SpacSec 2026

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

Robustness of Smart Grid Load Forecasting Models to Adversarial Attacks

Presenter — Gleb Semenov, NCFU

Gleb Semenov[1], Maria Lapina[1], Amjad Gawanmeh[2], Vitalii Lapin[1] and Oleg Sanamyan[1]

[1] North-Caucasus Federal University, Stavropol, Russia

[2] University of Dubai, Dubai, UAE

Why the next 15 minutes matter

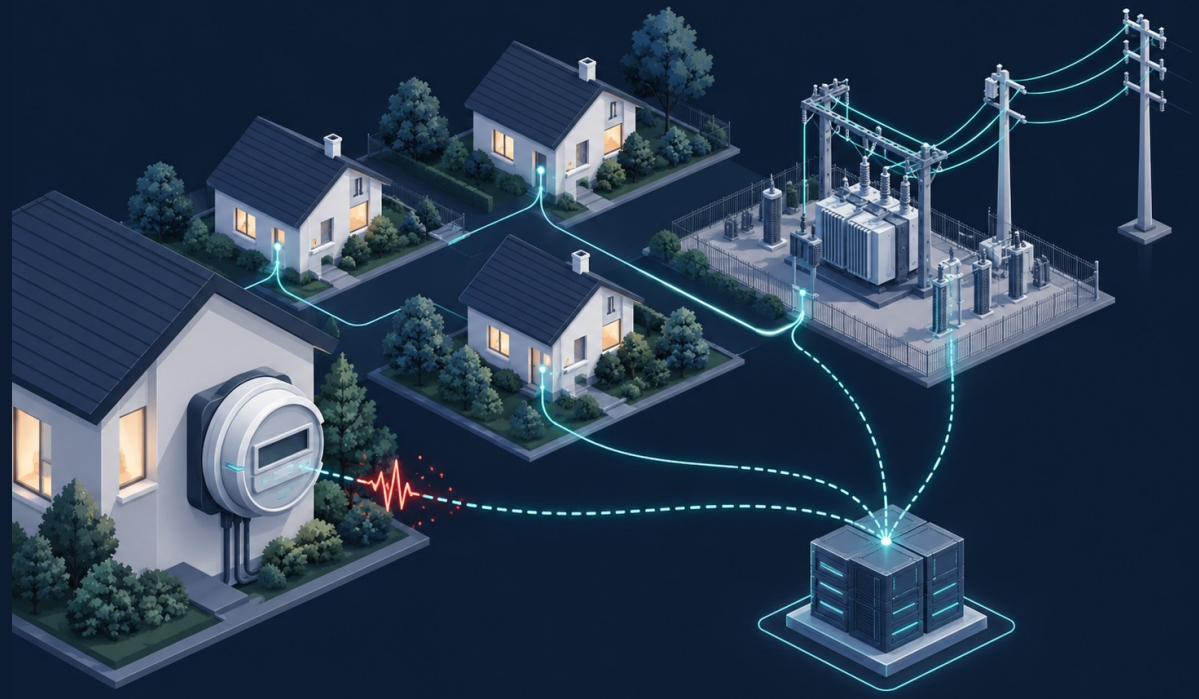
NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

An operator’s decision

Prepare reserve
Schedule battery use
Coordinate demand

Input reliability

Corrupted measurements
can distort the next forecast



Illustrative use case. We test the forecasting model.

The scope of our experiment

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

One forecasting module, tested offline

Scope	What is tested
Data	One historical dataset from Portugal
Targets	One meter and a 20-client aggregate
Forecast	96 past values to one next load value
Evidence	Forecast errors after numerical input changes

From yesterday to the next 15 minutes

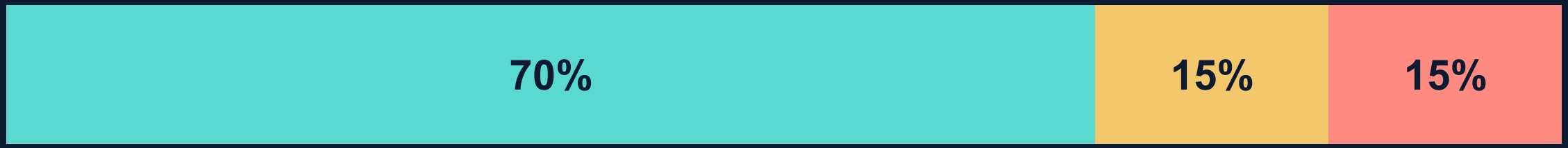
NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

96 measurements

Past 24 hours

1 measurement

Forecast 15 minutes ahead



Training	Validation	Test
98 112 windows	21 024 windows	21 024 windows

UCI Portugal data. Chronological split, without shuffling.

Adversarial changes to the input

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

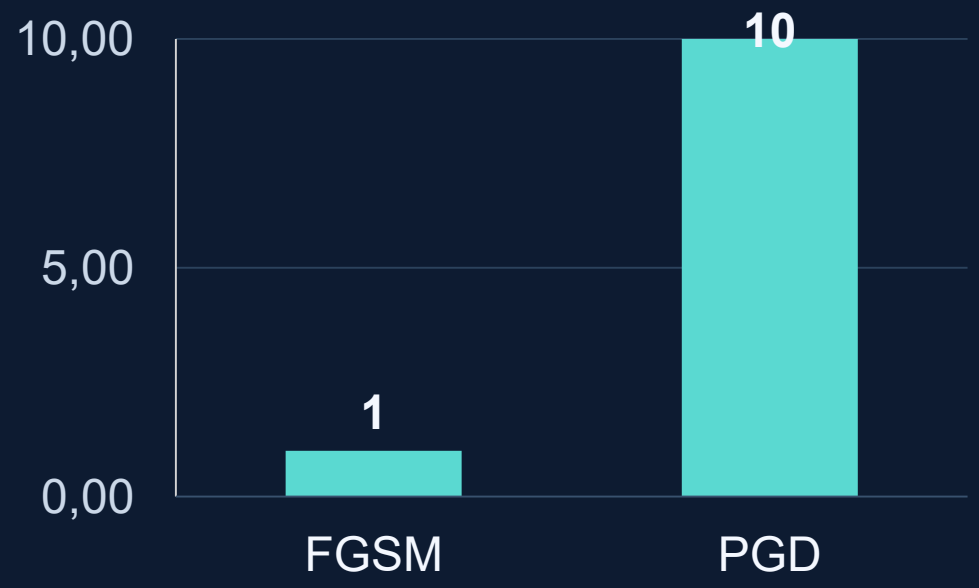
Question	Our test
What changes?	The history of 96 measurements
How is it chosen?	To increase the forecast MSE
What stays fixed?	Model weights during attack generation
How is it bounded?	Per-point ε budget and the [0,1] range
What is the output?	One number: the load forecast

Diagnostic test using the model, gradients and true next load

How we test the forecast

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

Attack steps



FGSM

One step using the gradient sign

PGD

α = ε / 4, with projection
No random start

ε budgets: 0.001 / 0.003 / 0.005 / 0.01 / 0.02

Similar FGSM and PGD results do not prove a complete attack search

What we compare

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

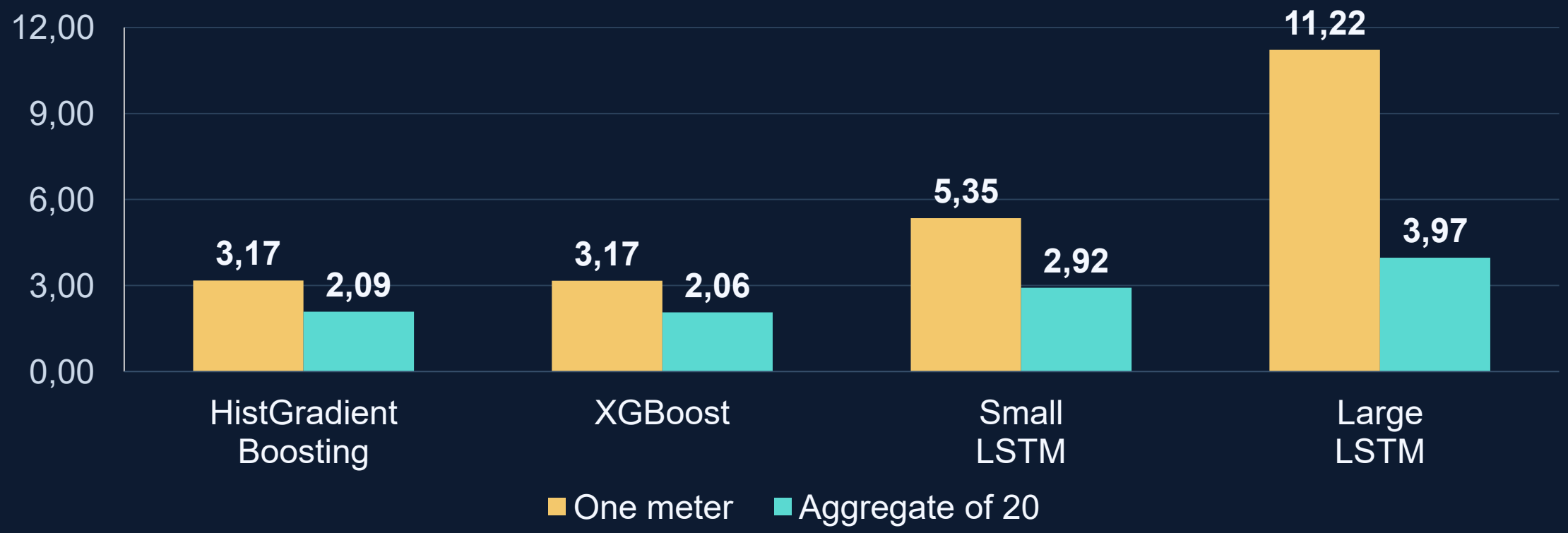
Comparison	Models	What it tells us
Clean baseline	HGB / XGBoost / LSTM	Ordinary forecast error
Robustness	LSTM / BiLSTM	Error after FGSM and PGD
Learning method	Standard / FGSM AT	Clean and attacked accuracy

MAPE in %. MAE and RMSE in kW. Lower errors are better.

Boosting robustness was not evaluated.

Accuracy on ordinary data

MAPE, %. Individual meter and aggregate, paper Table 5

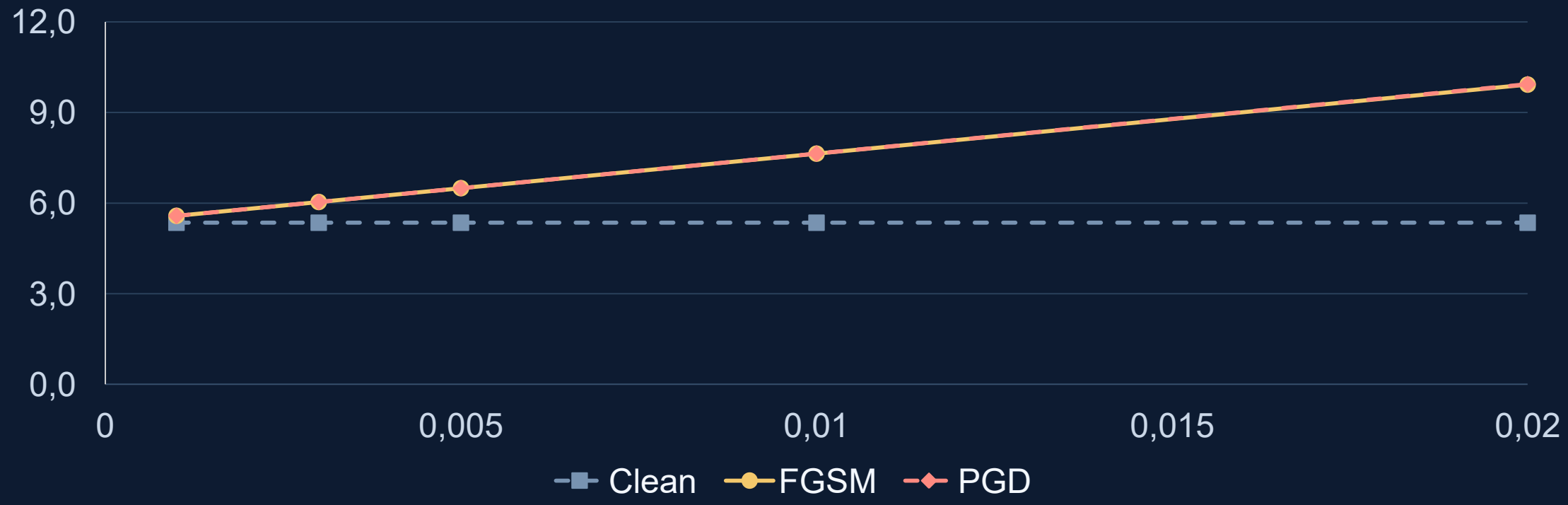


Boosting has low clean error, but its robustness was not evaluated

What happens after input changes

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

MAPE, %. Horizontal axis: attack budget ε

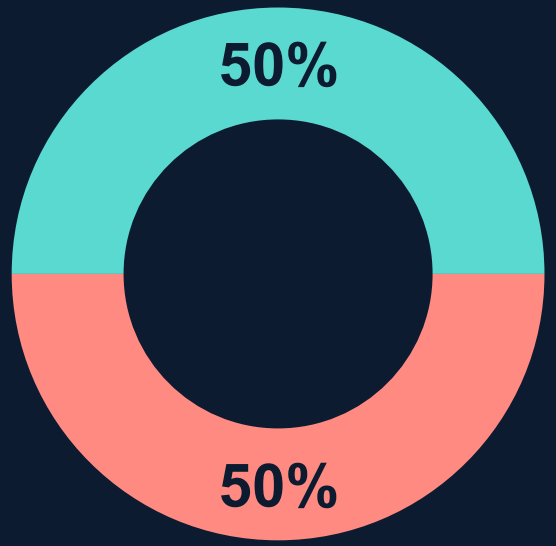


At ε = 0.02, PGD increases MAPE from 5.3508% to 9.9388%

Learning from difficult examples

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

Weights of the two loss terms



0.5 × clean MSE

0.5 × adversarial MSE

Generate an FGSM variant for every mini-batch

Early stopping uses clean validation data

Main suite: training ε = 0.003 / 0.005 / 0.01

Independent sweep: training ε from 0.001 to 0.005

Adversarial training in the main suite

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

MAPE, %. FGSM and PGD evaluation budget: ε = 0.02

Configuration	Clean	FGSM	PGD
LSTM AT-0.003	2.8376	7.0928	7.1024
Standard BiLSTM	3.4771	7.9266	7.9474
LSTM AT-0.010	3.9336	8.0137	8.0279
BiLSTM AT-0.010	4.0131	8.0082	8.0240
BiLSTM AT-0.005	3.5132	8.6589	8.6532
Standard LSTM	3.9731	8.7095	8.7626

LSTM AT-0.003 has the lowest MAPE in this suite

Why the baseline numbers differ

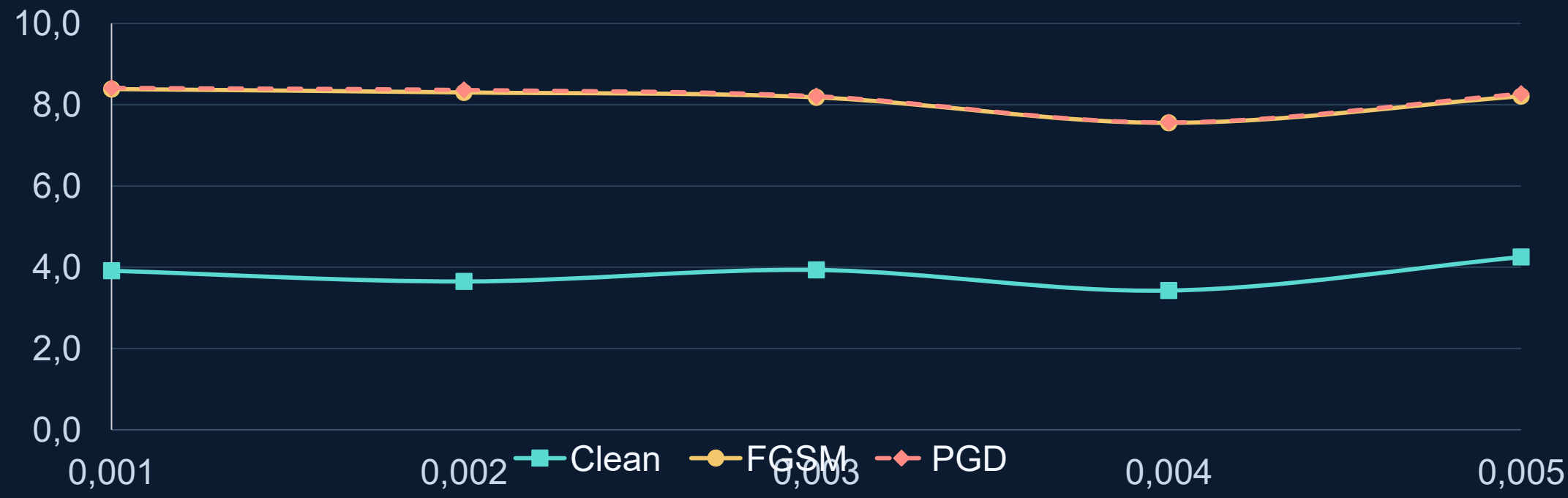
NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

Aggregate experiment	Clean MAPE, %
Small clean-baseline LSTM, Table 5	2.9215
Attack-evaluation run, Table 6	3.3207
Standard LSTM in robustness suite, Table 7	3.9731
AT-0.003 in main suite, Table 7	2.8376
AT-0.003 in independent sweep, Table 8a	3.9355

Comparisons must remain within their respective experiment

Training strength needs tuning

Independent sweep. MAPE, %. Evaluation budget: ε = 0.02



Training ε = 0.004: PGD MAPE 7.5664%, versus 8.2698% at 0.005

How far these results apply

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

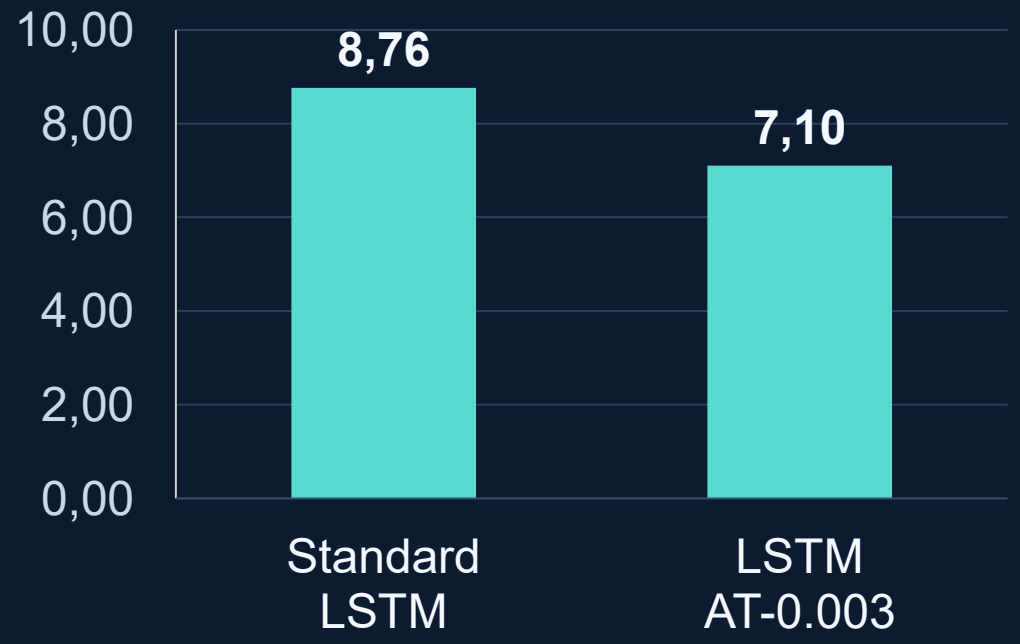
Limitation	Next validation
One dataset, 15-minute horizon	Other data and forecasting horizons
Full-series scaling, one run	Train-only scaling and repeated runs
True next load used for attacks	More realistic attacker information
PGD without a random start	Broader attack evaluation
No physical or dispatch validation	Feasibility and operational studies

Forecast errors measured. Outages and financial losses not measured.

Who can use this evidence

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

PGD MAPE, %, ε = 0.02, main suite



Forecasting developers

Test input sensitivity
and tune training strength

Teams preparing deployment

Validate data integrity
and local operating limits

Operational benefit still needs validation.

Thank you

NCFU
NORTH-CAUCASUS FEDERAL UNIVERSITY

Questions and discussion

Gleb Semenov

North-Caucasus Federal University

glebsemenov2003@gmail.com