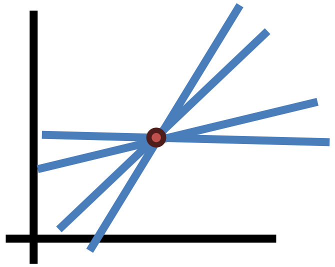
A double information shortage

Shortage #1: Real-Time Data

Fast & Sparse sensors: SCADA, μ-PMUs
Not enough real-time measurements to make the system observable

Slow & Abundant sensors: Smart Meters
Most data arrives with delays

Result:
no observability →



Shortage #2: Historical Data

Historical data is limited, often outdated, and non-representative

Networks evolve: new solar panels, topology changes, EV adoption shifts demand patterns

Result: statistical models trained on past data degrade quickly

$\mathbf{z}^a$ – available in real time

$\mathbf{z}^d$ – delayed, must be inferred from history

Our goal: a state estimator explicitly designed for both shortages simultaneously

Let similar past moments fill the gaps

Optimization and Analytics for Sustainable energY Systems

1 → **2** → **3**

Observe

Collect available real-time measurements $\mathbf{z}^a$

Search

Find K nearest neighbors in the historical dataset

Estimate

Use neighbors as pseudo-measurements for $\mathbf{z}^d$

Why K-Nearest Neighbors?

- Non-parametric: no model to train, no distributional assumptions
- Contextual: uses current real-time data (the context) to select relevant history
- Data-driven & Interpretable approximation of $P(\mathbf{z}^d|\mathbf{z}^a)$

The baseline: Vanilla K-NN state estimator

WLS-SE on the real-time data and the average of the K pseudo-measurement scenarios

Each delayed measurement in the training set → a plausible "scenario" for the missing data in the test

The weights reflect the sample variance of the neighbors: high variability across neighbors → low confidence → downweighted.

Lazy learning method: no offline training needed

Non-Robust (Vanilla) Estimator

$$\hat{\mathbf{x}}^v_{t'} \in \arg\min_{\mathbf{x}} \, (\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x}))^\top \mathbf{W}^a (\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x})) + \frac{1}{K} \sum_{t \in \mathcal{T}^K} (\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x}))^\top \widehat{\mathbf{W}}^d (\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x}))$$

$$\text{where } \mathcal{T}^K := \arg\min_{\substack{\mathcal{K} \subseteq \mathcal{T} \\ |\mathcal{K}| = K}} \sum_{k \in \mathcal{K}} (\mathbf{z}^a_{t'} - \mathbf{z}^a_k)^\top \widehat{\mathbf{W}}^a (\mathbf{z}^a_{t'} - \mathbf{z}^a_k)$$

⚠ But wait... This treats pseudo-measurements as if they were reliable. With scarce data, they carry significant statistical uncertainty!

The Contextual Robust State Estimator (CR-SE)

Core idea:
min over states & max over pseudo-measurement weights

$$
\begin{aligned}
\hat{\mathbf{x}}^r_{t'} \in \arg\min_{\mathbf{x}} \max_{w_t \geq 0} & \sum_{t \in \mathcal{T}^T} w_t \left(\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x})\right)^\top \mathbf{W}^a (\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x})) \\
& + \sum_{t \in \mathcal{T}^T} w_t \left(\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x})\right)^\top \widehat{\mathbf{W}}^d (\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x})) \\
& \text{s.t. } w_t \leq \frac{1}{K}, \ \forall t \in \mathcal{T}^T \quad : \lambda_t \\
& \sum_{t \in \mathcal{T}^T} w_t = 1 \quad : \mu_1 \\
& w_{t+1} - w_t \leq 0, \ \forall t \in \mathcal{T}^T \setminus \{t_{|\mathcal{T}^T|}\} \quad : \phi_t \\
& \sum_{t \in \mathcal{T}^T} w_t \left(\mathbf{z}^a_{t'} - \mathbf{z}^a_t\right)^\top \widehat{\mathbf{W}}^a (\mathbf{z}^a_{t'} - \mathbf{z}^a_t) \leq \rho \quad : \mu_2.
\end{aligned}
$$

- Protects against misspecification of the distribution
- At least *K* neighbors
- Valid distribution with finite discrete support
- Monotonically decresing
- Bounds the contextual distance
- Reformulated into an efficient equivalent single-level minimization problem

The Contextual Robust State Estimator (CR-SE)

Core idea:
min over states & max over pseudo-measurement weights

$$\hat{\mathbf{x}}^r_{t'} \in \underset{\mathbf{x},\,\lambda_t,\,\phi_t,\,\mu_1,\,\mu_2}{\arg\min} \quad \frac{1}{K}\sum_{t\in\mathcal{T}^T} \lambda_t + \mu_1 + \rho\,\mu_2$$

$$\begin{aligned}
\text{s.t. } & \lambda_t + \mu_1 + \mu_2\,(\mathbf{z}^a_{t'} - \mathbf{z}^a_t)^\top \mathbf{W}^a (\mathbf{z}^a_{t'} - \mathbf{z}^a_t) + f_t(\phi) \geq \\
& \quad (\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x}))^\top \mathbf{W}^a (\mathbf{z}^a_{t'} - \mathbf{h}^a(\mathbf{x})) \\
& \quad + (\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x}))^\top \widehat{\mathbf{W}}^d (\mathbf{z}^d_t - \mathbf{h}^d(\mathbf{x})),\ \forall t \in \mathcal{T}^T \\
& \mu_2 \geq 0 \\
& \lambda_t \geq 0,\ \forall t \in \mathcal{T}^T \\
& \phi_t \geq 0,\ \forall t \in \mathcal{T}^T \setminus \{t_{|\mathcal{T}^T|}\}
\end{aligned}$$

$$f_t(\phi) = \begin{cases} -\phi_t & \text{if } t = t_1 \\ \phi_{t-1} - \phi_t & \text{if } t_1 < t < t_{|\mathcal{T}^T|} \\ \phi_{t-1} & \text{if } t = t_{|\mathcal{T}^T|}. \end{cases}$$

- Protects against misspecification of the distribution
- At least *K* neighbors
- Valid distribution with finite discrete support
- Monotonically decresing
- Bounds the contextual distance
- Reformulated into an efficient equivalent single-level minimization problem

Overhead: $2|\mathcal{T}| + 1$ variables, $3|\mathcal{T}|$ constraints

Intuition: what does ρ control?

ρ governs the size of the ambiguity set:
¿how much the adversary can deviate from the empirical $1/K$ distribution?

$\rho = \rho_{min}$ $\rho = \rho_{max}$

Small ρ → Trust the data

- Ambiguity set is tight around the $1/K$ distribution.
- Behaves close to the Vanilla estimator.
- Efficient when data is representative.

Large ρ → Hedge your bets

- Wider ambiguity set allows more redistribution.
- Penalizes contextually distant neighbors.
- More conservative: better tail performance.

The sweet spot? We select ρ automatically using online validation

Selecting ρ: Data-driven + online

Define the search range

Grid search between ρ_{min} and ρ_{max}

Validate on the most recent instance

Retrospective state as ground truth

Pick the winner

Min validation error

Solve on the test instance

Optimization and Analytics for Sustainable energY Systems

Case study setup

IEEE 38-bus

Smaller, tractable baseline

+ IEEE 123-bus

Larger, tests scalability

Time-series with real residential data (Pecan Street)

Variables tested

- SCADA measurements: 1, 3, 5
- Training set size: 10, 20, 50
- Load factor: 0.5, 0.8, 1, 1.2
- 100 test instances per scenario

Metrics

RMSE – Root Mean Squared Error

- Overall estimation accuracy

RCVaR – Root Conditional Value at Risk ($\alpha = 0.95$)

- Tail performance: how bad are the worst 5% of errors?

Benchmarks

Vanilla K-NN + WLS no robustness

Persistent Most recent instance's state

Retrospective All sensors available

Oracle CR-SE validating on test

Results: CR-SE strictly outperforms Vanilla

Baseline scenario: 3 SCADA, Load = 0.5, training = 20, K = 5

	38-bus Network		123-bus Network	
Method	**RMSE ($\times 10^{-3}$)**	**RCVaR ($\times 10^{-3}$)**	**RMSE ($\times 10^{-3}$)**	**RCVaR ($\times 10^{-3}$)**
Vanilla	2.18	6.96	2.21	6.03
CR-SE	**2.13**	**6.70**	**2.16**	**5.62**
Persistent	4.45	14.62	3.69	9.85
Oracle	2.03	6.51	1.91	5.03
Retrospective	1.15	2.63	1.81	4.09

Does CR-SE close the gap between Vanilla and the Oracle?
RMSE = 19–30%, RCVaR = 41–57%

Where CR-SE shines most

Optimization and Analytics for Sustainable energY Systems

Least real-time measurements

Up to 9% RMSE
Up to 11% RCVaR

For the same context:

- More possible pseudo-measurements
- More different states

Growing training sets

Gap widens with more data

Larger window hurt Vanilla as farther operating conditions are included

CR-SE scales beneficially: it penalizes them via ρ

Heavy loading

Consistent performance across load scenarios

Stronger nonlinearities

Both methods worsen, CR-SE contains the degradation

The performance gap widens precisely in the configurations where the Vanilla is most vulnerable

Key takeaways

1. **Pseudo-measurements have uncertainty – treat it explicitly**
   Existing methods sistematically overlook pseudo-measurements statistical misspecification

2. **Distributionally robust optimization provides a principled fix**
   A min-max formulation hedges against the worst-case reweighting of pseudo-measurements within an ambiguity set

3. **Same tractability as the baseline**
   The robust formulation admits a single-level reformulation with the same structure and comparable computational cost

4. **Best gains where they matter most**
   CR-SE shines under the hardest scenarios: extreme unobservability, larger historical data, and heavy loading

Thank you!

Questions, ideas, collaborations?

Contextual Robust State Estimation in Distribution Systems

with Scarce Real-Time and Historical Data

Jose Gómez de la Varga
J. M. Morales · S. Pineda

OASYS