Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics—Near-Real-Time Cyber-Physical Resiliency Through Machine Learning

Michael Ingram,[1] Anthony Florita,[2] Maurice Martin,[1] Kenny Gruchalla,[1] Xin Fang,[1] Mengmeng Cai,[1] Graham Johnson,[1] Nalinrat Guba,[1] Adarsh Hasandka,[1] Meghan Mooney,[1] Monte Lunacek,[1] Nicholas Gilroy,[1] Taylor Langan,[2] Scott Caruso,[3] Robert Cruickshank,[3] Bri-Mathias Hodge,[4] and Marija Marković[4]

1 National Renewable Energy Laboratory
2 Independent contractor
3 Cable Television Laboratories, Inc.
4 University of Colorado Boulder

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-5D00-85187
February 2023

Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics—Near-Real-Time Cyber-Physical Resiliency Through Machine Learning

Michael Ingram,[1] Anthony Florita,[2] Maurice Martin,[1] Kenny Gruchalla,[1] Xin Fang,[1] Mengmeng Cai,[1] Graham Johnson,[1] Nalinrat Guba,[1] Adarsh Hasandka,[1] Meghan Mooney,[1] Monte Lunacek,[1] Nicholas Gilroy,[1] Taylor Langan,[2] Scott Caruso,[3] Robert Cruickshank,[3] Bri-Mathias Hodge,[4] and Marija Marković[4]

1 National Renewable Energy Laboratory
2 Independent contractor
3 Cable Television Laboratories, Inc.
4 University of Colorado Boulder

Suggested Citation
Ingram, Michael, Anthony Florita, Maurice Martin, Kenny Gruchalla, Xin Fang, Mengmeng Cai, Graham Johnson, Nalinrat Guba, Adarsh Hasandka, Meghan Mooney, Monte Lunacek, Nicholas Gilroy, Taylor Langan, Scott Caruso, Robert Cruickshank, Bri-Mathias Hodge, and Marija Marković. 2023. *Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics—Near-Real-Time Cyber-Physical Resiliency Through Machine Learning*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5D00-85187. https://www.nrel.gov/docs/fy23osti/85187.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-5D00-85187
February 2023

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Technology Transitions, Technology Commercialization Fund (TCF) TCF-20-20213: Advanced Power Distribution Sensing and Communications through the Cable TV Broadband Network. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

CESER CEDS Award M619000162:

Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics—Near-Real-Time Cyber-Physical Resiliency Through Machine Learning

Final Report

Performance Period: Sept. 2019 to Jan. 2023

Michael Ingram (PI, NREL), **Anthony Florita** (Contractor), **Maurice Martin** (NREL),
Kenny Gruchalla (NREL), **Xin Fang** (NREL), **Mengmeng Cai** (NREL), **Graham Johnson** (NREL),
Nalinrat Guba (NREL), **Adarsh Hasandka** (NREL), **Meghan Mooney** (NREL),
Monte Lunacek (NREL), **Nicholas Gilroy** (NREL), **Taylor Langan** (Contractor),
Scott Caruso (CableLabs), **Robert Cruickshank** (CableLabs),
Bri-Mathias Hodge (University of Colorado), and **Marija Marković** (University of Colorado)

January 31, 2023

Executive Summary

The Situational Awareness of Grid Anomalies (SAGA) project built upon foundational power system tools developed at the National Renewable Energy Laboratory (NREL) integrated with an ever-increasing set of Gridmetrics[1] data extracted from the cable television (CATV) broadband network infrastructure while assimilating other time-series geospatial data and information, such as weather and cyber-physical phenomena, to demonstrate a disruptive technology for power system data analytics relying on existing infrastructure.

The objective of the project was to put into practice a graphical user interface for utility system operators along with a versatile bidirectional application programming interface (API) to allow field-validated visual analytics, machine learning, and human-in-the loop decision support to integrate cyber-physical data from CATV broadband power supplies and utility information systems. The result was an enhanced distribution grid visibility and operational situational awareness system that can be used by utilities to assist in detecting patterns of operation indicative of cyber incidents and other issues that have the potential to affect power availability and quality, distribution system resiliency, and electric service restoration.

The scope of the project was to create and demonstrate the viability of the envisioned SAGA prototype system with our technical review committee (TRC) member utilities from Holy Cross Energy, Fort Collins Light & Power, and Northern Lights, Inc. Specifically, the scope included streaming broadband sensor observations and metadata processed by CableLabs and NREL servers from an initial batch mode to analytics-based importing, analyzing, visualizing, and alerting, i.e., Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics: Near-Real-Time Cyber-Physical Resiliency Through Machine Learning.

Research thrusts and results
Three research thrusts supported the project: (1) visual analytics, (2) cyber-physical power system simulation, and (3) anomaly detection:

1. **Visual analytics:** Geospatial data visualization and alerting to support analyst-based operational decision making and standardization of a new algorithm-agnostic interface for machine learning.
   Results: The evolution of the initial SAGA/Gridmetrics API provided a foundation on which to define custom cable sensor groups and performance indices that could be made available with associated time-series data. The integration of this API and its sensor data were the focus and motivation for the development of the project's *Core Data Services*—three data services to access metadata, near-real-time data streams, and historical batches—enabling the data and visualization team to deliver a scalable and cohesive data pipeline that directly supports novel grid visualizations and future machine learning workflows. Additionally, these *Core Data Services* are containerized to provide

[1] Gridmetrics is a commercial subsidiary of Cable Television Laboratories, Inc. (CableLabs), the research and development association of the cable broadband industry. CableLabs provided SAGA data and research and development collaboration.

flexibility with deployment environments and scalable interfaces to Gridmetrics sensor data.
To subsequently ingest, explore, and analyze these rich data, the team designed and integrated interactive visualization components with the *Core Data Services*. To this end, three advanced visual analytic components were prototyped to provide user support for comprehensive system overviews, geospatial exploration, and advanced time-series analysis capabilities. Further details are included in Section 2, Appendix A1, and Appendix A2.

2. **Cyber-physical power system simulation:** Cyber-physical power system modeling and simulation for a better understanding of grid impacts caused by prototype cyber-physical events.
   Results: Both steady-state and dynamic cyber-physical power system simulation models were developed leveraging open-source software and real-world distribution feeder data. Cyber-physical events were designed and implemented considering a forward-looking system operating setting in which distributed energy resources (DERs) are aggregated to provide grid services. Further details are included in Section 3.

3. **Anomaly detection:** Enabling Gridmetrics sensor data to provide improved cyber situational awareness through machine learning.
   Results: The Gridmetrics sensor data-enabled anomaly detection algorithm is developed and tested with simulation data collected from the cyber-physical power system simulations to demonstrate how the Gridmetrics sensor data can be used to automate cyber anomaly detection in real-time operation. Both deterministic and probabilistic prediction-based anomaly detection models are developed and compared. Further details are included in Section 4.

Work was performed over 3 years plus a 4-month documentation and reporting period. During the project, we convened and met with our valued TRC member utility advisors on four occasions. The focus of Phase 1 (Year 1) was on SAGA alpha testing, in which we built our first geospatial visualization platform, developed models of electric distribution grids, and defined the cybersecurity issues of greatest importance to cyber-physical resiliency. The focus of Phase 2 (Year 2) was on beta testing, in which we evolved the alpha version capabilities to explore cybersecurity anomaly detection built off a robust analytical backend. We also extended the geospatial visualization platform to observe and contextualize anomalies, creating a visual analytics framework to provide a new state of the art in situational awareness across geographically disparate areas at high spatial fidelity. The focus of Phase 3 (Year 3) was on demonstrations of cyber event simulations and exploring new indices for arbitrary collections of sensors and time spans.

Learnings and improvements identified and completed in parallel with the SAGA project
After starting the SAGA project, the team identified increased grid cybersecurity benefits that could be realized by improving sensing beyond the capabilities of the millions of broadband power quality sensors already in service worldwide. We learned that improved sensing would help promote the rapid widespread commercial availability of secondary distribution grid voltage and phase angle data by providing higher fidelity grid sensing. As a SAGA-enhancing activity, the team proposed and completed a technology commercialization project to advance the state of

the art in sensing of the grid. Funding was provided by the U.S. Department of Energy (DOE) Office of Technology Transitions, Technology Commercialization Fund (TCF) for TCF-20-20213: Advanced Power Distribution Sensing and Communications Through the Cable TV Broadband Network. The TCF project was designed to fill the need for better standards-based grid sensors to support SAGA. In the TCF project, we defined a grid power quality sensor with 0.2% (p.u.) precision and 10-kHz sampling, created the ANSI/SCTE 271 grid sensing American National Standard, built dozens of prototype sensors, and put them in the field—all within the timeline of the SAGA project. The TCF final report includes a copy of the new ANSI standard, the patent we filed that was required for our application for DOE TCF funding, and a letter of support for continued commercialization efforts from SAGA utility partner Holy Cross Energy; please see https://www.nrel.gov/docs/fy22osti/83624.pdf.

Important next steps to take advantage of the value created

In September 2022, the team responded to a request from the DOE Office of Cybersecurity, Energy Security, and Emergency Response (CESER) with a draft SAGA Ops proposal to build upon the SAGA work to provide operational benefits. The main objective of SAGA Ops is to provide CESER with the ability to take advantage of the increasing set out-of-band cable sensor data from Gridmetrics on a nationwide scale to directly support CESER's responsibilities regarding grid security and resiliency. At a high level, SAGA Ops provides two categories of support: (1) real-time monitoring of sensor data and (2) reporting of sensor data over defined periods of time. The abridged SAGA Ops proposal is included in Appendix B.

Conclusions

The SAGA project created technology that leverages, couples, and fortifies two vastly different realms—power and broadband—to increase the resiliency of the power grid in the face of increasing cyberattacks and operational challenges related to integrating DERs. Our exploration of potential synergies of broadband-enabled grids resulted in identifying a mutually beneficial symbiosis that can increase the resiliency of both power and broadband services. Broadband networks perform better with reliable power and are good at providing real-time measurements that identify where the grid is under attack, is failing, or is weak. Likewise, sensor-starved distribution grids perform better and can be more reliable when their operation is buttressed with observations of broadband-detected anomalies.

Further, although new communication technologies can be rate-payer financed and deployed over years and decades to assist in grid operations, existing in-service broadband networks are unique in that they already pass within 1,000 feet of 97% of homes in North America. Moreover, broadband's gigabit speeds and millisecond latencies create new grid observability and control paradigms that cannot be accomplished with existing networks that provide lower speeds and higher latencies. Broadband-enabled, grid situational awareness and control create entirely new possibilities for actively managing the grid to prevent cascading outages.

In terms of saving lives and preventing human suffering, broadband's contribution to improving grid resiliency, reliability, and cost-effective operation should be further developed as quickly as possible. In addition to supporting the objectives of CESER, SAGA follow-on work can have a positive impact on many objectives across various DOE offices.

Table of Contents

Executive Summary iii

1 Task and Milestone Summary 1

2 Core Data Services and Visual Analytics 2

2.1 Metadata Service 3

2.2 Historical Service 5

2.3 Streaming Service 7

2.4 Visual Analytics 8

2.5 Learnings, Conclusions, Recommendations 17

3 Cyber-Physical Power System Simulation 18

3.1 Dynamic Simulation With DER-Enabled Frequency Regulation 18

3.2 Steady-State Simulation with DER-Enabled Voltage Regulation 19

3.3 Cyber Anomaly Scenarios 22

3.4 Simulation Results 23

3.4.1 Synchronous Versus Asynchronous Communication Latency 24

3.4.2 Measurement Channel Versus Control Channel Events 25

3.4.3 False Data Injection Attack 26

3.5 Learnings, Conclusions, Recommendations 29

4 Anomaly Detection 30

4.1 Distributed Anomaly Detection Framework 30

4.2 Deterministic Versus Probabilistic Approaches 31

4.3 Model Fitting Procedure 32

4.4 Performance Evaluation 33

4.4.1 Performance of the Deterministic Approach 33

4.4.2 Performance of the Probabilistic Approach 35

4.4.3 Performance Comparison 36

4.5 Learnings, Conclusions, Recommendations 37

References to SAGA-Related Papers 38

Appendix A1: Using the SAGA APIv3 39

Appendix A2: SAGA APIv3 40

Appendix B: Abridged SAGA Ops Proposal 52

Appendix C: Letters From Utilities Supporting Continued SAGA Development 68

List of Figures

Figure 2.1. Overview of SAGA's core data services architecture 2
Figure 2.2. Sample *Metadata Service* API documentation available from the containerized service 4
Figure 2.3. Sample *Historical Service* API documentation available from the containerized service 6
Figure 2.4. Sample *Historical Service* API documentation available from the containerized service 7
Figure 2.5. Visual analytics geospatial component in the initial view 9
Figure 2.6. Visual analytics geospatial component with the data layer management menu selected 10
Figure 2.7. Visual analytics geospatial component focused on the Houston, Texas, metropolitan area by selecting a "fly-to" option in the location menu located in the top left corner 11
Figure 2.8. Visual analytics geospatial component focused on Glenwood Springs, Colorado, by selecting a "fly-to" option in the location menu located in the top left corner 12
Figure 2.9. Visual analytics geospatial component focused on Fort Collins, Colorado, by using the location menu in the top left corner 13
Figure 2.10. Visual analytics geospatial component focused on Fort Collins, Colorado 14
Figure 2.11. Visual analytics geospatial component focused on Fort Collins, Colorado. 15
Figure 2.12. Sample of a visual analytics time-series component displaying a selected focus area's sensor voltage distribution on the left and a selected sensor's voltage time series on the right 15
Figure 2.13. Sample visual analytics time-series component for interactively exploring sensor voltage patterns across long periods 16
Figure 3.1. Framework of the dynamic cyber-physical simulation tool 19
Figure 3.2. Secondary frequency regulation control diagram 19
Figure 3.3. Framework of the steady-state cyber-physical simulation tool 20
Figure 3.4. Closed-loop control diagram of the frequency regulation 22
Figure 3.5. Closed-loop control diagram of the voltage regulation 22
Figure 3.6. Frequency trajectories under different levels of synchronous communication latencies 24
Figure 3.7. Frequency trajectories under different levels of asynchronous communication latencies 25
Figure 3.8. Frequency trajectories under different levels of packet dropouts occurring at the control channels 25
Figure 3.9. Frequency trajectories under different levels of packet dropouts occurring at the measurement channels 26
Figure 3.10. Gridmetrics voltage measurements under normal operation 27
Figure 3.11. Gridmetrics voltage measurements under a control channel data deception event 27
Figure 3.12. Gridmetrics voltage measurements under a measurement channel data deception event 28
Figure 3.13. Daily system load and PV profiles 28
Figure 4.1. Illustration of the distributed anomaly detection framework 30
Figure 4.2. Structure of the prediction model 31
Figure 4.3. Illustration of the pinball loss when $\boldsymbol{\tau} > 0.5$ 32
Figure 4.4. Distribution of R-square values across 148 deterministic prediction models 34
Figure 4.5. Anomaly detection results of the deterministic approach when the threshold equals 0.15 34
Figure 4.6. Anomaly detection performance of the deterministic approach varies against the threshold value 35
Figure 4.7. Output generated from a signal probabilistic prediction model 35
Figure 4.8. Result of the probabilistic anomaly detection 36
Figure 4.9. Histogram of the observing frequencies across 148 probabilistic prediction models 36
Figure 4.10. Anomaly detection performance of the deterministic approach under different numbers of attack targets 37
Figure 4.11. Anomaly detection performance of the probabilistic approach under different numbers of attack targets 37

List of Tables

Table 1.1. SAGA Planned and Actual Milestone Activities 1
Table 4.1. Difference Between the Deterministic and Probabilistic Anomaly Detection Approaches....... 32

1 Task and Milestone Summary

Table 1.1 presents a summary of the Situational Awareness of Grid Anomalies (SAGA) project planned versus actual milestone accomplishments based on the U.S. Department of Energy (DOE) Office of Cybersecurity, Energy Security, and Emergency Response (CESER) Cybersecurity for Energy Delivery Systems (CEDS) Award M619000162 fieldwork proposal.

Table 1.1. SAGA Planned and Actual Milestone Activities

Milestone	Planned Completion and Activity	Actual Completion and Activity
Milestone Q2-2020	(Month 6) TRC feedback on data sources and SAGA features	Due to COVID-19, delayed until 11/23/21 TRC project review of: • Years 1 and 2 accomplishments • Next-gen sensors TCF project • Year 3 plan. TRC feedback: "Good use of AI to figure what's going on: tree, squirrel, recloser, sensor drift, etc."
Milestone Q3-2020	(Month 12) SAGA alpha version delivered and go/no-go decision	11/4/20: Alpha wireframes
Milestone Q2-2021	(Month 18) TRC feedback on empirical demonstration and SAGA features	Due to COVID-19, delayed until 2/17/22 TRC project review: • Year 3 plans and progress. TRC Feedback 1: Consider incident response; what do we do if utility is attacked/compromised? TRC Feedback 2: Support next phase: • Simul vs. actual • API for OMS • Restoration priority.
Milestone Q4-2021	(Month 24) SAGA beta version delivered and go/no-go decision	2Q21: Beta working system. See https://app.box.com/s/2qh91fxihoy9d1pwnzit2adiyuqn1ou6.
Milestone Q2-2022	(Month 30) TRC feedback on emulation and SAGA features	5/19/22: TRC project review: • Emulation • Simulation • Anomaly detection.
Milestone Q4-2022	(Month 36) Deliver SAGA software for visual analytics: final software, technical report, and presentation	1/31/23
New unplanned milestone	(Month 39) TRC feedback on: • Final reports for SAGA project and TCF project • SAGA Ops proposal.	12/15/22 TRC multi-project review: • Next-gen sensor results and use cases • Recommendations for further research and commercialization.

2 Core Data Services and Visual Analytics

With the initial development of the Gridmetrics application programming interface (API) for accessing sensor-level metadata and time-series data, a fundamental goal of the project was to design, develop, and prototype a scalable data systems architecture that enables continued data set expansion, high availability and performance, advanced interactive visualization, and real-time machine learning workflows. To this end, the data and visualization team designed SAGA's *Core Data Services.* This initial core consists of three composable data services—*Metadata Service, Historical Service,* and *Streaming Service*—with each service providing the appropriate functionality for exploring, monitoring, analyzing, and providing data made available by Gridmetrics sensors. The following sections provide an overview of the prototyped *Core Data Services* and each service's current context. A visual representation is also shown in Figure 2.1.

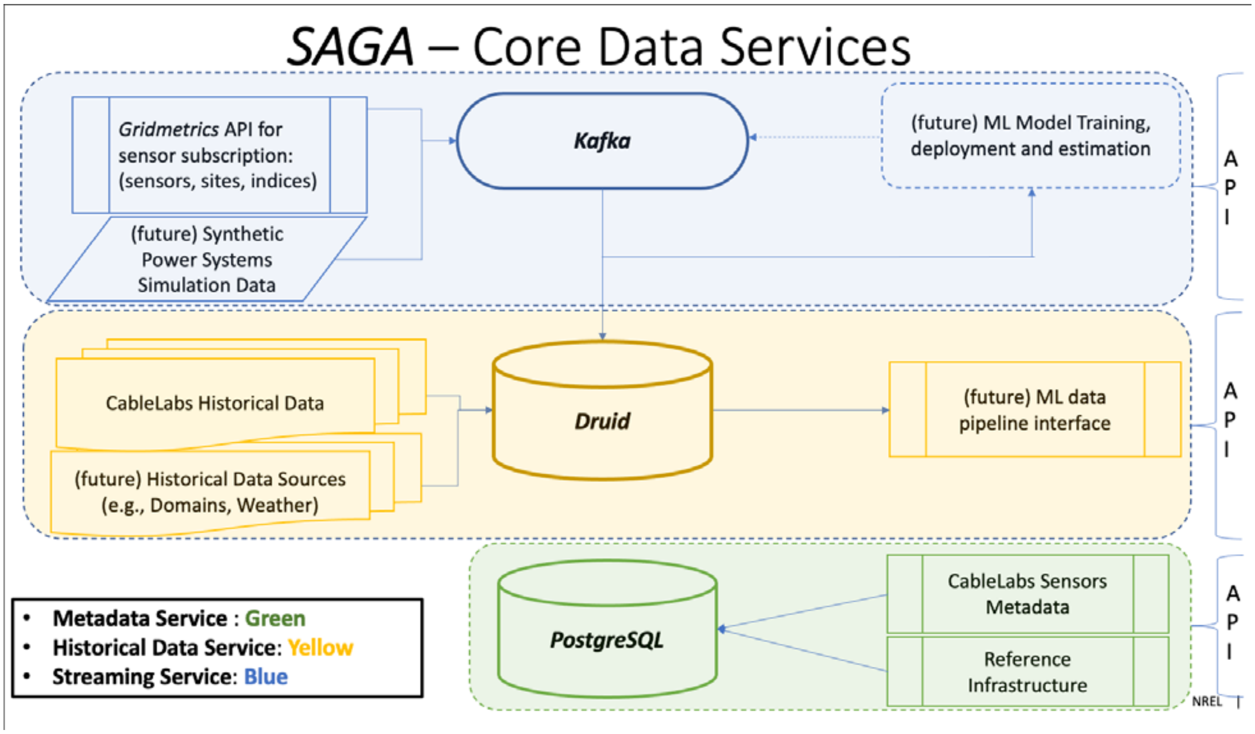
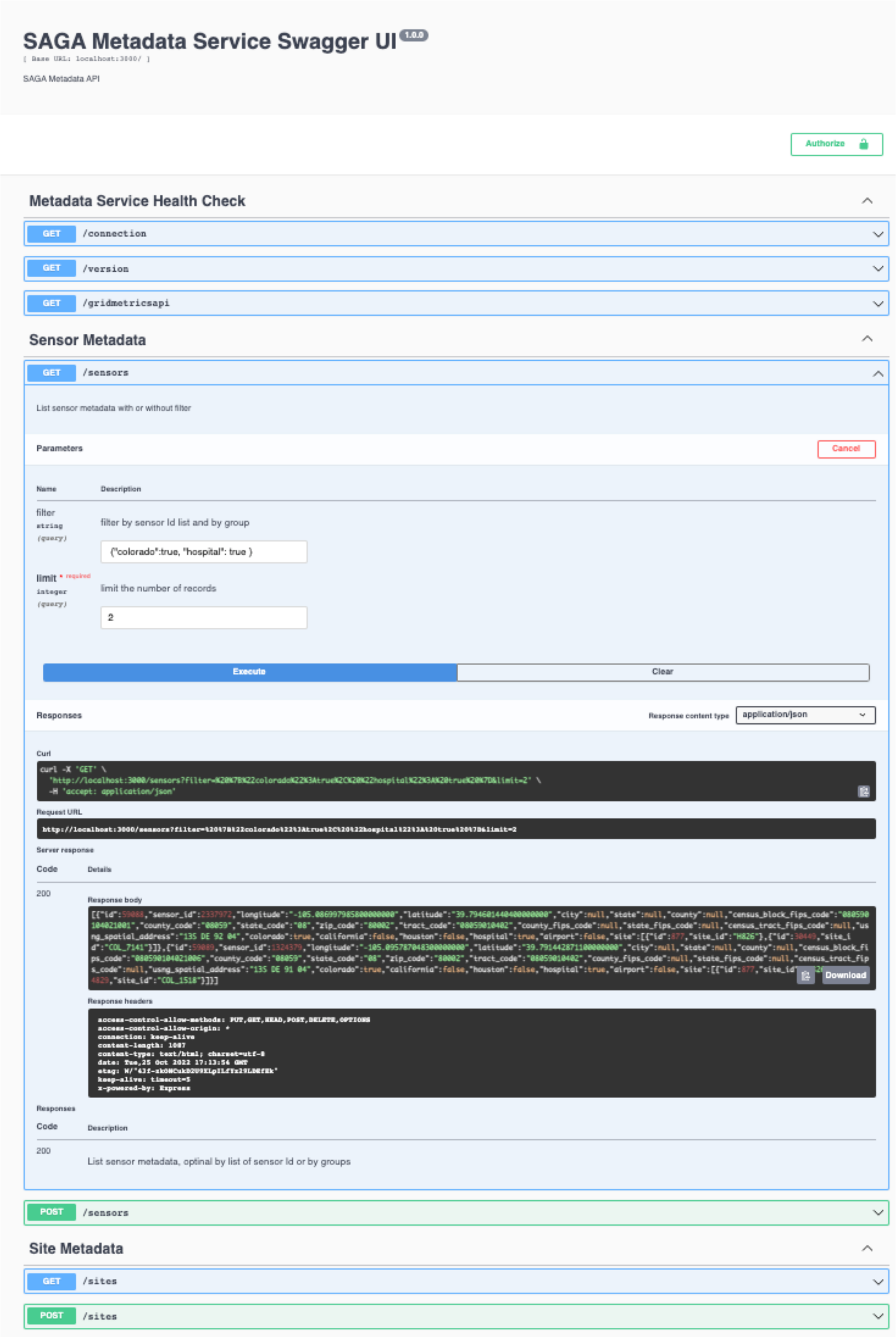

SAGA – Core Data Services
Gridmetrics API for sensor subscription: (sensors, sites, indices)
Kafka
(future) ML Model Training, deployment and estimation
(future) Synthetic Power Systems Simulation Data
API
CableLabs Historical Data
Druid
(future) ML data pipeline interface
(future) Historical Data Sources (e.g., Domains, Weather)
API
Metadata Service : Green
Historical Data Service: Yellow
Streaming Service: Blue
PostgreSQL
CableLabs Sensors Metadata
Reference Infrastructure
API
NREL

Figure 2.1. Overview of SAGA's core data services architecture

Three services were designed and prototyped to provide scalable interfaces to access, manage, and analyze Gridmetrics sensor data. The *Metadata Service* (green in Figure 2.1) handles the fetching, composing, and providing of static sensor properties, such as identification (ID), location, and group inclusion. This sits on top of a PostgreSQL database instance housed at the National Renewable Energy Laboratory (NREL). The *Historical Service* (yellow) handles the storage, access, and analysis of batches of Gridmetrics sensor data, e.g., intervals of raw data, outage counts, average voltages, and group statistics. This service is built using an Apache Druid instance housed at NREL. The *Streaming Service* (blue) handles the subscription and continuous ingestion of near-real-time Gridmetrics sensor data via the defined API specification and available routes provided in Appendix A2. The collected streaming data are directly piped into the Druid data store for access via the *Historical Service*. Future extensions of the services include hooks for online machine learning training and inference. Each of the three prototyped services is available as a Docker container to provide flexible deployment options.

2.1 Metadata Service

The primary function of the *Metadata Service* is to provide access to subscribed sensor metadata, such as its geospatial location, group inclusions, and other static properties. The service's capabilities and routes are documented via an interactive Swagger interface within its Docker container. A sample of the documentation is provided in Figure 2.2.



SAGA Metadata Service Swagger UI 1.0.0

[Base URL: localhost:3000/]

SAGA Metadata API

Metadata Service Health Check

- GET /connection
- GET /version
- GET /gridmetricsapi

Sensor Metadata

GET /sensors

List sensor metadata with or without filter

Parameters

Name	Description
filter string (query)	filter by sensor Id list and by group {"colorado":true, "hospital": true }
limit * required integer (query)	limit the number of records 2

Responses — Response content type: application/json

Curl

```
curl -X 'GET' \
  'http://localhost:3000/sensors?filter=%20%7B%22colorado%22%3Atrue%2C%20%22hospital%22%3A%20true%20%7D&limit=2' \
  -H 'accept: application/json'
```

Request URL

```
http://localhost:3000/sensors?filter=%20%7B%22colorado%22%3Atrue%2C%20%22hospital%22%3A%20true%20%7D&limit=2
```

Server response

Code: 200

Details — Response body

```
[{"id":59088,"sensor_id":2337972,"longitude":"-105.086997985800000000","latitude":"39.794601440400000000","city":null,"state":null,"county":null,"census_block_fips_code":"080590
104021001","county_code":"08059","state_code":"08","zip_code":"80002","tract_code":"08059010402","county_fips_code":null,"state_fips_code":null,"census_tract_fips_code":null,"us
ng_spatial_address":"13S DE 92 04","colorado":true,"california":false,"houston":false,"hospital":true,"airport":false,"site":[{"id":877,"site_id":"H826"},{"id":30449,"site_i
d":"COL_7141"}]},{"id":59089,"sensor_id":1324379,"longitude":"-105.095787048300000000","latitude":"39.791442871100000000","city":null,"state":null,"county":null,"census_block_fi
ps_code":"080590104021006","county_code":"08059","state_code":"08","zip_code":"80002","tract_code":"08059010402","county_fips_code":null,"state_fips_code":null,"census_tract_fip
s_code":null,"usng_spatial_address":"13S DE 91 04","colorado":true,"california":false,"houston":false,"hospital":true,"airport":false,"site":[{"id":877,"site_id"
4829,"site_id":"COL_1518"}]}]
```

Response headers

```
access-control-allow-methods: PUT,GET,HEAD,POST,DELETE,OPTIONS
access-control-allow-origin: *
connection: keep-alive
content-length: 1087
content-type: text/html; charset=utf-8
date: Tue,25 Oct 2022 17:13:54 GMT
etag: W/"43f-zk0HCukB2U9XLpILfYz29LB8fEk"
keep-alive: timeout=5
x-powered-by: Express
```

Responses

Code	Description
200	List sensor metadata, optinal by list of sensor Id or by groups

- POST /sensors

Site Metadata

- GET /sites
- POST /sites

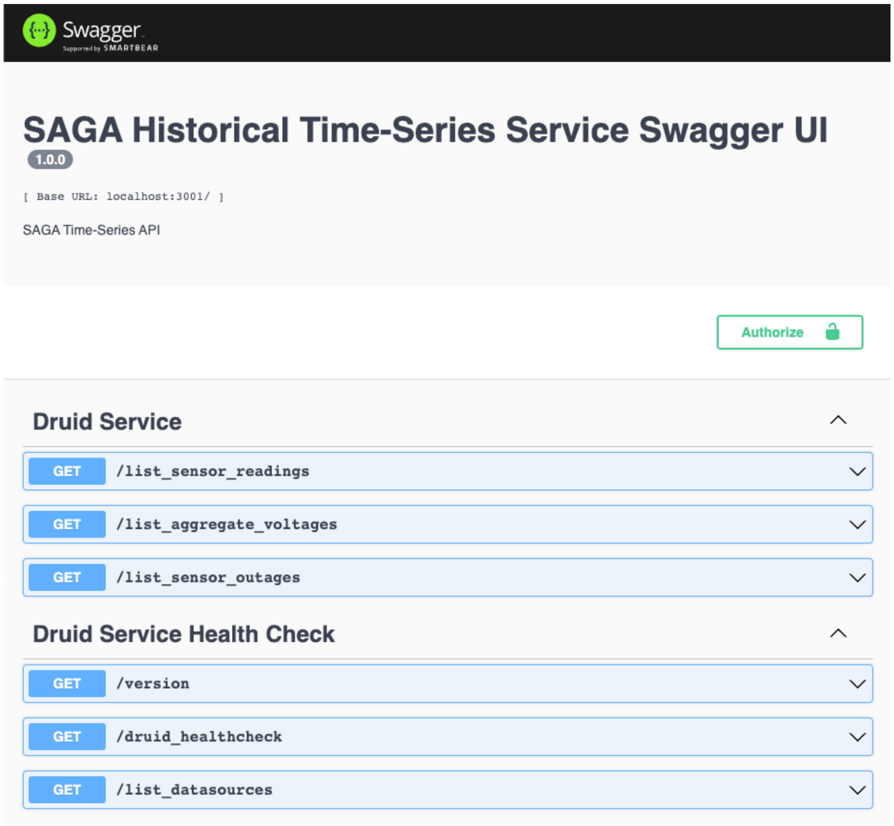
Figure 2.2. Sample *Metadata Service* API documentation available from the containerized service

Displays all available routes for querying along with sample interfaces to check live API statuses, route parameters, and sample data response schemas. Currently, the main routes provide access to all subscribed sensors (via the /sensors route) as well as to all Gridmetrics sensor sites (i.e., locations of groups of sensors, via the /sites route). Service API documentation is available within in each service's running container and is similar for both the *Historical Service* and the *Streaming Service*.

Additionally, the *Metadata Service* is intended to be used to populate the visual analytics geospatial component—providing users the ability to explore available sensor asset details and associated groups of interest. The *Metadata Service* and the *Historical Service* work in conjunction with one another as sensors are selected within the visual analytics geospatial component. The selected sensors provide the unique sensor IDs from the *Metadata Service* as a query parameter for requests made to the *Historical Service* for sensor data over a specified date-time range.

2.2 Historical Service

The primary function of the *Historical Service* is to handle requests for Gridmetrics sensor data over specified date-time ranges. Similar to the *Metadata Service*, documentation of routes and functionality is provided via an interactive Swagger interface within its Docker container. Currently, there are three main routes that return various transforms of Gridmetrics sensor data: raw sensor readings, aggregate/statistical voltages, and sensor outage intervals. Each route accepts parameters with sensor IDs, a date-time range, and a response size limit. The aggregate/statistical voltage route accepts an additional parameter for the desired statistic (e.g., mean, standard deviation) and a subinterval, e.g., to provide the average of a sensor's daily voltage over that last month. An overview of the Swagger documentation is shown in Figure 2.3, with sample fields from the sensor outage route provided in Figure 2.4.



SAGA Historical Time-Series Service Swagger UI

1.0.0

[Base URL: localhost:3001/]

SAGA Time-Series API

Druid Service

GET /list_sensor_readings

GET /list_aggregate_voltages

GET /list_sensor_outages

Druid Service Health Check

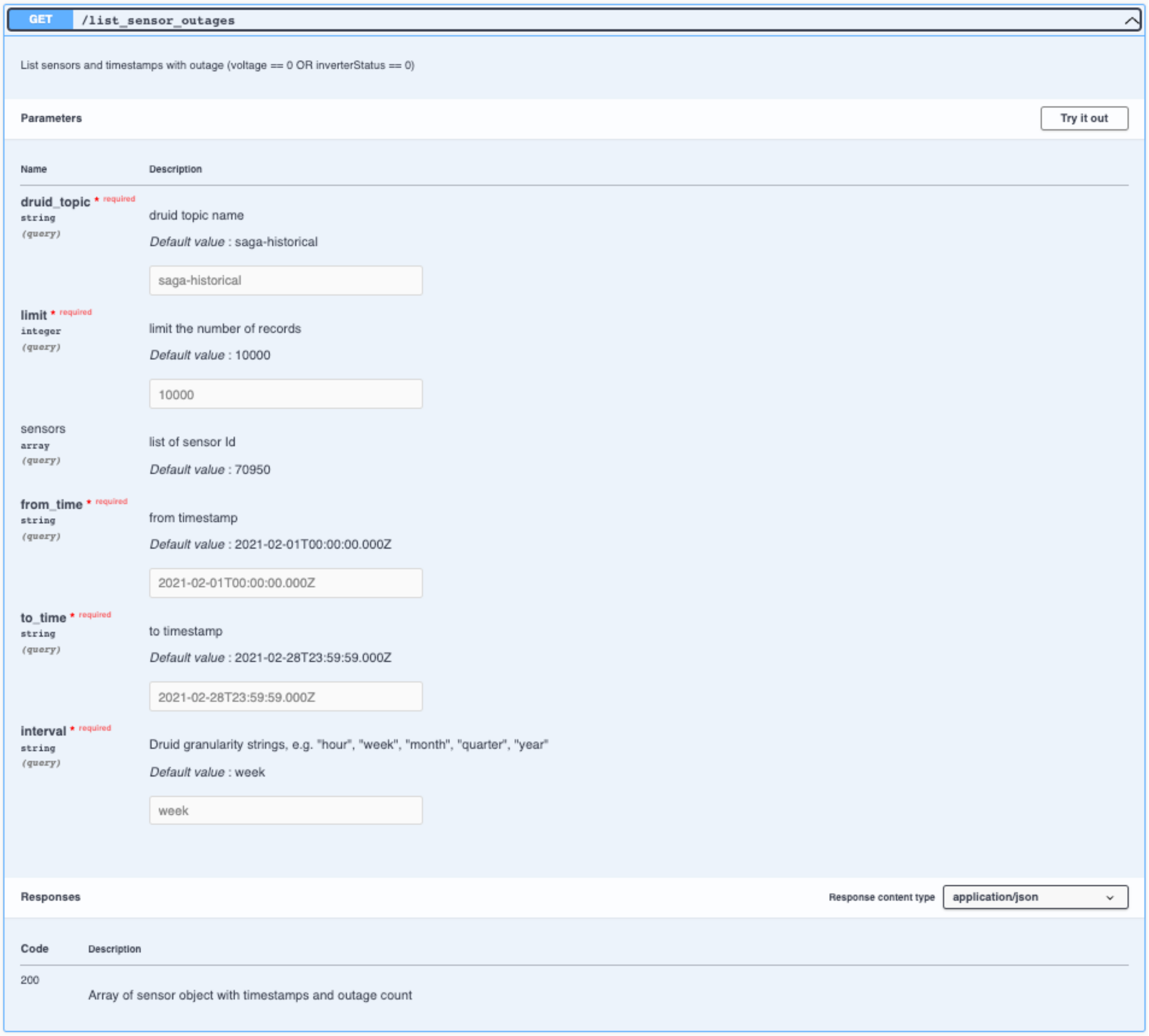
GET /version

GET /druid_healthcheck

GET /list_datasources

Figure 2.3. Sample *Historical Service* API documentation available from the containerized service

Displays include all available routes for querying along with sample interfaces to check live API statuses, route parameters, and sample data response schemas. The main data routes are the first three listed; they return raw, statistical, and outage data, respectively, over defined intervals. A sample of the query parameters is shown in Figure 2.4



GET /list_sensor_outages

List sensors and timestamps with outage (voltage == 0 OR inverterStatus == 0)

Parameters

Name	Description
druid_topic * required string (query)	druid topic name *Default value* : saga-historical saga-historical
limit * required integer (query)	limit the number of records *Default value* : 10000 10000
sensors array (query)	list of sensor Id *Default value* : 70950
from_time * required string (query)	from timestamp *Default value* : 2021-02-01T00:00:00.000Z 2021-02-01T00:00:00.000Z
to_time * required string (query)	to timestamp *Default value* : 2021-02-28T23:59:59.000Z 2021-02-28T23:59:59.000Z
interval * required string (query)	Druid granularity strings, e.g. "hour", "week", "month", "quarter", "year" *Default value* : week week

Responses

Response content type: application/json

Code	Description
200	Array of sensor object with timestamps and outage count

Figure 2.4. Sample *Historical Service* API documentation available from the containerized service

This view shows the available query parameters for the route returning outage intervals over the defined date-time range. This interactive documentation enables users to test example queries and explore response schemas before integrating *Core Data Services* calls into applications.

The responses from the *Historical Service's* routes are used by the visual analytics components to generate various time-series representations for system behaviors, supporting the visualization of system distributions, raw sensor streams, rolling averages, outage periods, or system pattern exploration. The data currently ingested in the *Historical Service* comprise both batches of historical periods as well as current 5-minute intervals that are piped from the *Streaming Service*.

2.3 Streaming Service

The primary function of the *Streaming Service* is to continuously ingest available Gridmetrics sensor data and pass them into the *Historical Service's* Apache Druid data store. This was accomplished using an NREL Kafka instance with associated producers and consumers, written in Python, that long-poll the developed Gridmetrics API every 5 minutes to place all responses into NREL's Druid instance using Apache StreamSets. Although this prototype architecture supports the current minute-level data rates, updates to Gridmetrics data interfaces using direct Kafka connections to NREL's *Core Data Services* will support future efforts with next-generation sensor streams that provide data at higher orders of magnitude (e.g., sensor streams at 10 kHz).

2.4 Visual Analytics

The development of the *Core Data Services* directly supported the team's visual analytics efforts by providing capabilities for users to interactively fetch associated sensor metadata and time series to analyze system behaviors. To this end, the team designed and prototyped an interactive web-based visual analytics platform that directly connects to the *Core Data Services* to fetch/populate sensor locations and query for time-series values. In its current form, the visual analytics platform enables users to interactively visualize geospatial distributions of Gridmetrics sensor locations and their proximity to various infrastructure assets, select sensors/regions of interest, and subsequently query for real-time or historical batches of sensor data (i.e., voltage, inverter status) at up to 1-minute granularity. Overviews of the visual analytics capabilities, various views, components, and areas for extension are shown in the following figures.

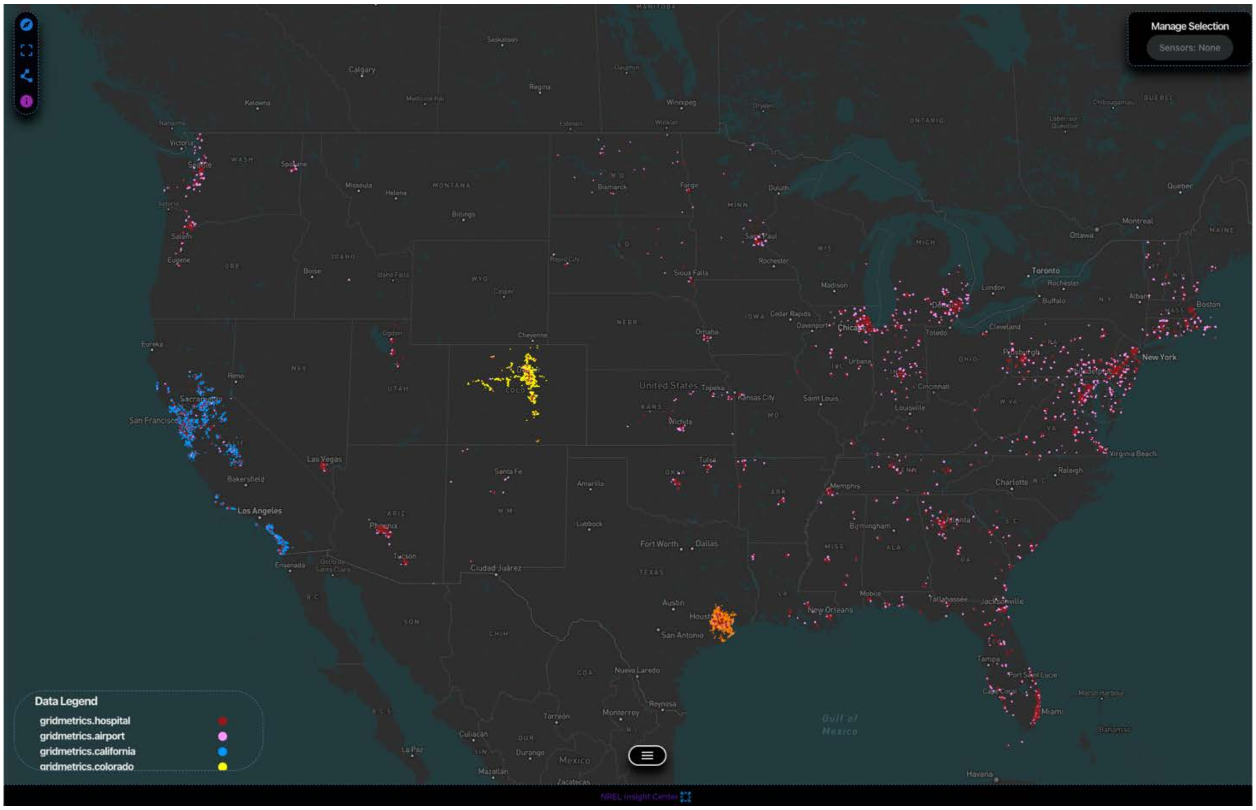
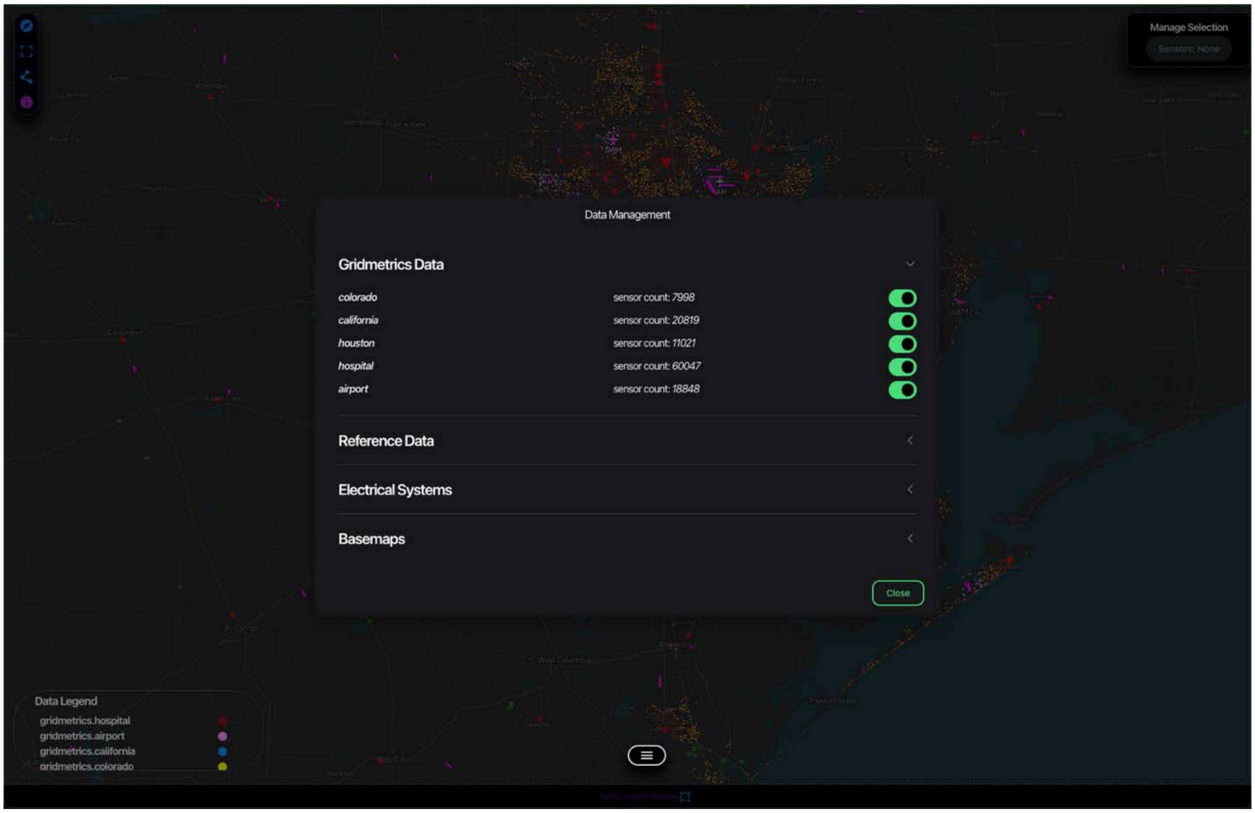


Figure 2.5. Visual analytics geospatial component in the initial view

This component allows users to interactively explore geospatial distributions of Gridmetrics sensors and reference infrastructure. The different Gridmetrics sensor groups are shown and colored by their group membership. For example, red: hospital group, purple: airport group, yellow: Colorado group, orange: Houston group, and blue: California group, with some sensors belonging to multiple groups and subsequently styled by the regional group. A data layer management menu is accessible from the bottom center, a navigation/“fly-to” location and selection-tool menu is provided in the top left, a manage selection menu/table is provided in the top right, and an active data layer legend is viewable in the bottom left.



Data Management

Gridmetrics Data

colorado	sensor count: 7998	●
california	sensor count: 20819	●
houston	sensor count: 11021	●
hospital	sensor count: 60047	●
airport	sensor count: 18848	●

Reference Data

Electrical Systems

Basemaps

Data Legend

gridmetrics.hospital	●
gridmetrics.airport	●
gridmetrics.california	●
gridmetrics.colorado	●

Figure 2.6. Visual analytics geospatial component with the data layer management menu selected

This menu allows users to toggle different data groups, with the sensor counts within each group shown, in addition to layering available reference data from HIFLD[2] (e.g., hospitals, airports, power plants), as well as different base map options. In the future, distribution and transmission electric systems can be added and subsequently viewed.

[2] HIFLD (Homeland Infrastructure Foundation-Level Data) Open Data provides national foundation-level geospatial data within the open public domain that can be useful to support community preparedness, resiliency, research, and more. Further details are available at https://hifld-geoplatform.opendata.arcgis.com/.

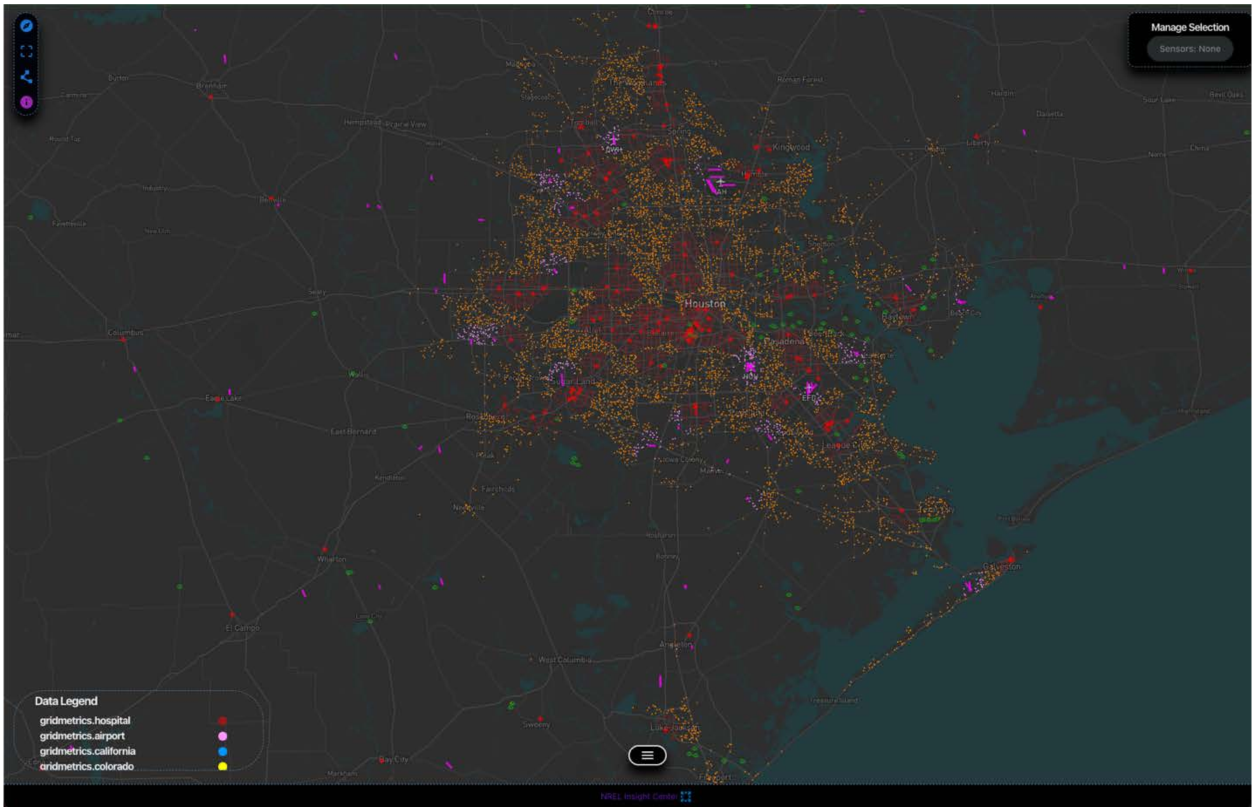


Figure 2.7. Visual analytics geospatial component focused on the Houston, Texas, metropolitan area by selecting a “fly-to” option in the location menu located in the top left corner

This current view provides an overview of the distributions of more than 11,000 Gridmetrics sensors in the major metropolitan area, in addition to HIFLD data for hospitals (red cross icon), airports (purple runway lines), and power plants (green power icons), with red points corresponding to hospital groups, purple points corresponding to airport groups, and orange points corresponding to the greater Houston metropolitan group.

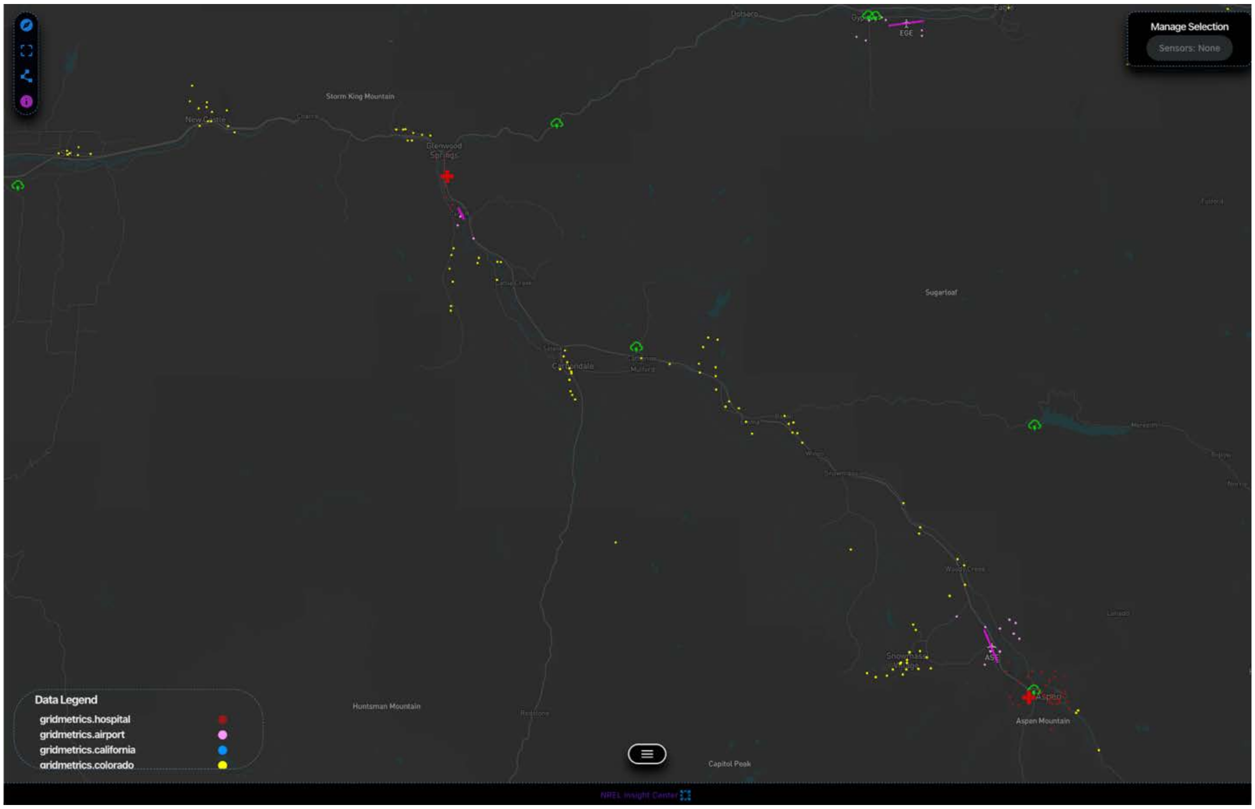


Figure 2.8. Visual analytics geospatial component focused on Glenwood Springs, Colorado, by selecting a "fly-to" option in the location menu located in the top left corner

This current view provides an overview of the distributions of approximately 200 Gridmetrics sensors in the rural area, in addition to HIFLD data for hospitals (red cross icon), airports (purple runway lines), and power plants (green power icons), with red points corresponding to hospital groups, purple points corresponding to airport groups, and yellow points corresponding to the Colorado group.

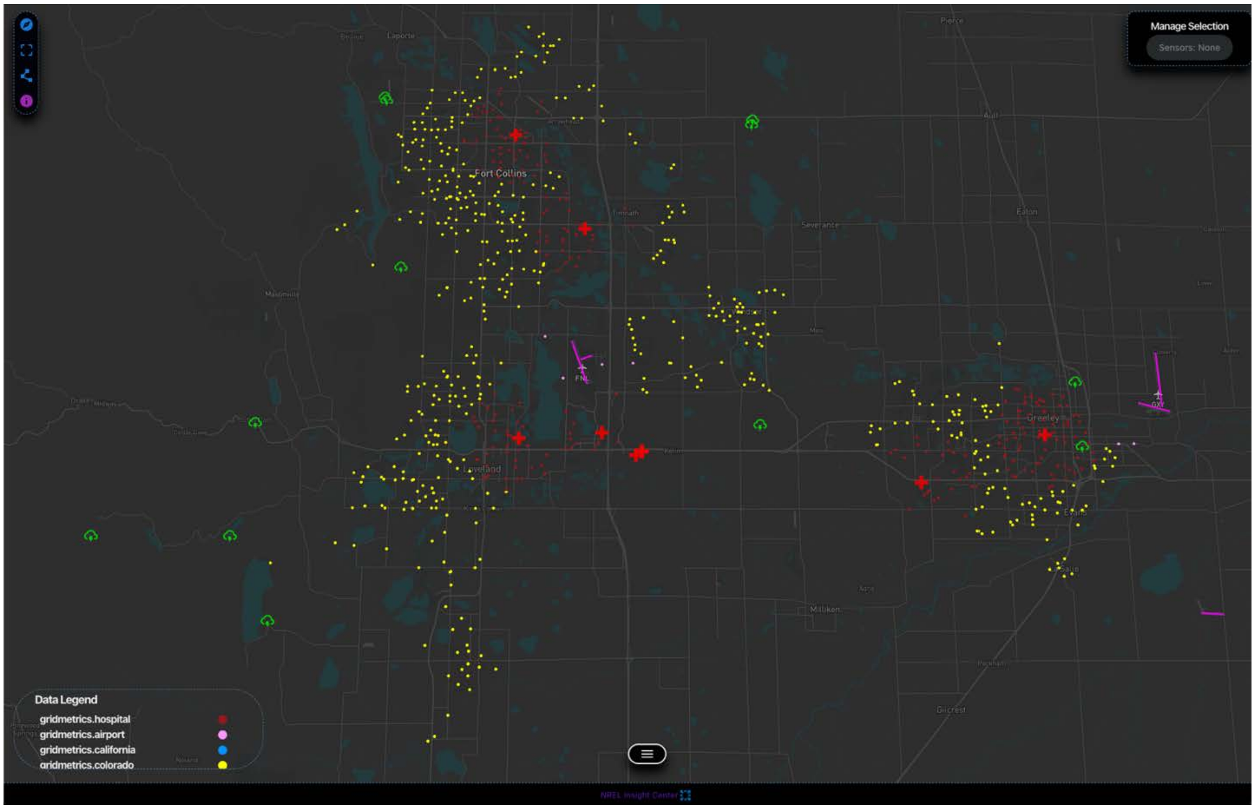


Figure 2.9. Visual analytics geospatial component focused on Fort Collins, Colorado, by using the location menu in the top left corner

The current view provides an overview of the distributions of approximately 300 Gridmetrics sensors in the suburban area, in addition to HIFLD data for hospitals (red cross icon), airports (purple runway lines), and power plants (green power icons), with red points corresponding to hospital groups, purple points corresponding to airport groups, and yellow points corresponding to the Colorado group.

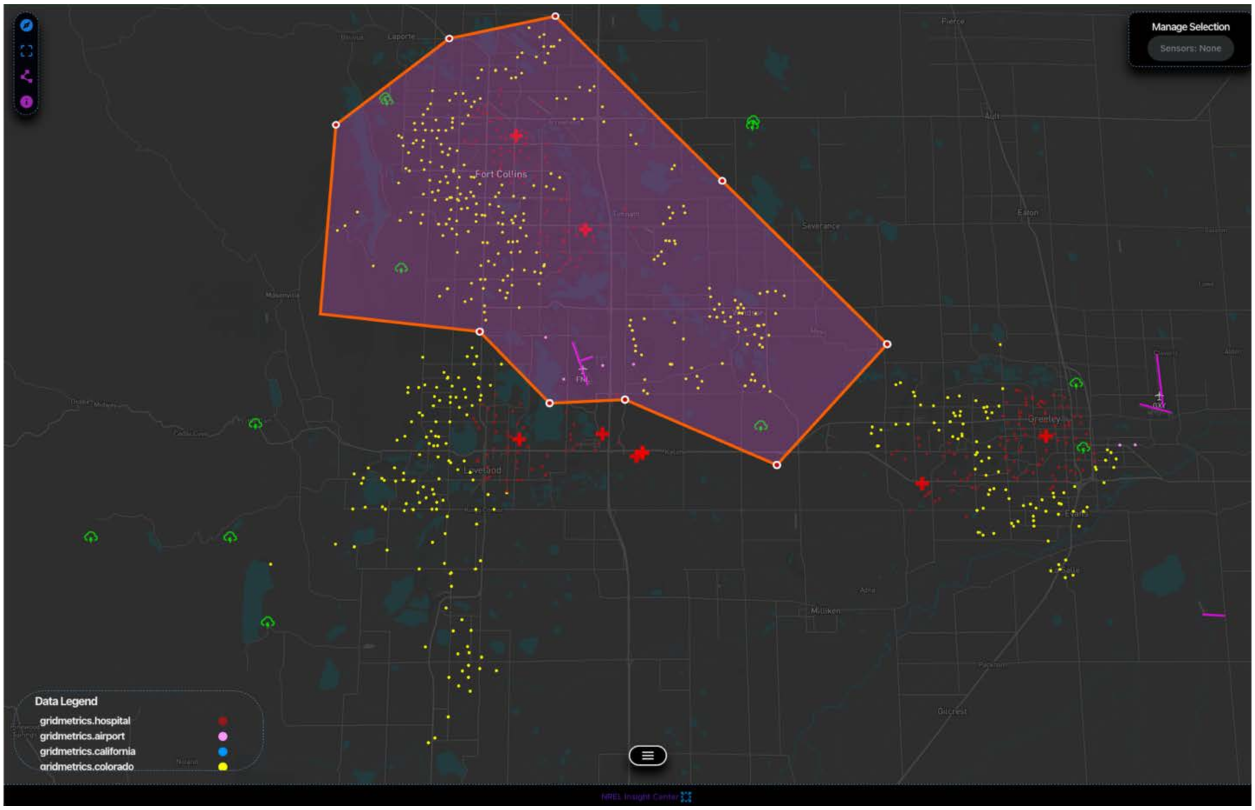
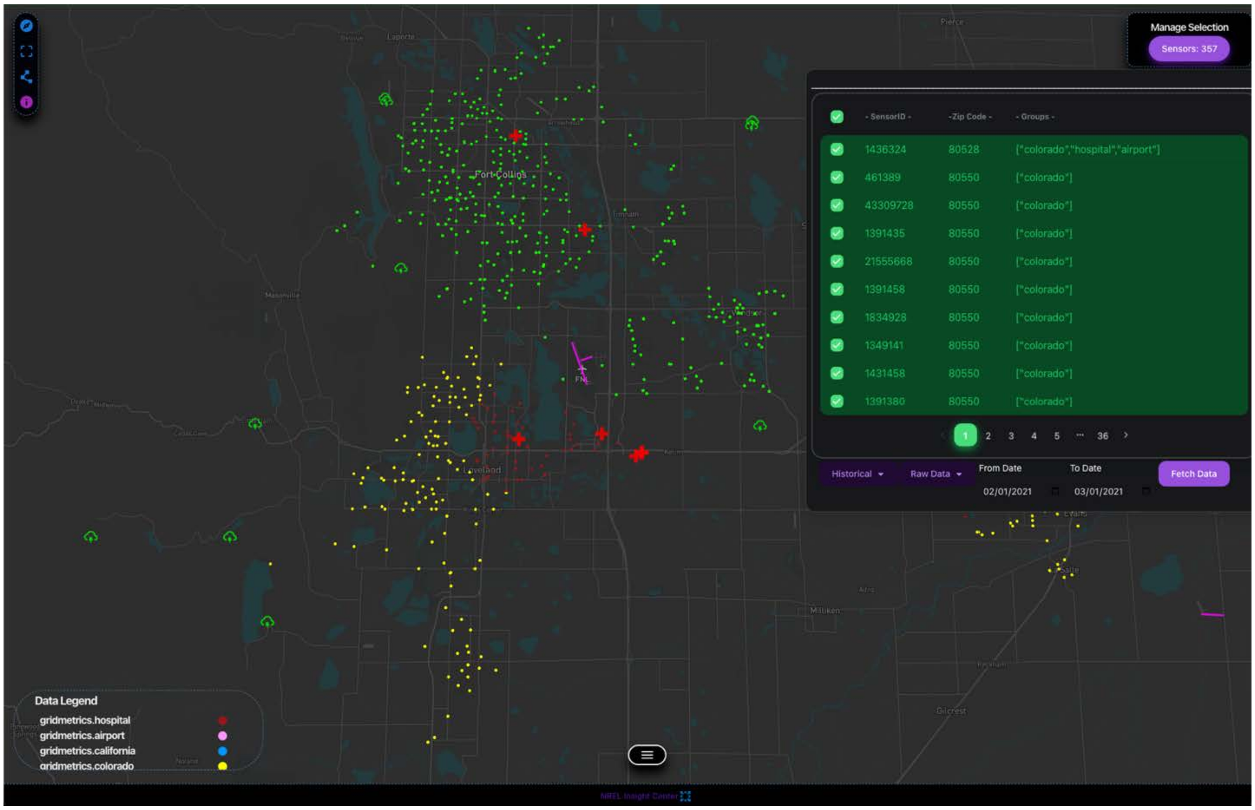


Figure 2.10. Visual analytics geospatial component focused on Fort Collins, Colorado

Sensor selections can be chosen from the polygon selection or rectangular selection in the top left menu. Here, a polygon-selection tool is being used to select sensors of interest. Selected sensors subsequently activate/populate the Manage Selections menu in the top right, shown in detail in Figure 2.12.



☑	- SensorID -	-Zip Code -	- Groups -
☑	1436324	80528	["colorado","hospital","airport"]
☑	461389	80550	["colorado"]
☑	43309728	80550	["colorado"]
☑	1391435	80550	["colorado"]
☑	21555668	80550	["colorado"]
☑	1391458	80550	["colorado"]
☑	1834928	80550	["colorado"]
☑	1349141	80550	["colorado"]
☑	1431458	80550	["colorado"]
☑	1391380	80550	["colorado"]

From Date 02/01/2021
To Date 03/01/2021

Figure 2.11. Visual analytics geospatial component focused on Fort Collins, Colorado.

Polygon-selected sensors are highlighted in green and populate the Manage Selections menu in the top right, which provides a tabular view for further sensor detail inspection and sub-selection. Selected sensors within the table can be subsequently submitted, with date-time query parameters, to the *Historical Data Service*, with service responses able to populate various time-series visualization components (samples shown in Figure 2.13 and Figure 2.14).

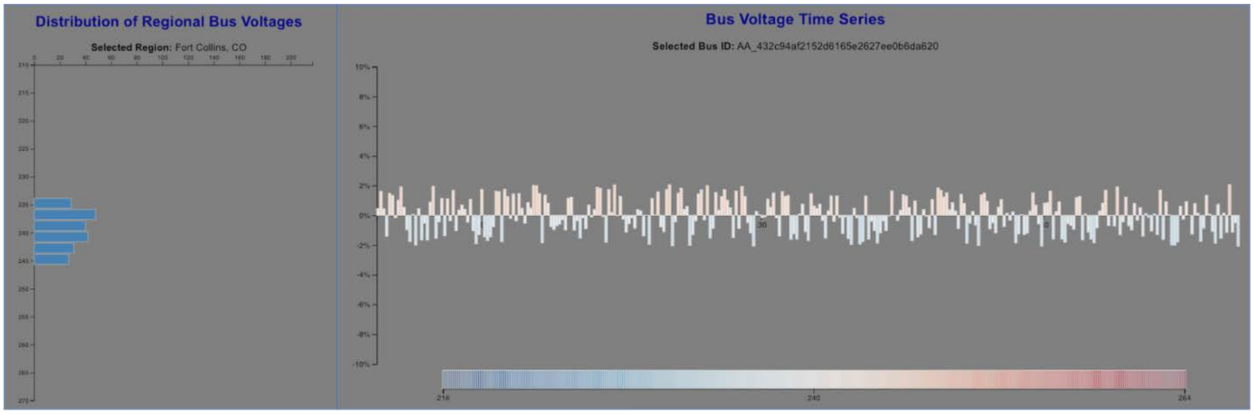
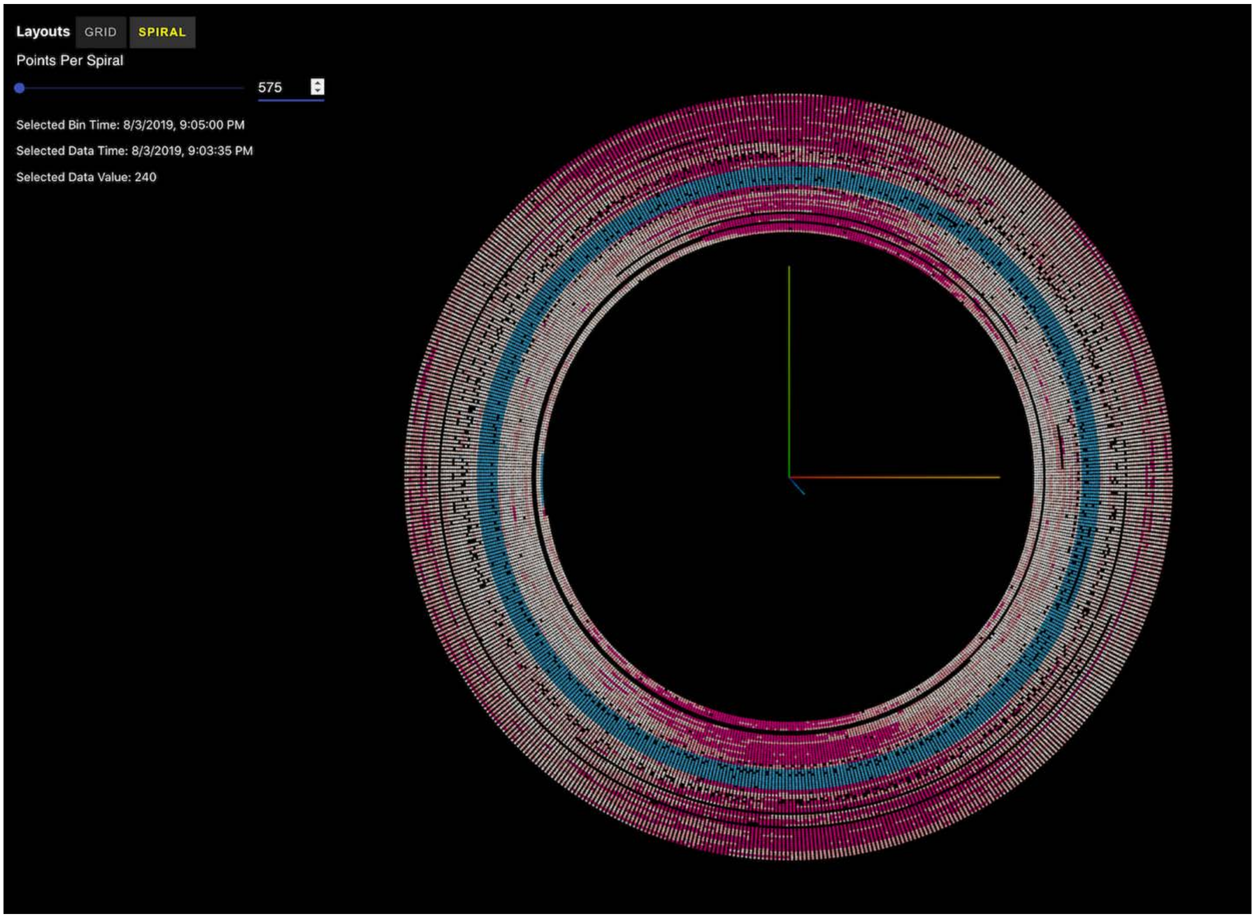


Figure 2.12. Sample of a visual analytics time-series component displaying a selected focus area's sensor voltage distribution on the left and a selected sensor's voltage time series on the right

The selected sensor time series is plotted as a diverging bar chart, with the y-axis as a percentage over/under nominal. Bars above the x-axis are over nominal, and bars under the x-axis are below nominal, with the color corresponding to the voltage value (blue for under, red for over, and white for nominal voltage value).



Points Per Spiral

575

Selected Bin Time: 8/3/2019, 9:05:00 PM

Selected Data Time: 8/3/2019, 9:03:35 PM

Selected Data Value: 240

Figure 2.13. Sample visual analytics time-series component for interactively exploring sensor voltage patterns across long periods

Here, a sensor's voltage time series from a 4-month period is rendered in a spiral pattern, with the innermost radii marking the beginning and the outermost radii marking the end of the chosen interval, looping outward counterclockwise. A user can interactively change the number of points/intervals packed per circumference/loop to alter the desired period in which to search for any patterns that would be present along any given ray outward. In this case, the current points per loop is set to 576 (where each point represents a 5-minute interval of data, of which there are 288 5-minute intervals per day); thus, a full loop counterclockwise, from the x-axis, at any radii represents successive 2-day cycles. The points are colored by their voltage value in a red-white-blue diverging pattern, where red represents an overvoltage, white represents a nominal voltage, and blue represents an undervoltage (with missing data in black). In this case, there is a period once per day when the observed voltage spikes (in the evening) and a large outage period within the band of blue (where the voltage was zero); hence, the two regions shown in darker red. This visualization technique can aid human perception to quickly uncover visual patterns in large amounts of data across a large set of possible periodicities (e.g., hour, day, week, month, quarter, season, year). Further, any identified patterns can be cross-referenced to determine their commonality. Future work will extend this technique for multiple sensors in which users could quickly and interactively compare patterns between sensors/groups and geographic areas.

The *Core Data Services* and visual analytics prototypes offer a reference into the architecture, scalability, and overall holistic utility provided by focusing on the availability of Gridmetrics cable sensor data. Although current data availability is at 1-minute intervals for roughly 100,000 sensors, the prototyped design and implementations are influenced by future plans to provide more than 600,000 sensors at sub-minute frequencies. Although there will need to be adaptations and architectural updates to support data storage size, throughput, querying, load balancing, and downstream ingestion (e.g., such as with a machine learning platform), the *Core Data Services* and web-based visual analytics tools provide a foundation that future operational technologies can reference and build upon.

2.5 Learnings, Conclusions, Recommendations

We learned that we could be successful in designing, developing, and prototyping a broadband grid scalable data systems architecture that enables continued data set expansion, high-availability and performance, advanced interactive visualization, and real-time machine learning workflows. We conclude that three services are essential—*Metadata Service, Historical Service,* and *Streaming Service*—with each service providing the appropriate functionality for exploring, monitoring, analyzing, and providing data made available by Gridmetrics sensors. We recommend funding further development of broadband-enabled, grid situational awareness and control to create entirely new possibilities for actively managing the grid to prevent cascading outages and to limit the loss of life and property.

3 Cyber-Physical Power System Simulation

High-fidelity cyber-physical power system modeling and simulation tools are essential to support preventative risk analyses on cyber threats against system stability and reliability. Such tools provide an inexpensive and risk-free environment to test various cyber-relevant events and to collect labeled system responses, which are valuable for conducting research on early cyber anomaly detection, optimal protective resource allocation, and mitigation measures. In SAGA, both dynamic and steady-state cyber-physical power system simulation tools were developed to capture the system impacts under typical cyber-physical events given a forward-looking operating condition in which distributed energy resources (DERs) are aggregated to provide frequency and voltage regulation.

3.1 Dynamic Simulation With DER-Enabled Frequency Regulation

The dynamic cyber-physical transmission-and-distribution simulation tool was built using three open-source software systems: ANDES,[3] OpenDSS,[4] and HELICS.[5] Figure 3.1 illustrates the overall simulation architecture. The colored blocks represent five types of simulation agents (modeled as HELICS federates), with processes managed and synchronized through a HELICS broker. The thick and thin arrows indicate the physical information and communication message exchanges among federates.

Transmission and distribution agents work jointly to define the grid topology, model the system components, and simulate the power flow and system dynamics. Whereas ANDES has a built-in transmission agent that solves the system electromechanical transient dynamic in the time domain every 33.3 ms, OpenDSS has a built-in distribution agent that updates the quasi-static power flow in the phasor domain every second. These two processes are coupled via the exchanges of boundary physical variables, enabled by the HELICS publish/subscribe, at the distribution feeder heads. As illustrated in Figure 3.1, the transmission agent sends voltage magnitudes to each distribution agent, and each distribution agent sends the active and reactive power to the transmission agent every second. Unlike most existing work, in which DERs are often modeled as negative loads, we applied the Western Electricity Coordinating Council distributed photovoltaics (PV) model to capture the fast dynamics of DERs enabled by the smart inverters.

Aggregator and control center agents work jointly to enable the control and automation function of the simulation tool. Agents implement the secondary frequency regulation (SFR) service procedure enabled by a centralized automatic generation control (AGC) model for restoring the system frequency to its nominal value in two steps, as illustrated in Figure 3.2:

1. In Step 1, the area control error (ACE) is calculated based on the area frequency measurement, which is then translated into the transmission-level SFR power generation

[3] ANDES is an open-source Python library for power system modeling, computation, analysis, and control; see https://docs.andes.app/en/latest/.

[4] OpenDSS is an open-source electric power distribution system simulator; see https://smartgrid.epri.com/SimulationTool.aspx.

[5] HELICS is the Hierarchical Engine for Large-scale Infrastructure Co-Simulation; see https://helics.org.

reference through an AGC proportional-integral (PI) controller inside the transmission system control center.

2. In Step 2, the substation-level SFR power generation references are calculated and sent to individual DER aggregators, which are then further disaggregated to the device level based on the participation factors of DERs.

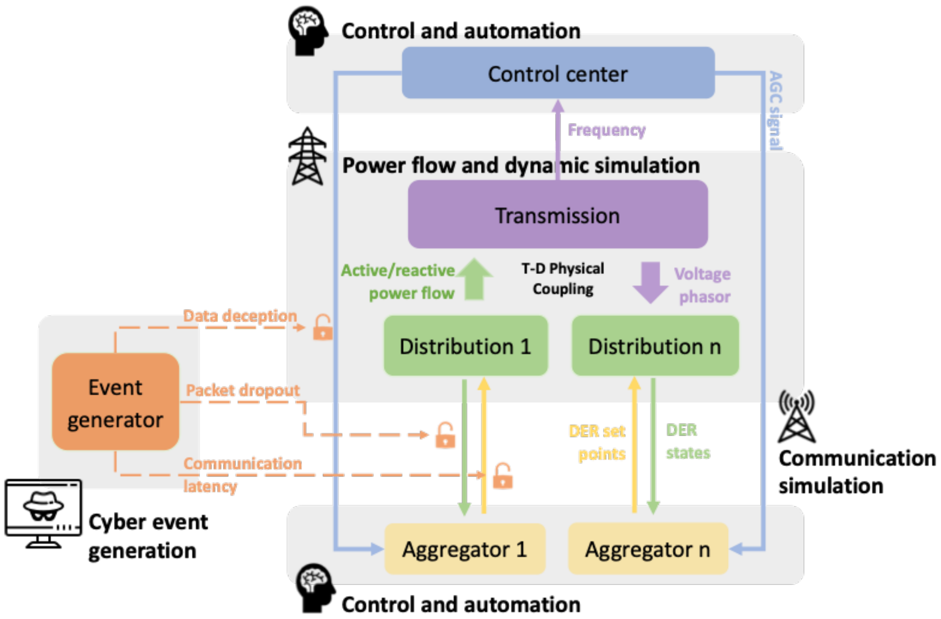


Figure 3.1. Framework of the dynamic cyber-physical simulation tool

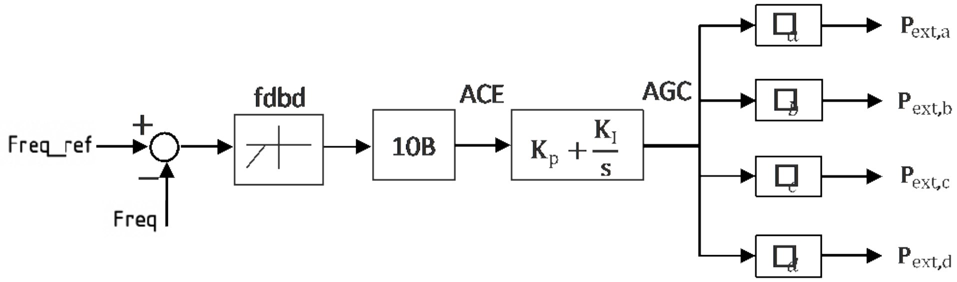


Figure 3.2. Secondary frequency regulation control diagram

Communications among the agents are modeled by HELICS end points. Each end point represents one node in a communication network. The pairwise communications between end points are defined by registering the end points at different federates and specifying their message receivers. The thin arrows in Figure 3.1 depict such pairwise relations defined in this study. Messages sent from the end points can be delayed, intercepted, or picked up by any federate to simulate the communication latency, false data injection, and packet loss by the event generator agent, as shown by the dashed arrows in Figure 3.1.

3.2 Steady-State Simulation with DER-Enabled Voltage Regulation

Figure 3.3 illustrates the framework of the steady-state cyber-physical distribution-only simulation tool. Similar to the dynamic simulation tool, it follows a multi-agent structure with four functionalities: power flow simulation, control and automation, communication simulation, and cyber event generation. The colored solid arrows represent the data flows occurring at the measurement and control communication channels supporting the voltage regulation. Note that v_i^t, $p_{curtail,j}^t$, $|v|_{ref,j}^t$, and $q_{PV,j}^t$ denote the primary voltage phasor measurements collected by the micro-phasor measurement unit located at the ith primary node, along with PV curtailment, the secondary voltage magnitude reference, and the reactive output set points sent to the smart inverter located at the jth secondary node.

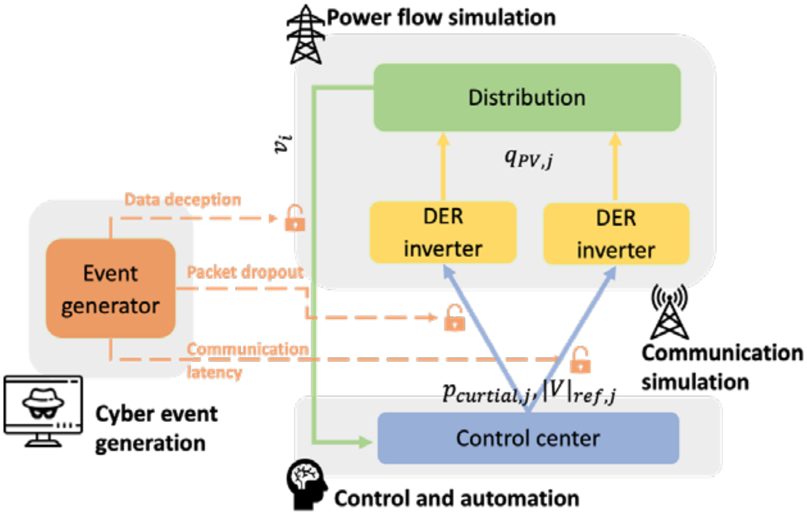


Figure 3.3. Framework of the steady-state cyber-physical simulation tool

The centralized voltage regulation control scheme is executed in a two-level fashion. At the upper level, an optimal power flow module runs inside the distribution system control center, taking primary measurements as inputs and generating PV control set points as outputs. Equations 3.1–3.7 outline the problem formulation of the optimal power flow module. Equation 3.1 aims to minimize a weighted sum of total PV active power curtailment (first summation term in Equation 3.1), the total cross-time voltage variation (second summation term in Equation 3.1), and the total voltage violation (third summation term in Equation 3.1). Equation 3.2 states a three-phase linearized power flow model that governs the relationship between the voltage magnitude (p.u.) at each primary node with the active (p.u.) and reactive power injections (p.u.) at each primary node. Equation 3.3 defines the operating boundary of the PV inverters according to the aggregated PV capacity associated with each primary node, i. Equation 3.4 restricts the primary voltage magnitude to between 0.95 p.u. and 1.05 p.u. with a soft margin $\delta_{v,i}^t$. Equation 3.5 imposes the soft margin to be positive. Inequality constraints in equations 3.6 and 3.7 are introduced to linearize the absolute voltage fluctuation, $|v|_{delta,i}^t$, between two consecutive time intervals.

$$\min_{p_{curtial}^t, |v|_{ref}^t, q_{PV}^t} \omega_{curtail} \sum_{i\in\mathcal{N}_p} p_{curtail,i}^t + \omega_v \sum_{i\in\mathcal{N}_p} |v|_{delta,i}^t + \omega_\delta \sum_{i\in\mathcal{N}_p} \delta_{v,i}^t \tag{3.1}$$

Subject to:

$$|v|_{ref}^t = C_p(-p_{load}^t + p_{PV}^t - p_{curtail}^t) + C_q(-q_{load}^t + q_{PV}^t) + c \tag{3.2}$$

$$\left(q_{PV,i}^t\right)^2 + \left(p_{PV,i}^t - p_{curtail,i}^t\right)^2 \le s_{PV,i}^2 \tag{3.3}$$

$$0.95 - \delta_{v,i}^t < |v|_{ref,i}^t < 1.05 + \delta_{v,i}^t \qquad \forall i \in \mathcal{N}_p \tag{3.4}$$

$$\delta_v^t \ge 0 \tag{3.5}$$

$$|v|^{t}_{delta,i} \geq |v|^{t}_{ref,i} - |v|^{t-1}_{se,i} \tag{3.6}$$

$$|v|^{t}_{delta,i} \geq |v|^{t-1}_{se,i} - |v|^{t}_{ref,i} \tag{3.7}$$

Notations in bold represent matrices or vectors, and non-bold indicates scalars. $\omega_{curtail}$, ω_{v}, and ω_{δ} denote configurable weights that sum to unity and can be set based on system operator preferences. $\mathcal{N}_p$ denotes the set of primary distribution nodes. p^{t}_{load}, p^{t}_{PV}, and q^{t}_{load} assemble the forecasted active loads, the active PV generation, and the reactive loads for all primary nodes. $p^{t}_{curtail}$, q^{t}_{PV}, and $|v|^{t}_{ref}$ denote the decision variables controlling the aggregated PV curtailments, the PV reactive power generation, and the reference voltage magnitudes at different primary nodes. C_p, C_q, and c are linear coefficient matrices and are vector derived based on the linear power flow model. $S_{PV,i}$ represents the sum of the apparent power capacity for distributed PV connected to the same primary node, i. Finally, $|v|^{t-1}_{se,i}$ and $|v|^{t}_{delta,i}$ represent the estimated voltage magnitude at the primary node, i, measured from the previous time interval, $t-1$, and its distance from the reference voltage magnitude, $|v|^{t}_{ref,i}$, at the current time step, t. Although this minimization problem would become computationally expensive as the system grows, the focus of SAGA is on detection, not control.

Once the problem is solved, the solution yields the optimal aggregated set points associated with each primary node, i, including the PV curtailment, $p^{t}_{curtail,i}$, and the reference voltage magnitude, $|v|^{t}_{ref,i}$. Given that all these control variables are calculated at the aggregated level, an additional disaggregation process is required to translate the set points associated with each node at the primary side to those for individual PV units at the secondary side. Equation 3.8 and Equation 3.9 describe such disaggregation rules.

$$p^{t}_{curtail,j} = \frac{s_{PV,j}}{\sum_{\mathcal{N}_{s,i}} s_{PV,j}}\, p^{t}_{curtail,i} \tag{3.8}$$

$$|v|^{t}_{ref,j} = \frac{|v|^{t-1}_{PV,j}}{|v|^{t-1}_{se,i}}\, |v|^{t}_{ref,i} \qquad j \in \mathcal{N}_{s,i} \tag{3.9}$$

Note that i and j indices of the primary and secondary nodes, respectively. $\mathcal{N}_{s,i}$ assembles secondary nodes located in the same secondary system connected to the primary node, i. According to Equation 3.8 and Equation 3.9, the aggregated PV curtailment set point, $p^{t}_{curtail,i}$, is distributed among PV units located in the same secondary system based on its apparent power capacity, $s_{PV,j}$. The ratio between the smart inverter local voltage measurement, $|v|^{t-1}_{PV,j}$, and the primary voltage magnitude measurement, $|v|^{t-1}_{se,i}$, from the previous time step, $t-1$, is applied to rescale the reference voltage magnitude, $|v|^{t}_{ref,j}$, for individual PV units.

At the lower level, each smart inverter performs the localized integral control to track with the reference voltage magnitude, $|v|^{t}_{ref,j}$, determined at the upper layer, via actively adjusting its reactive power generation, $p^{t}_{PV,i}$, following the rule given in Equation 3.10. Note that I_j indicates the integral gain of the local controller located at the secondary node, j. Such a localized control is used for hedging uncertainties introduced by the load/PV forecast and the set point disaggregation.

$$p^{t}_{PV,i} = p^{t-1}_{PV,i} - I_j\left(|v|^{t-1}_{PV,j} - |v|^{t}_{ref,j}\right) \tag{3.10}$$

Jointly, the upper and lower levels controllers coordinate distributed PV units located in the secondary systems to support the primary system voltage. Such a voltage regulation process relies on bidirectional communications, via the utility network, between physical devices located in the distribution circuit and the software program located remotely in the control center, as illustrated in Figure 3.3.

3.3 Cyber Anomaly Scenarios

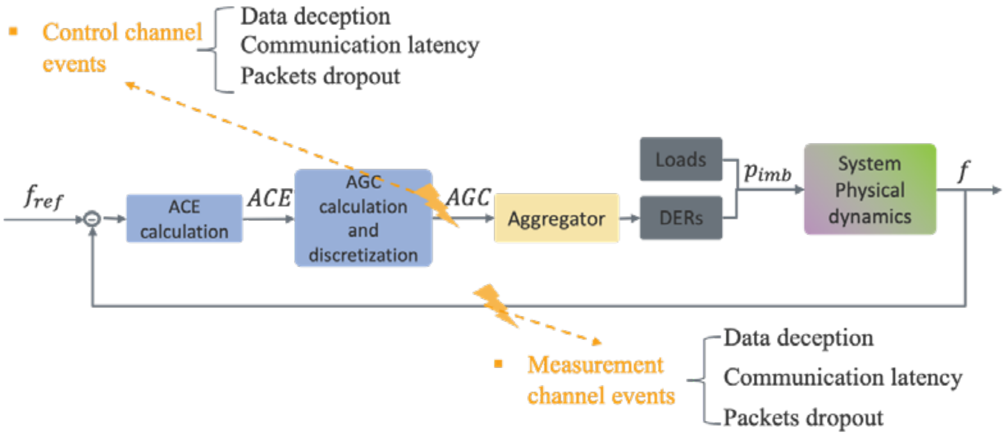


Figure 3.4. Closed-loop control diagram of the frequency regulation

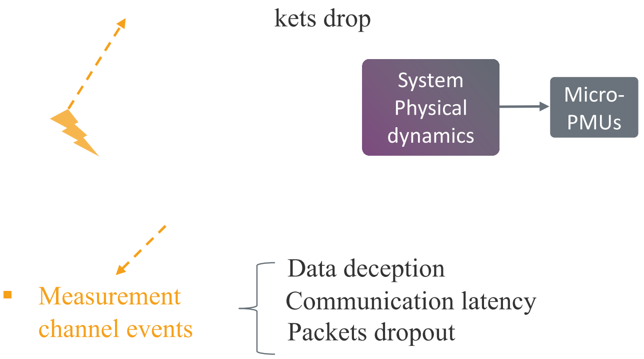


Figure 3.5. Closed-loop control diagram of the voltage regulation

The communication dependency of the frequency and voltage regulations, as discussed in the previous two subsections, result in two categories of cyber threats: measurement channel events and control channel events, based on the entry point of the cyber anomaly, as shown in Figure 3.4 and Figure 3.5. Measurement channels are used for transmitting frequency or voltage measurements, collected by remote terminal units at multiple physical locations, to the central station of a supervisory control and data acquisition (SCADA) system. Control channels are used for transmitting DER set points, calculated at the central station of the SCADA system, to the aggregators and further to individual DER controllers at the grid edge. Under each category, three types of cyber events are of interest to be investigated in SAGA:

1. Data deception: can be triggered by malicious packets injected into communication channels through man-in-the-middle attack techniques, such a manipulation, which can be mathematically expressed as:

$$G_j(t) = G_j^0(t) + \gamma_j(t) \tag{3.11}$$

   where $G_j(t)$ denotes the breached control signal received by DER j at time t, which equals the authentic control signal, $G_j^0(t)$, superposed by an attack vector, $\gamma_j(t)$.
2. Communication latency: refers to the length of time it takes for data fed into one end of a network to emerge at the other end. Significant communication latencies can be introduced by common denial-of-service attacks.
3. Packet dropout: when one or more communication packets fails to reach its intended destination. Significant packet dropout can be another major consequence of a denial-of-service attack.

Event generator agents have been developed and integrated into both simulation tools to enable customized simulations of cyber events. Agents take scripts (formatted in JSON) defining metadata relevant to the cyber-physical events as inputs and launch the data injection to the main time-series simulation in an event-driven manner. The metadata required for defining a cyber-physical event include:

- *TargetFeeder*: Defines the distribution feeder targeted by the adversary
- *EventType*: Defines the type of the event. Currently, four types of events are supported by the test bed: load perturbation events, DER availability events, AGC measurement channel events, and AGC control channel events. Details regarding each type can be found in Section III. Additional types of events can be added by interested developers.
- *TargetNum*: Defines the number of system components that need to be simultaneously compromised by the adversary to successfully create the event
- *TargetName*: Defines the names of all system components being compromised during the event
- *AttackMagnitude*: Defines the magnitudes of the disruptions added to each system component
- *StartTime*: Defines the time when the event occurs
- *EndTime*: Defines the time when the event is cleared.

During a time series simulation, multiple events with overlapping/nonoverlapping time windows can be injected by customizing the *StartTime* and *EndTime* of all events.

3.4 Simulation Results

Three sets of comparative case studies have been conducted to illustrate how the proposed cyber-physical power system simulation tools can be leveraged for assessing the system impacts under various cyber-physical events. To demonstrate the scalability of the simulation tools, a transmission-and-distribution network has been modeled in the dynamic simulation tool with a

2,000-bus synthetic transmission network[6] connected with 122 detailed distribution feeders,[7] covering the footprint of Texas; and a 1,600-node distribution network has been modeled in the steady-state simulation tool based on real-world system operating data. It was assumed that 148 Gridmetrics sensors are installed in the secondary system. All the simulations were performed on the high-performance computer Eagle at NREL.

3.4.1 Synchronous Versus Asynchronous Communication Latency

In the first case study, we compared the system impacts caused by the synchronous versus asynchronous communication latency events aimed at compromising the frequency regulation performance in response to a typical transmission contingency (generator trip). Figure 3.6 illustrates the frequency trajectories under synchronous communication latencies (i.e., the same communication delays are applied to all targeted communication channels) varying from 4 s to 16 s at 4-s increments. Figure 3.7 illustrates frequency trajectories under asynchronous communication latencies (i.e., various communication delays are applied across targeted communication channels; note that in this study we assume that the communication delays follow a uniform distribution). The asynchronous communication latencies shown in Figure 3.7 are with the same mean, 8 s, yet various variances, ranging from 3 s (6–12 s uniformly distributed) to 27 s (0–18 s uniformly distributed). Note that these cyber events coincided with a transmission contingency (generation trip) that occurred at 8 s. According to Figure 3.6, the greater the synchronous latency, the longer the frequency fluctuates before settling. Comparing the frequency trajectory under the 8-s synchronous communication latencies in Figure 3.6 with the frequency trajectories in Figure 3.7 clearly shows that synchronous communication latencies are riskier than asynchronous communication latencies. In addition, asynchronous communication latencies could even help suppress the frequency overshoots when the coefficients of the PI controllers are not well designed.

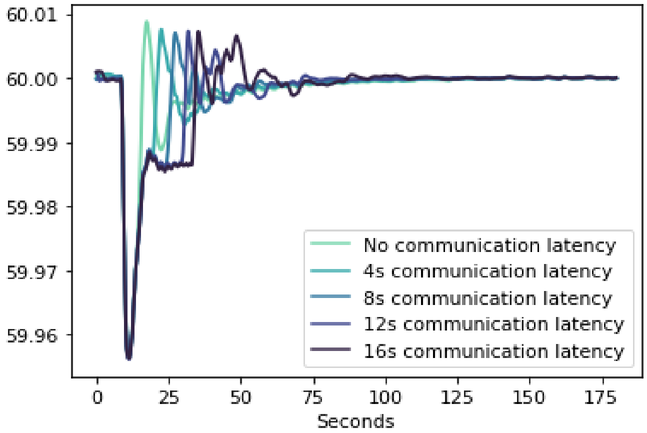


Figure 3.6. Frequency trajectories under different levels of synchronous communication latencies

[6] Texas A&M University Electric Grid Datasets; see https://electricgrids.engr.tamu.edu.

[7] Krishnan, Venkat K., Bryan S. Palmintier, Bri-Mathias. Hodge, Elaine T. Hale, Tarek Elgindy, Bruce Bugbee, Michael N. Rossol, Anthony J. Lopez, Dheepak Krishnamurthy, Claudio Vergara et al. 2017. “SMART-DS: Synthetic Models for Advanced, Realistic Testing—Distribution Systems and Scenarios.” Golden, CO: National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy17osti/68764.pdf.

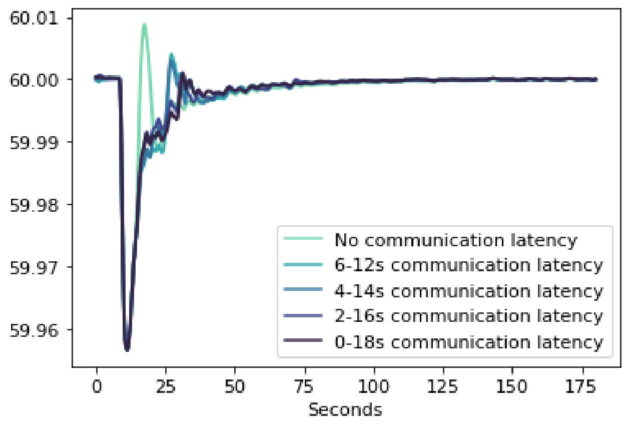


Figure 3.7. Frequency trajectories under different levels of asynchronous communication latencies

3.4.2 Measurement Channel Versus Control Channel Events

To understand different influences of cyber events occurring on different communication links of a closed-loop frequency control in response to a transmission contingency (generator trip), we further performed a comparison between cyber events, i.e., communication latency and packet dropout, occurring at the measurement and control channels. The same frequency trajectories were obtained for the communication latency events whether the attack entered at the control channels or at the measurement channels. In other words, system impacts under communication latency events were not affected by the location of the attack; however, this is not the case for the packet dropout events, as shown in Figure 3.8 and Figure 3.9, which plots frequency trajectories under various packet dropout rates for events happening on the measurement channels versus the control channels. Results show that system impacts under the same dropout rate are more significant for events happening at the control channels than the measurement channels. This is because we implemented a PI controller to translate the ACE signals to the AGC signals, so the loss of an AGC packet results in a greater compensation deficiency than the loss of a frequency measurement/ACE packet. This demonstrated that system impacts under packet dropouts (losses) are significantly affected by the location of the attack.

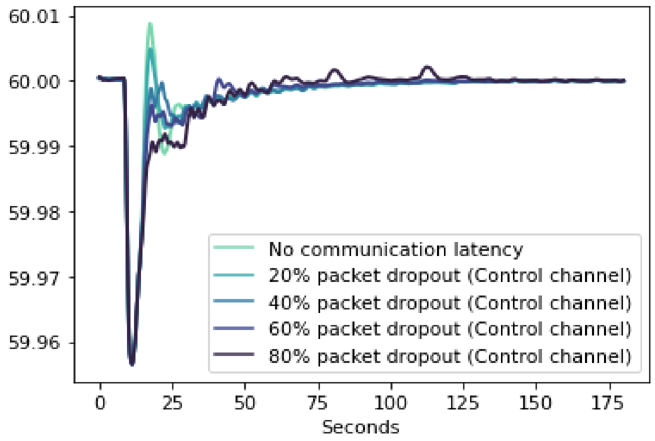


Figure 3.8. Frequency trajectories under different levels of packet dropouts occurring at the control channels

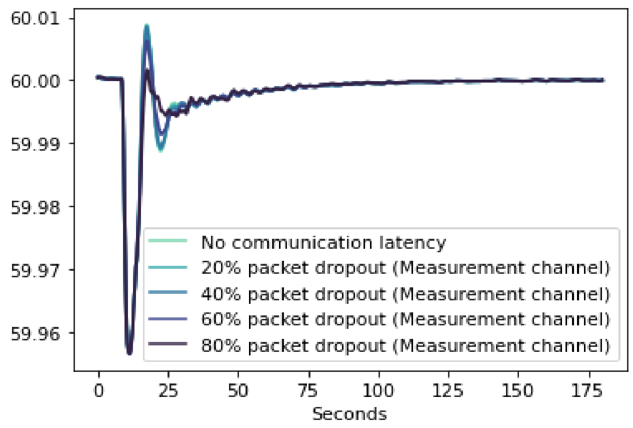


Figure 3.9. Frequency trajectories under different levels of packet dropouts occurring at the measurement channels

3.4.3 False Data Injection Attack

Figures 3.10–3.12 compare simulated Gridmetrics sensor measurements under different scenarios, e.g., normal operation, a control channel data deception event, and a measurement channel data deception event. To ensure fairness of the comparison, we ran the simulation for 1 day at a 1-minute time resolution under the same load/PV time-series profiles and forecasts. It was assumed that only the measurement/control channels directly connected with 1 of 148 secondary systems were targeted by the adversary under the control channel and the measurement channel data deception events. Further, these two attacks enter the system control loop at the same time. Each colored line in figures 3.10–3.12 denotes simulated observations for a given sensor. Figure 3.10 depicts the load variability in normal operation. The subtle differences among figures 3.10–3.11 are shown in the top green trace, which shows approximately 15 downward spikes in the presence of a *control channel* data deception event. The nearly imperceptible differences between Figure 3.10 and Figure 3.12 suggest that the *measurement channel* data deception events are more difficult to detect. The system-level active load, reactive load, and PV generation profiles applied in the daily simulation are illustrated in Figure 3.13.

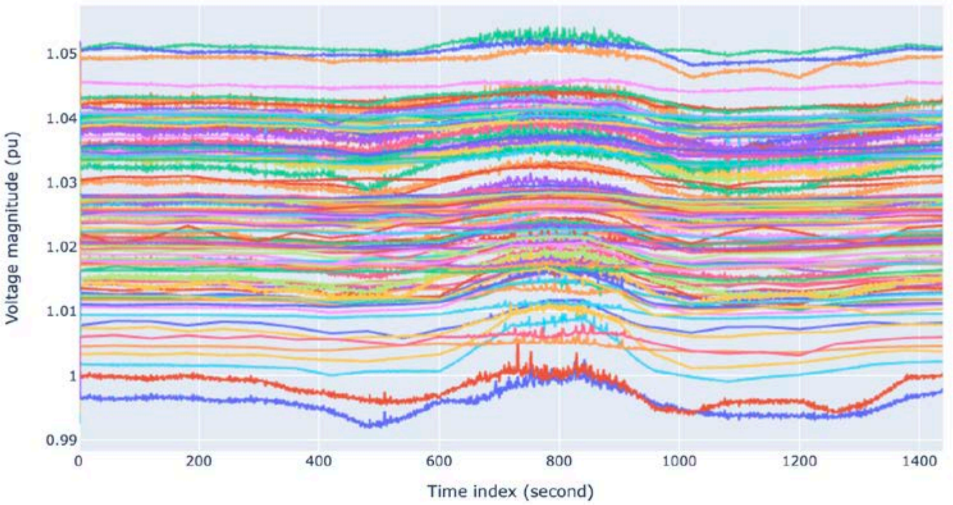


Figure 3.10. Gridmetrics voltage measurements under normal operation

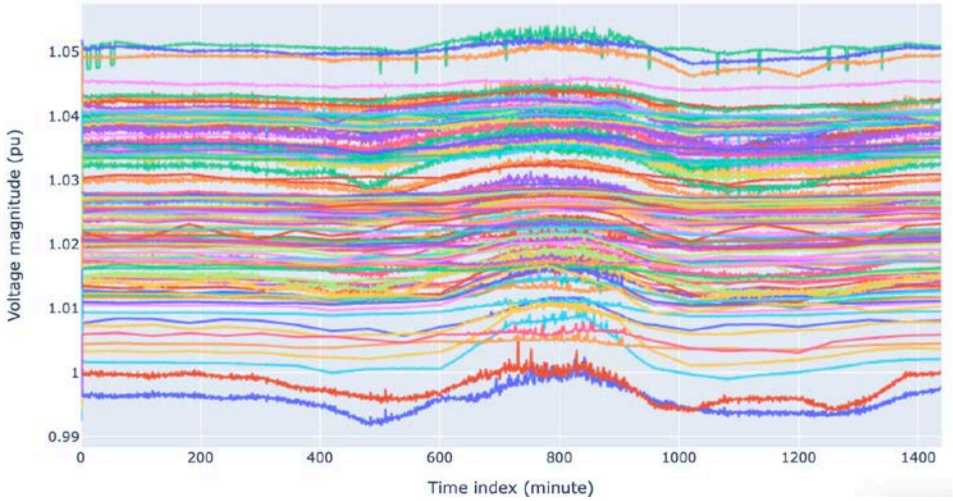


Figure 3.11. Gridmetrics voltage measurements under a control channel data deception event

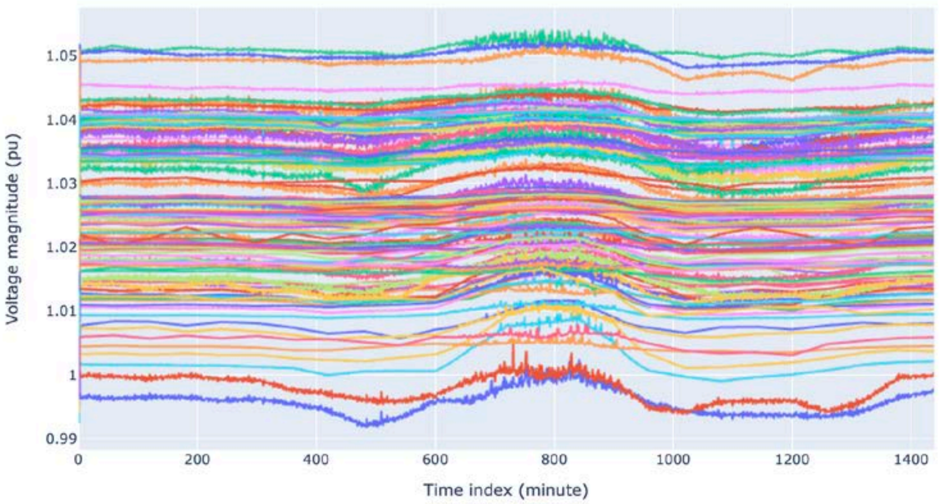


Figure 3.12. Gridmetrics voltage measurements under a measurement channel data deception event

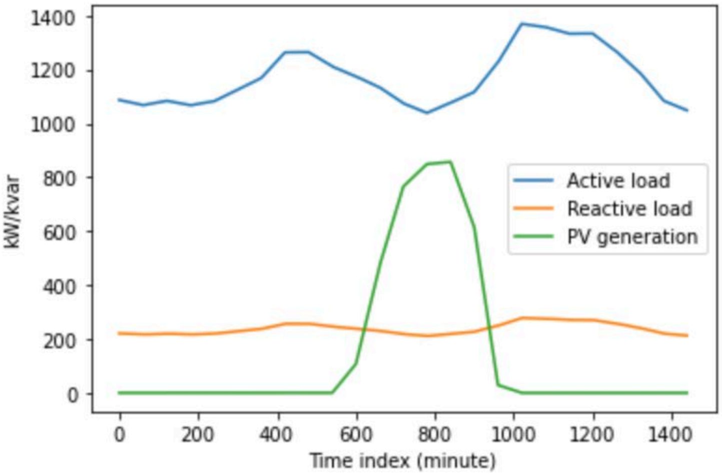


Figure 3.13. Daily system load and PV profiles

According to figures 3.10–3.13, the following observations are made:

1. The Gridmetrics measurements stay relatively flat within the day and increase slightly during midday due to higher PV generation.
2. Despite the overall smooth voltage variation, small voltage fluctuations are observed across the day due to uncertainties associated with the load/PV forecasting and set point disaggregation. Fluctuations are greater during the middle of the day, possibly because greater PV forecasting errors occur.
3. Differences between the Gridmetrics measurement profiles observed under normal operation versus under cyber events are subtle, provided that the simulated cyber events are at small scale (only a small number of communication channels are being affected). It is hard to tell via the naked human eye directly from the voltage profile when an event occurs, which justifies the necessity of a data analytics tool that automates anomaly detection.

3.5 Learnings, Conclusions, Recommendations

We learned that high-fidelity cyber-physical power system modeling and simulation tools are essential to support preventative risk analyses on cyber threats against system stability and reliability. We conclude that such tools provide an inexpensive and risk-free environment to test various cyber-relevant events and to collect labeled system responses, which are valuable for conducting research on cyber anomaly detection, optimal protective resource allocation, and mitigation measures. We recommend funding further exploration and quantification of the impacts on grid reliability due to data deception, communication latency, and packet dropouts.

4 Anomaly Detection

DER integration is gaining momentum on a worldwide scale. The active involvement of DERs in system operation entails efficient and robust communication support. Given the large communication surface, cyber anomalies—whether caused by network-induced delays or malicious data injection—are exposed to the DER-grid integrated control network and are inevitable. Such anomalies could compromise grid control performance, damage physical devices, or even jeopardize the system stability. One key preventative strategy to secure the system from cyber vulnerabilities is through proactive anomaly detection. In this section, we investigated how Gridmetrics voltage measurements, as an additional system situational awareness resource located outside the utility communication network, can be leveraged to detect cyber anomalies occurring in the utility communication network. In particular, a data-driven anomaly detection model was developed to detect cyber anomalies exposed to the utility communication network when providing DER-enabled voltage regulation.

4.1 Distributed Anomaly Detection Framework

To reduce the communication burden of the algorithm, a data-driven anomaly detection framework was designed following a distributed architecture, as illustrated in Fig 4.1. Specifically, each Gridmetrics sensor was equipped with an edge intelligence unit with a built-in prediction model. The prediction model estimated the expected value or range of the associated Gridmetrics voltage measurement based on historical operating data, assuming that the system will continue operate under normal conditions. As such, cyber anomalies could be captured according to deviations between the incoming voltage measurements and the expected voltage measurements at different locations. At each time step, each edge intelligence function provided an anomaly score, measuring the likelihood of a cyber anomaly based on the deviation. Anomaly scores provided by all edge intelligences were collected by an upstream centralized anomaly detector to determine a global score via the calculation of a weighting function, as shown in Figure 4.1. In our study, the global anomaly score was calculated as the average of all unit-level anomaly scores, and a threshold-based anomaly detection scheme was applied.

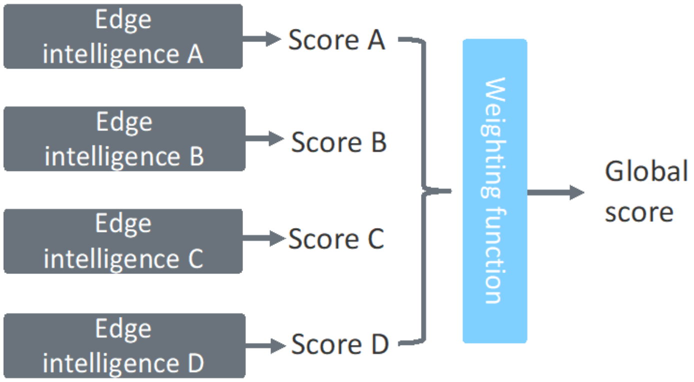


Figure 4.1. Illustration of the distributed anomaly detection framework

Figure 4.2 depicts what are inside each edge intelligence function. Each function contains a prediction model and a scoring function. The prediction model takes active/reactive load forecasts ($P_{forecast}{}_{t+1}^{j}/Q_{forecast}{}_{t+1}^{j}$), PV generation forecasts ($PV_{forecast}{}_{t+1}^{j}$), and local voltage measurements (V_t^j) at the current time step, *t*, as inputs and generates the estimated voltage measurements ($\tilde{V}_{t+1}^j$) for the next time step, *t+1*, as the output. To reduce the communication burden during the real-time implementation, a physics-based feature selection method is applied. Each node *i* only needs to collect the forecasted load and PV generation data from the acting nodes ($j \in N_i$) that are directly connected to the observing node *i*. When time advances to *t+1*, the scoring function is executed to measure how the incoming actual voltage measurement deviates from its expected value.

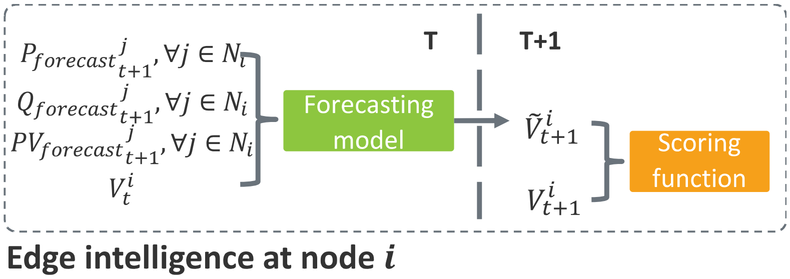

$P_{forecast}{}_{t+1}^{j}, \forall j \in N_i$
$Q_{forecast}{}_{t+1}^{j}, \forall j \in N_i$
$PV_{forecast}{}_{t+1}^{j}, \forall j \in N_i$
V_t^i
Forecasting model
T
T+1
$\tilde{V}_{t+1}^i$
V_{t+1}^i
Scoring function
Edge intelligence at node i

Figure 4.2. Structure of the prediction model

Based on the problem formulation previously introduced in equations 3.1–3.7:

- Under normal operation, the input parameters p_{load}^t, p_{PV}^t, q_{load}^t, and $|v|_{se}^{t-1}$ uniquely determine the optimal solutions, $p_{curtial}^t$, $|v|_{ref}^t$, and q_{PV}^t for the optimal power flow problem, and by following which, the system runs at an optimal operating point.
- Under the control channel data deception event, $|v|_{ref}^t$ will be modified to $\widetilde{|v|}_{ref}^t$, which could directly misguide the system to a suboptimal operating point.
- Under the communication channel data deception event, $|v|_{se}^{t-1}$ will be modified to $\widetilde{|v|}_{se}^{t-1}$, which results in suboptimal solutions, $\tilde{p}_{curtial}^t$, $\widetilde{|v|}_{ref}^t$, $\tilde{q}_{PV}^t$, and consequently misleads the system to a suboptimal operating point.

When the system operates under optimal versus suboptimal operating points, it is expected that the relationship between the forecasted load and the PV generation, p_{load}^t, q_{load}^t, and p_{PV}^t, Gridmetrics measurements from the previous and current time steps, $|v|_{Gridmetrics}^{t-1}$ and $|v|_{Gridmetrics}^t$, follow different patterns; therefore, by fitting a data-driven predictive model, $f(\cdot)$, as shown in Equation 4.1, that can capture such an underlying relationship in the optimal operating space and predict the expected value of the Gridmetrics measurement, $\widehat{|v|}_{Gridmetrics}^t$, any deviation from the expected measurement could indicate an anomaly.

$$\widehat{|v|}_{Gridmetrics,k}^t = f\left(p_{load}^t, p_{PV}^t, q_{load}^t, |v|_{Gridmetrics,k}^{t-1} \right) \tag{4.1}$$

4.2 Deterministic Versus Probabilistic Approaches

Both a deterministic approach and a probabilistic approach have been applied for fitting the predictive model, $f(\cdot)$, and defining the scoring function. Table 4.1 summarizes the difference between the approaches.

Table 4.1. Difference Between the Deterministic and Probabilistic Anomaly Detection Approaches

Component	Deterministic Approach	Probabilistic Approach
Forecasting model	Linear regression models are trained to fit the expected value ($\bar{v}_{t+1}^{j}$) of the voltage measurements.	Linear quantile regression models are trained to fit the 2.5 ($v_{2.5,t+1}^{j}$), 50 ($v_{50,t+1}^{j}$), and 97.5 ($v_{97.5,t+1}^{j}$) percentiles of the voltage measurements.
Scoring function	$s_d^i = \lvert v_t^i - \bar{v}_{t+1}^i \rvert$	$s_p^i = \dfrac{\lvert v_t^i - \bar{v}_{50,t+1}^i \rvert}{\lvert v_{97.5,t+1}^j - v_{2.5,t+1}^j \rvert}$

The linear regression model and the linear quantile regression model differ in the loss function. The loss function applied in the linear regression problem is the mean squared error. When we train a linear model to minimize the mean squared error, it will try to fit the expected mean value of the output distribution. To fit a linear quantile regression model that predicts quantiles of the output distribution, we applied the pinball loss function, as stated in Equation (4.2), instead.

$$\mathcal{L}(\xi_i|\tau) = \begin{cases} \tau\xi_i & if\ \xi_i \geq 0 \\ (\tau - 1)\xi_i & if\ \xi_i < 0 \end{cases} \tag{4.2}$$

where τ is the required quantile (a value between 0 and 1), and:

$$\xi_i = y - \hat{y} \tag{4.3}$$

Figure 4.3 illustrates the pinball loss with respect to the ξ_i when τ > 0.5.

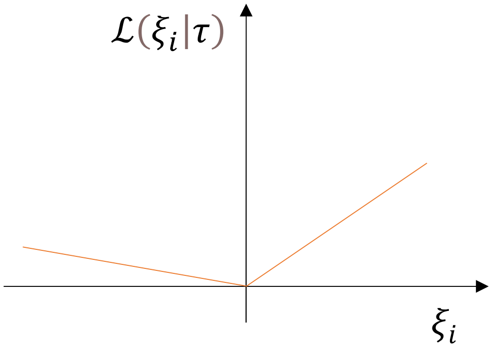


Figure 4.3. Illustration of the pinball loss when τ > 0.5

Regarding the scoring function, whereas the deterministic score captures the relative deviation between the actual voltage measurement and its expectation, the probabilistic score measures the relative deviation between the actual voltage measurement and its medium with respect to the width of the 95% confidence interval.

4.3 Model Fitting Procedure

The following describes the steps taken for training and testing both the deterministic and the probabilistic prediction models.

1. <u>Training and testing data generation and splitting:</u> We ran voltage regulation control on the test bed distribution model using OpenDSS for 1 month at a 1-minute time resolution

under two different operating scenarios: normal operation and a control channel data deception event. We then collected two sets of load/PV forecasts and Gridmetrics measurements. Time series collected under the normal operation were split by a 7:3 ratio as the training and testing data sets for evaluating the model prediction accuracy. Time series collected under the cyber events were applied for evaluating the anomaly detection accuracy.

2. Model training: We applied the *LinearRegression* function built into the scikit-learn Python package for model fitting along with the *StandardScaler* function for data standardization.
3. Modeling testing: Testing data sets were fed into trained predictive models for evaluating the prediction accuracy. R-square was used as the performance metric for the deterministic predictive model, and we evaluated the performance of the probabilistic predictive model by measuring the observing frequency with which the testing data points were within the 95% confidence intervals.
4. Anomaly detection accuracy: Precision, Recall, and F-score were applied as the performance metrics for the anomaly detection, which was calculated based on the True positive rate (TP), false positive rate (FP), true negative rate (TN), and false negative rate (FN), as given in equations 4.4–4.6.

$$Precision = \frac{TP}{TP + FP} \tag{4.4}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.5}$$

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.6}$$

These were applied to provide a quantitative comparison between the deterministic and probabilistic anomaly approaches. Whereas precision quantifies the percentage of true events among all identified events, recall measures the percentage of events that are being identified out of all true events in the testing data set. The F-score reflects the harmonic mean between the precision and recall.

4.4 Performance Evaluation

4.4.1 Performance of the Deterministic Approach

Figure 4.4 illustrates the distribution of the R-square values across 148 deterministic prediction models (using 148 Gridmetrics sensors in the test bed system).

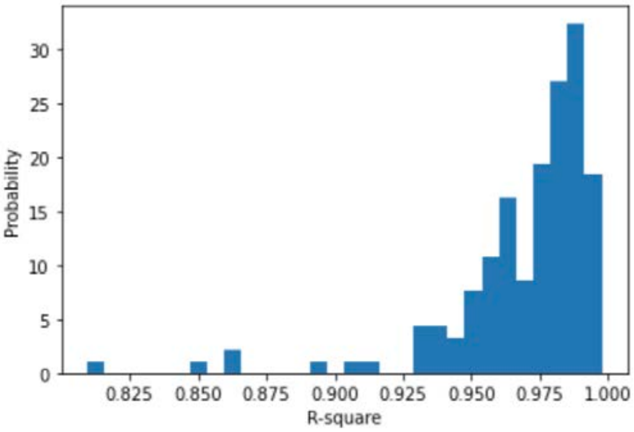


Figure 4.4. Distribution of R-square values across 148 deterministic prediction models

According to Figure 4.4, most deterministic prediction models have an R-square value above 0.9, which indicates an overall good performance of the deterministic models in capturing the expected Gridmetrics measurements at different locations under normal operation. Figure 4.5 depicts the anomaly detection results when the threshold is set at 0.15. Purple dots indicate the time when an anomaly alarm is triggered. Green dots, on the other hand, indicate the ground truth attack time.

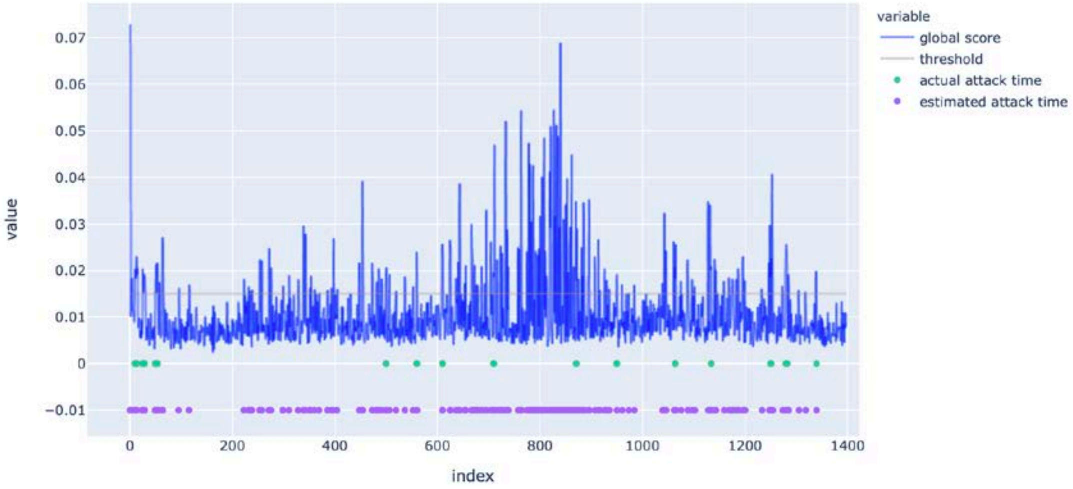


Figure 4.5. Anomaly detection results of the deterministic approach when the threshold equals 0.15

Figure 4.6 depicts how the three anomaly detection metrics vary against different threshold values. We observed: (1) The lower the threshold value, the higher the recall, because the anomaly detection is more sensitive to the deviations with a lower threshold. (2) Precision shows a lower value with either a high threshold or a low threshold, and it reached its highest value when the threshold is set as 0.015; the highest F-score value was when the threshold equaled 0.015.

However, no matter how we varied the threshold, the F-score value stayed below 0.4. This means that the anomaly detector tended to raise false alarms frequently. Figure 3.10 shows that even without the influence from cyber events, the natural voltage variation changes over time;

therefore, a fixed threshold won't work all the time because it cannot effectively consider the different uncertainty levels associated with the voltage variations. Missed cyberattacks could be very problematic. As such, future development should focus on better reliability of probability forecasts so as not to miss cyberattacks.

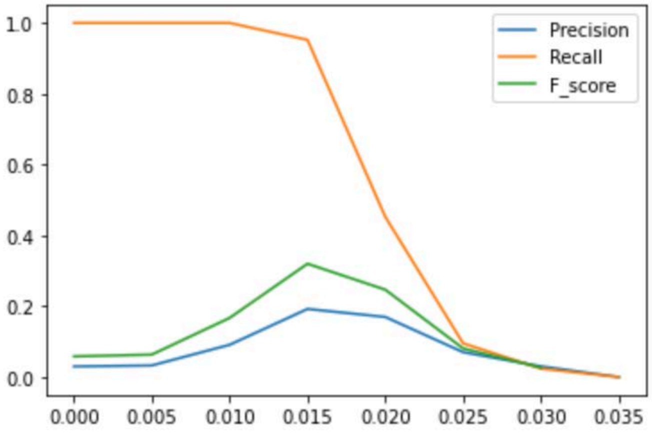


Figure 4.6. Anomaly detection performance of the deterministic approach varies against the threshold value

4.4.2 Performance of the Probabilistic Approach

Figure 4.7 shows the output generated from a single probabilistic prediction model. The gray line plots the forecasted medium, and the green shaded area indicates the 95% confidence interval.

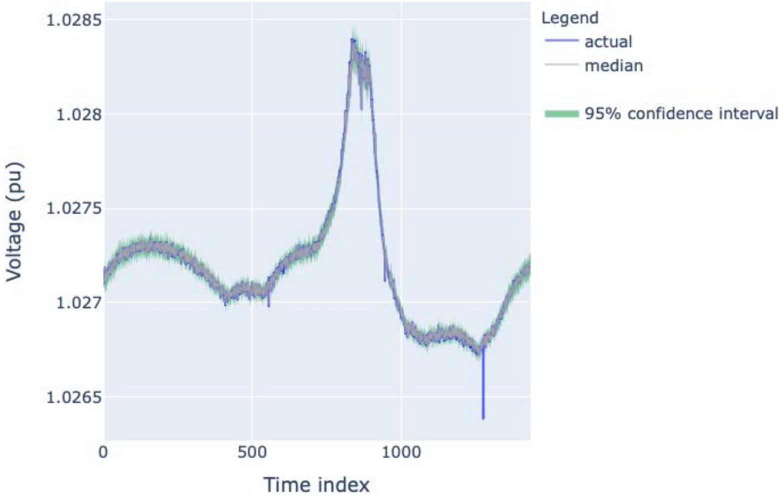


Figure 4.7. Output generated from a signal probabilistic prediction model

Figure 4.8 shows the final output obtained by the centralized anomaly detector. The red points in Figure 4.8 indicate the actual attack time. The gray line plots the global anomaly scores calculated based on an averaging weighting function. The estimated attack times (indicated by the green dots) are identified in this case based on a 0.5 threshold value.

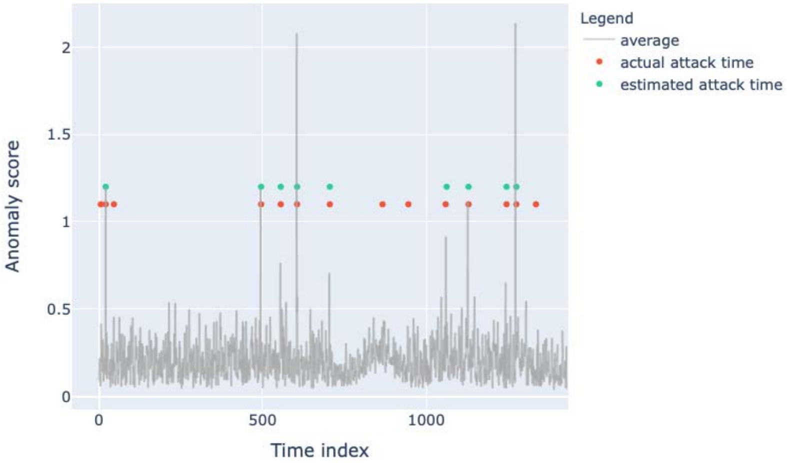


Figure 4.8. Result of the probabilistic anomaly detection

Figure 4.9 illustrates the histogram of observedfrequencies, e.g., the frequency at which the actual Gridmetrics measurements are between the 2.5 and 97.5 percentiles of forecast across the 148 probabilistic prediction models. The majority fall into the [92.5%, 97.5%] interval, with the mode (93.9%) and mean (93.3%) values being slightly lower than 95%.

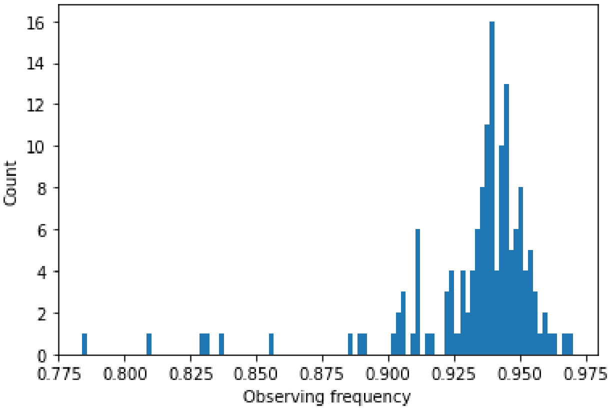


Figure 4.9. Histogram of the observing frequencies across 148 probabilistic prediction models

4.4.3 Performance Comparison

To provide a comprehensive comparison between the deterministic and probabilistic approaches, we further conducted a sensitivity analysis by varying the number of attack targets from 1 to 6, and we obtained the performance results shown in figures 4.10 and 4.11 for both the deterministic and probabilistic approaches. It was clearly demonstrated that the probabilistic approach outperformed the deterministic approach by resulting in overall higher precision, recall and, F-score values.

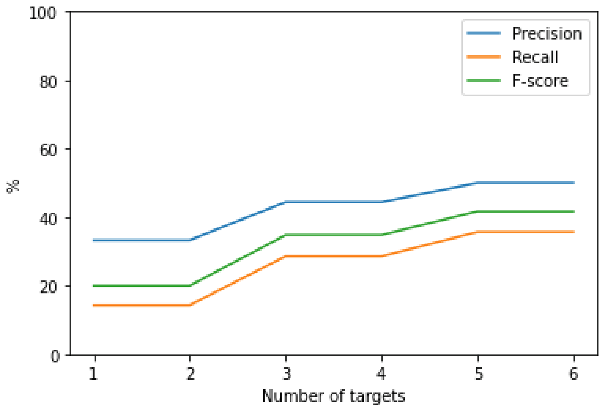


Figure 4.10. Anomaly detection performance of the deterministic approach under different numbers of attack targets

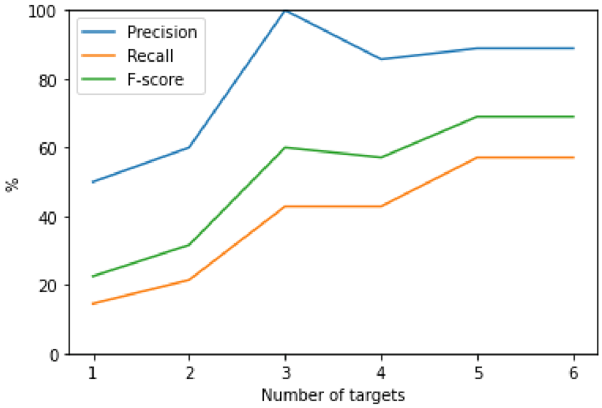


Figure 4.11. Anomaly detection performance of the probabilistic approach under different numbers of attack targets

4.5 Learnings, Conclusions, Recommendations

We learned that the active involvement of DERs in system operation requires efficient and robust communication support to detect and mitigate cyberattacks and gaps in operational situational awareness. We conclude that given the large communication surface, cyber anomalies, whether caused by network-induced delays or malicious data injections, will be exposed to DER-grid integrated control networks and are inevitable. Further, anomalies could compromise grid control performance, damage physical devices, or even jeopardize the system stability. We recommend funding the development of preventative strategies to secure the system from cyber vulnerabilities through proactive anomaly detection.

References to SAGA-Related Papers

American National Standards Institute. 2021. “American National Standard ANSI/SCTE 271: Requirements for Power Sensing in Cable and Utility Networks.” Society of Cable Telecommunications Engineers.

Cable Television Laboratories, Inc. 2022. “Gridmetrics White Paper.”

Cai, M., W. Wang, X. Fang, and A. Florita. 2023. “A Scalable Cyber-Physical Dynamic Simulation Test Bed for Electric Grid Impact Analytics with Compromised Grid Edge Communications.” Submitted to *CSEE Journal of Power and Energy Systems*.

Cai, M., X. Fang, and A. Florita. 2022. “A Medium-/Low-Voltage Joint State Estimator Through Linear Uncertainty Propagation.” *IEEE Power & Energy Society General Meeting (PESGM)*: 1–5. https://doi.org/10.1109/PESGM48719.2022.9917189.

Cruickshank, R., N. Metts, P. Schauer, and C. Snyder. 2020. “Gridmetrics Data Provide Insights and Improve Situational Awareness of the Electric Power Grid.” Presented at the Society of Cable Telecom Engineers, Cable-Tec Expo, Orlando, Florida, October 2020.

Gridmetrics. 2021. “Houston, We Have a Problem.” Cable Television Laboratories, Inc. https://gridmetrics.io/resources/.

Gridmetrics. 2022. “PENS Outage-Reliability-Stability-Quality (PENS-ORSQ) and GridCON: The PENS Power Intelligence Indexes.” Cable Television Laboratories, Inc.

Marković, M., A. Florita, and B.-M. Hodge. 2021. “Matrix Completion for Improved Observability in Low-Voltage Distribution Grids.” *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Achen, Germany, October 2021. https://doi.org/10.1109/SmartGridComm51999.2021.9632334.

Marković, M., A. Sajadi, A. Florita, R. Cruickshank, and B.-M. Hodge. 2021. “Voltage Estimation in Low-Voltage Distribution Grids with Distributed Energy Resources.” *IEEE Transactions on Sustainable Energy* 12 (3): 1640–50. https://doi.org/10.1109/TSTE.2021.3060546.

Wang, W., X. Fang, and A. Florita. 2021. “Impact of DER Communication Delay in AGC: Cyber-Physical Dynamic Co-simulation.” *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*: 2616–20. https://doi.org/10.1109/PVSC43889.2021.9518779.

Yang, R., M. Ingram, and M. Cai. 2022. “Situational Awareness of Grid Anomalies (SAGA).” Presented at the Israel-U.S. Initiative on Cybersecurity Research and Development for Energy (ICRDE), October 18, 2022. https://www.nrel.gov/docs/fy23osti/84465.pdf.

Appendix A1: Using the SAGA APIv3

Gridmetrics calculates a number of useful voltage performance metrics for every sensor on a daily basis. In particular, Gridmetrics offers the Power Event Notification System (PENS) Outage Index (POI), the PENS Reliability Index (PRI), the PENS Stability Index (PSI), and the PENS Quality Index (PQI). Each index measures a different aspect of the voltage performance as measured by a sensor. Gridmetrics PENS indices can be selected for arbitrary collections of sensors and time spans.

For the purposes of the Situational Awareness of Grid Anomalies (SAGA) project, specific portfolios of sensors might be of particular interest, e.g., sensors related to critical infrastructure, such as hospitals. The procedure for collecting and reporting these site-specific data is known as the Fetch Region Average and is described as follows:

1. Start with a set of sensors, with each sensor belonging to a subset, e.g., the set of sensors within 1 km of an airfield or a hospital—in this case, each sensor subset consists of the sensors within 1 km of a particular hospital or an airfield. Particular geographies can also be selected as sites, such as counties or U.S. National Grid cells.
2. Collect the index data from the set of sensors in (1) over a particular time range, which can span multiple days.
3. Collect the index data from (2) and average it across the sensors within each subset. For example, if a particular hospital has five sensors within 1 km, then the particular index value (from POI, PRI, PSI, and/or PQI) is calculated as the average from those five sensors for the total number of days that are included in the time range. If the time range is 3 days, then there are 5 x 3 = 15 measurements to be averaged for each index for this case.

In the current implementation, the indices are calculated daily, but in future work they could be calculated on much shorter time frames, such as by the hour.

- Fetch Region Average accepts: limit, offset, and additionally:
- Fetch Region Average requires:
  - Date—starting date to query from
  - Days—number of days to gather data for calculations
  - Region_type—one of: state, county, census_block_fips_code, zip_code, tract_code, usng_spatial_address (formatted as shown in metadata)
  - Region_code—a valid region code for corresponding region_type field.

The final SAGA application programing interface is included in Appendix A2.

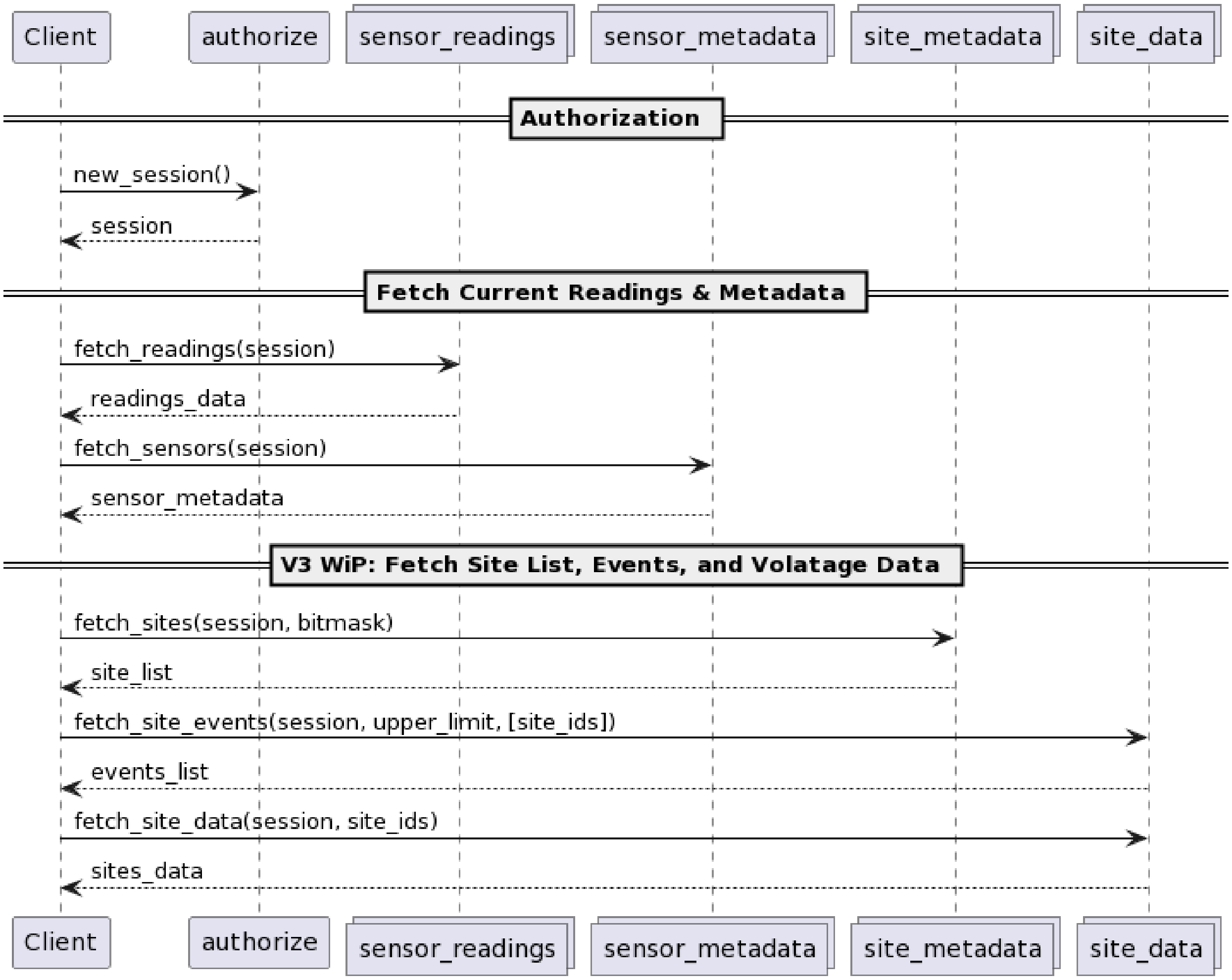
Appendix A2: SAGA APIv3

Begins on the next page.

SAGA APIv3

Overview

As part of the Situational Awareness of Grid Anomalies (SAGA) project, the National Renewable Energy Laboratory (NREL) worked with Cable Television Laboratories, Inc. (CableLabs) and subcontractor teams to develop the ability to query sensor data in total or by predefined groups and sites. The following table shows an overview of the use of the application programming interface (API).



APIv3—Monitoring APIs

The v3 interface has calls for opening an authenticated session and then querying for sensor metadata and the most-recent sensor readings. There are 3 API calls:

- authorize
- fetch_readings
- fetch_sensors

Name: authorize	**Purpose:** To get a current token for access to all other API calls.	

REST call (POST):
Endpoint (auth_url): https://admin-e8b207ca-eval-prod.apigee.net/oauth/client_credential/accesstoken?grant_type=client_credentials
	INPUTS:
	`{`
	`"client_id": id,`
	`"client_secret": secret`
	`}`
	which is placed in the request data field
		OUTPUTS:
		`{`
		`"access_token": token,`
		`"api_product_list_json":`
		`api_product_list,`
		`"status": status`
		`}`
		The status should be “approved” for a successful call.
		The token is used in subsequent calls.

Python lib call:
new_session()
Throws exception on authorization failure/error
	INPUTS:
	none:
	client_id and client_secret are hard-coded in Python lib
		OUTPUTS:
		session, which is of the form:
		`{"access_token": token,`
		`"client_name": client_name,`
		`"api_product_list": api_product_list}`
		(for use in subsequent query calls)

Name: **fetch_readings**	**Purpose:** Fetch 5-min reading data for all SAGA-allocated sensors. **NOTE:** Sensor limits: 50,000 sensors 10/1/19 until 9/30/20, 75,000 sensors 10/1/20 until 9/30/21, 100,000 sensors 10/1/21 until 9/30/22	

REST call (GET)
endpoint (fetch_sensor_readings): https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_readings
	INPUTS:
	Required:
	session[`'access_token'`]
	which is placed in the Authorization/Bearer field of the request header
	Optional*:
	limit (int)
	offset (int)
	* One or more constraints are required to keep payload <10 MB. Full metadata is 22 MB+.
		OUTPUTS:
		`{ "data": data }`
		The data array has the format:
		data: [{
		'batchTime': 1629221700000000,
		'block_bbox': [-104.926155, 39.529121, -104.923805, 39.532481],
		'census_block_fips_code': '080350141371010',
		'census_tract_fips_code': '08035014137',
		'city': 'Lone Tree',
		'county': 'Douglas',
		'county_fips_code': '08035',
		'inputVoltage': 117.6,
		'inverterStatus': '1.0',
		'latitude': '39.532444',
		'longitude': '-104.92455',
		'pollTime': 1629221541000000,
		'sensor_id': 1410686,
		'state': 'Colorado',
		'state_code': 'CO',
		'state_fips_code': '08',
		'state_name': 'Colorado',
		'usng_spatial_address': '13S ED 0648 7586',
		'zip_code': '80124'
		}]

Python lib call: TBD	**INPUTS:** **session**	**OUTPUTS:** **data** array (See the REST call above for format.)

Name: fetch_sensor_metadata	**Purpose**: Fetch metadata info. for all SAGA-allocated sensors.	

REST call (GET)
 	
endpoint (fetch_sensor_metadata):
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_metadata
	INPUTS:
	
	Required:
	session['access_token']
	
	which is placed in the Authorization field, concatenated with “Bearer” prefix in request header
	
	Optional*:
	site_ids
	comma-separated string of site IDs
	bitmask
	6-digit binary
	limit (int)
	offset (int)
	
	* One or more constraints are required to keep payload <10 MB. Full metadata is 22 MB+.
		OUTPUTS:
		
		`{ "data": metadata }`
		
		The metadata array has the format:
		
		metadata: [{
		"sensor_id": 1374226,
		"site_id": "H1",
		"latitude": 36.6760482788,
		"longitude": -121.6606292725,
		"county_code": 6053,
		"state_code": 6,
		"usng_spatial_address": "10S FF 19 59",
		},
		]

Python lib call: TBD	**INPUTS:** **session**	**OUTPUTS:** **metadata** array (See the REST call above for format.)

APIv3—Group APIs

The v3 interface has calls for fetching sites (with filtering by group) along with fetching site events and data:

- Five groups have been defined by SAGA (Gridmetrics-data-collection.saga_sensors.saga_v3_groups):
  - California
  - Colorado
  - Houston metro
  - Hospitals
  - Airports.
- The sensor metadata table now includes an additional column for sites (Booleans in the form of a single combined bitmask):
  - Canonical sensor ID
  - Site ID
  - Latitude
  - Longitude
  - County Federal Information Processing Standard (FIPS) code
  - State FIPS code
  - U.S. National Grid coordinate (1-km precision)
  - **Colorado (Boolean indicator for presence in Colorado)**
  - **California (Boolean indicator for presence in California)**
  - **Houston (Boolean indicator for presence in Houston)**
  - **Hospital (Boolean indicator for presence in/near a hospital)**
  - **Airport (Boolean indicator for presence in/near an airport).**
- The enumerated states are defined as:
  - Nominal, high (6%–25%),
  - Low (6%–25%),
  - Extreme high (>25%),
  - Extreme low (<6%)
  - Outage.

There are three API calls in APIv3:

- fetch_sites
- fetch_site_events
- fetch_site_data

Name: fetch_sites	**Purpose:** Fetch sensors for a site or list of sites (aka Groups).	

REST call (GET)
endpoint (fetch_sites):
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sites
	INPUTS:
	Required:
	session['access_token']
	which is placed in the Authorization field of the request header
	e.g., "Bearer 12345"
	Optional*:
	site_ids
	e.g., site_ids=H1,HOU156
	bitmask
	6-bit bitmask format: Incl/Excl,CO,CA,Hou,Airport, Hospital
	First bit: inclusive = 0, exclusive = 1. Inclusive ignores 0s and combines 1s; Exclusive selects 1s AND NOT 0s, excluding overlaps.
	Inclusive (0):
	e.g., bitmask = 01001
	Returns all of: CO AND H
	Exclusive (1):
	e.g., bitmask = 11001
	Returns: CO AND NOT H
	Logic:
	(CO || CA || HOU) && (H || A)
	limit (int)
	offset (int)
	* One or more constraints are required to keep payload <10 mb.
		OUTPUTS:
		`{`
		`"records": number_data_records,`
		`"query-time-me": query-time-ms,`
		`"data": data`
		`}`
		The data array has the format:
		data:
		[{
		"v": 115.2,
		"v_ref": 115.1999969482,
		"sensor_id": 1374226,
		"site_id": "H1",
		"latitude": 36.6760482788,
		"longitude": -121.6606292725,
		"county_code": 6053,
		"state_code": 6,
		"usng_spatial_address": "10S FF 19 59",
		"colorado": false,
		"california": true,
		"houston": false,
		"hospital": true,
		"airport": false,
		"pollTime": 1648943875000
		}...]

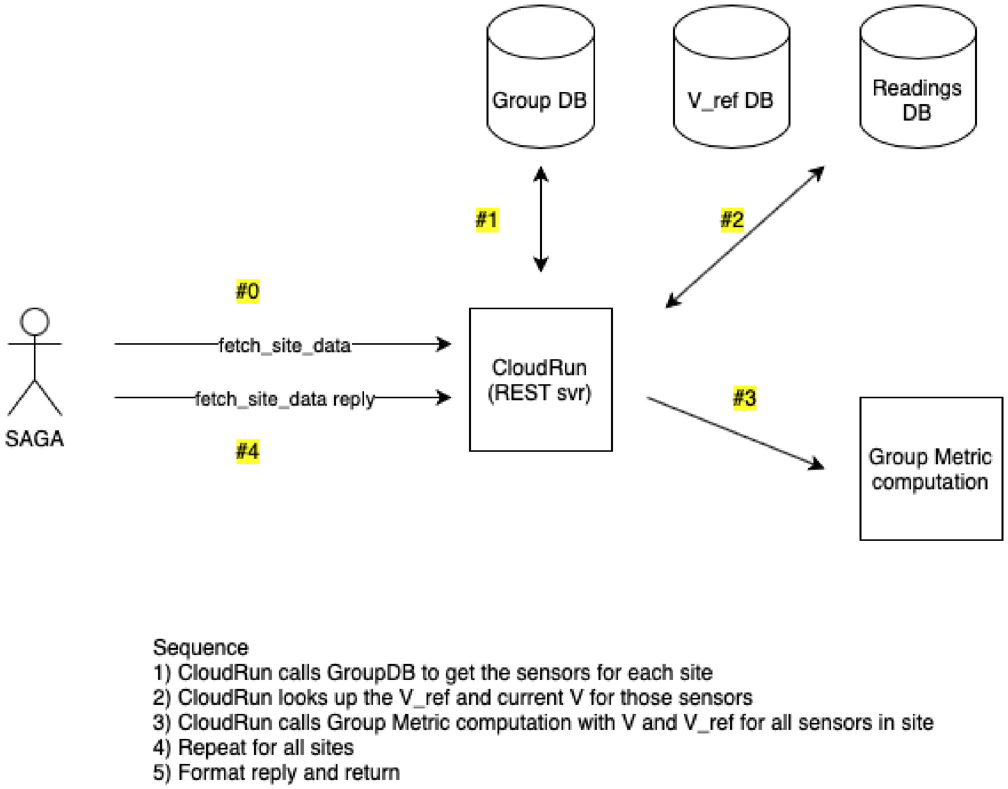
Name: fetch_sites	**Purpose:** Fetch sensors for a site or list of sites (aka Groups).	
Python lib call: TBD	**INPUTS:** **session** **site_ids** **bitmask** (See the REST call above for details of this input.)	**OUTPUTS:** **data** array (See the REST call above for format.)

Name: fetch_site_events	**Purpose:** Fetch event info. for sensor at the site.	
REST call (GET) **endpoint (fetch_site_events):** https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_events	**INPUTS:** **Required:** **session[`'access_token'`]** which is placed in the Authorization/Bearer field of the request header **lower (default .95)** Baseline metric undervoltage threshold Return all entries with metric <threshold; threshold range is from 0–1. **upper (default .95)** Baseline metric overvoltage threshold Return all entries with metric <threshold; threshold range is from 0–1. **period (default 1)** Lookback period (in hours) Max is 4?? GORP **Optional*:** **limit** **offset** **bitmask**	**OUTPUTS:** `{ "data": data }` The data array has the format: **data:** [{'baseline_metric_overvoltage': 0.8726945683451612, 'baseline_metric_undervoltage': 0.8726945683451612, 'sensor_id': 197011, 'v': [1.000000012975161, 1.000000012975161, ...]}, ...] “v” field is v/v_ref array over period at 5-min intervals.

Name: fetch_site_events	**Purpose:** Fetch event info. for sensor at the site.	
	site_ids * One or more constraints may be required to keep payload <10 mb.	
Python lib call: TBD	**INPUTS:** **session** **lower** Baseline metric undervoltage threshold Return all entries with metric <threshold; threshold range is from 0–1. **upper** Baseline metric overvoltage threshold Return all entries with metric >threshold; threshold range is from 0–1. **period** Lookback period (in hours) Max is 4?? GORP	**OUTPUTS:** **data** array (See the REST call above for format.)

Name: fetch_site_data	**Purpose:** Fetch data for sensors for sites. Returns data for SAGA-allocated sensors.	
REST call (GET) **endpoint (fetch_site_data):** https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_data	**INPUTS:** **Required**: **session[**`'access_token'`**]** which is placed in the Authorization/Bearer field of the request header **Optional*:** **limit** **offset** **bitmask** **site_ids** * One or more constraints may be required to keep payload <10 mb.	**OUTPUTS:** `{ "data": data }` The data array has the format: **data:** [{“sensor_id”:”sensor_id1”, “site_id”:”site_id1”, “V_ref”:120.1, “V”:121.3}, {sensor_id”:”sensor_id2”, “site_id”:”site_id1”, “V_ref”:120.5, “V”:122.3} ...]
Python lib call: TBD	**INPUTS:** **session** **Site_id_list** **bitmask** (See the REST call above for details of this input.)	**OUTPUTS:** **data** array (See the REST call above for format.)

Gridmetrics internal architecture for APIv3:



Sequence
1) CloudRun calls GroupDB to get the sensors for each site
2) CloudRun looks up the V_ref and current V for those sensors
3) CloudRun calls Group Metric computation with V and V_ref for all sensors in site
4) Repeat for all sites
5) Format reply and return

Note:

- Gridmetrics will return V_ref and V; caller will compute states.
- Suggested states: nominal, high (6%–25%), low (6%–25%), extreme high (>25%), extreme low (<6%%), outage (V = 0)
- Returning V/V_ref instead of states allows the caller flexibility to change state definitions to suit their needs.

Summary and References

Authorization:
All calls other than the authorization call to fetch a token require the authorization token to be passed in via the bearer field. See the first table above for more information on authorization. This call is POST; all others are GET.
URL: https://admin-e8b207ca-eval-prod.apigee.net/oauth/client_credential/accesstoken?grant_type=client_credentials

All parameters to these calls are optional (with hopefully sane defaults), but in cases where the return payload would exceed 10 MB, some constraints must be supplied to narrow or paginate the results.

Fetch sensor readings:
Accepts two parameters: *limit* (e.g., 10), *offset* (e.g., 100)

Limit and offset can be combined to limit results output and/or implement pagination.
URL: https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_readings

Example with query string parameters: https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_readings?limit=1&offset=100

Fetch sensor metadata, fetch site data, and fetch sites:
Accept *limit*, *offset* parameters defined above and additionally:

Bitmask: Six binary digits representing incl./excl., CO, CA, HOU, H, A, respectively.
First bit: inclusive = 0 or exclusive = 1. Inclusive ignores 0s and combines 1s; Exclusive selects 1s AND NOT 0s, excluding overlaps.

For example: inclusive (0): bitmask = 01001
Returns all of: CO AND H

Exclusive (1): bitmask = 11001
Returns: CO AND NOT H

site_urls: A comma-separated list of site IDs (e.g., H1, HOU156)

URLs:
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_data
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sites
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_metadata

Examples with query string parameters:
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_data?bitmask=001001&limit=10
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_sensor_metadata?site_ids=H1

Fetch site events:
Accepts *limit*, *offset, bitmask, site_ids* parameters defined above, and additionally:

- *lower*: baseline metric undervoltage threshold
- *upper*: baseline metric overvoltage threshold
- *period*: period over which to observer v/v_ref readings in hours (1 default, 4 max).

URL: https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_events

Example with query string parameters:
https://admin-e8b207ca-eval-prod.apigee.net/saga-py-fetch/fetch_site_events?bitmask=010000&limit=100&offset=200&lower=.90&upper=.90&period=2

Appendix B: Abridged SAGA Ops Proposal

Begins on the next page.

SAGA Operations

The next phase for the Situational Awareness of Grid Anomalies (SAGA) project

8 September 2022

DRAFT SOW

Kenny Gruchalla
Graham Johnson
Maurice Martin

NREL

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

Contract No. DE-AC36-08GO28308

NOTICE

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content

Table of Contents

1 Introduction: Situational Awareness of Grid Anomalies 56
2 Project Objectives 57
3 Scope of Work 58
4 Tasks 63
5 Deliverables 66

1 Introduction: Situational Awareness of Grid Anomalies

The ongoing Situational Awareness of Grid Anomalies (SAGA) project seeks to expand grid situational awareness by integrating data from sensors within the cable television broadband network infrastructure with capabilities for visualization and analysis. The sensor data is supplied by Gridmetrics Inc., a wholly owned subsidiary of Cable Television Laboratories (CableLabs), the research and development arm of the cable broadband industry. Gridmetrics has agreements with CableLabs' members to supply data from sensors that are attached to the cable network, via cable modems embedded in networked uninterruptable power supplies (UPS).

While there are existing systems in place to support grid situational awareness using SCADA data generated by the utility itself (this implies that SCADA at distribution system is ubiquitous it's not. There's a complimentary angle here as well), the use of sensor data provided by Gridmetrics offers unique advantages over utility-generated data.

- **Out-of-band.** The Gridmetrics data is collected by sensors not operated by the utility, providing "ground truth" for SCADA and AMI telemetry.
- **Higher spatial fidelity.** Data from Gridmetrics provides higher spatial resolution, as it is collected from sensors at the neighborhood level, when compared to traditional SCADA telemetry, which is gathered at the feeder level.
- **Higher temporal fidelity.** Gridmetrics provides up to 1-minute sensor data, offering more frequent readings than SCADA systems, or even most AMI systems.
- **Cross-service territories.** Data from Gridmetrics spans boundaries of utility service territories. This directly supports monitoring of the grid at city, county, state, regional, and national levels.
- **Independent from utility communications infrastructure.** Disruptions to the utilities' own networks will not impact Gridmetrics data transmission.
- **Independent from utility power.** Sensors are battery-backed and continue providing measurements during power outages.

SAGA has demonstrated the utility of capturing, exploring, and integrating the nation-wide out-of-band Cable Network sensor data, provided by Gridmetrics', for situational awareness settings of power systems. At a proof-of-concept level, the project explored the needs of different potential user groups – including operators, utilities, and government agencies, along with novel approaches for sensor data visualization and analysis. As such, an extension of the prototype developments to operational capabilities provides a rich path to supporting enhanced distribution grid situational awareness – enabling operators and agencies to leverage scalable data services for accessing historical and near real-time data and a platform to visually explore these datasets. Together, these capabilities can provide users the ability to view geospatial distributions of sensor states, identify sensor group behaviors, explore sensor timeseries, uncover temporal patterns, and generate summary reports. Ultimately, scalable data services and visualization capabilities of Gridmetrics' neighborhood-level 1-minute cable sensor data will provide an extensible basis for enhanced real-time grid awareness – forming a solid foundation for the continued developments of novel grid anomaly detection and state estimation algorithms, in distribution systems that currently have restricted observability.

The proposed follow-on project, SAGA Operations (SAGA Ops) will extend and scale the proof-of-concept capabilities developed by the current project and ready them for operational support of CESER's mission.

2 Project Objectives

The main objective of SAGA Ops is to provide CESER with the ability to leverage the out-of-band cable sensor data from Gridmetrics at scale, and directly support CESER's responsibilities regarding grid security and resilience. At a high-level, SAGA Ops will construct two categories of support: (1) Real-time monitoring of sensor data, and (2) Reporting of sensor data over defined periods of time.

SAGA Ops will empower users to incorporate cable sensor data for:

Real Time Monitoring

- Leverage 1-minute data from cable sensors distributed across the continental U.S.
- Geospatially explore sensor distributions, visualize grid performance, compare to historical trends, and inspect associated sensor timeseries.
- Analyze sensor states during fluctuating voltages, over/under-voltages, and outage events with indications of the scale.
- Define custom sensor groups to monitor behavioral baselines of facilities or regions of interest (e.g., hospitals, transportation hubs, proxies to critical infrastructure, counties, etc.).
- Define custom event scenarios to receive notifications (e.g., sensors operating outside of a threshold, or length of outage time). Apply sensor data filters to include or exclude certain event types (e.g., power deviation, outages, etc.).
- Review historical sensor behavior and create sensor, site, or group baselines.

Creation of Event Reports

- Leverage interactive visual analytics tools to isolate data sets of interest, compare behaviors to baselines or reference sets, and uncover deep insights to system operation under evolving conditions.
- Analyze system behaviors, investigate custom event notifications, and generate incident reports for geographic areas and/or custom sensor groups over a specific period.
- Export event reports, data references, visualizations, and supporting analyses.

Additionally, these goals are supported in part by Gridmetrics' efforts to improve the quality and scale of data it can provide:

- **Increasing the number of sensors contributing data.** When the SAGA project began in 2019, Gridmetrics could supply data from about 6,000 sensors. Today, Gridmetrics currently has access on the order of 300,000 sensors. The maximum possible number of U.S.-based sensors is 650,000 which Gridmetrics plans to achieve by 2023.

- **Increasing the frequency of sensor readings.** When the SAGA project began, the sensors where supplying readings nine times per day. With the helpful support of SAGA funding, currently, 90% of sensors are reporting data in one-minute intervals. SAGA Ops support will enable Gridmetrics to increase this to nearly 100%. (Note: Gridmetrics and the SAGA team pioneered a standard and developed hundreds of prototypes for a next generation of sensor technology that increases the voltage precision and sampling rate even more, but these next generation sensors have not yet been deployed in bulk and are outside the scope of this proposal).

- **Continued cleaning, storage, pre-processing, and distribution of expanded sensor data.** A major benefit to the cable sensor data provided by Gridmetrics is its overall quality and ease of access. While other data sources require vast web-scraping operations with data cleaning pipelines, Gridmetrics will work to continue supplying accurate and clean raw sensor data, in addition to derived sensor data metrics, across subscription sets.

To successfully enable these capabilities, we propose to extend the prototype software systems that were designed and developed for SAGA into operations-ready capabilities. This extension has four major areas of effort – (1) Design iterations with CESER stakeholders. (2) Refinement, scaling, and deployment of the Core Data Services and an Event Identification Platform. (3) Design, development, and integration of advanced visual analytics components. (4) Design, development, and deployment of a web-based, authenticated, Operations Platform. Additionally, we will include phases of testing of software components and their integration, initial NREL deployment monitoring, and comprehensive documentation of respective software systems. These phases support a rigorous inspection of an operational-ready software system and successfully position it for continued feature enhancement with program evolution. Furthermore, to support control of software access and data sources, we will create and implement both Software and Data Management Plans – engaging with stakeholders to define necessary requirements for data storage and access, as well as scoping software extensibility for follow-on efforts.

At the end of the project, SAGA Ops will be sufficiently robust to support CESER’s role in monitoring the state of the grid for cybersecurity and emergency response needs. In particular, the Core Data Services, visual analytics capabilities, and operational interface will enable users to leverage scalable data interfaces and interactively analyze areas for situations of interest. Additionally, component documentation will be created to include a clear path for managing data, software development and maintenance, and capability upgrading.

3 Scope of Work

To provide an operational capability that supports desired CESER missions, development will be driven through an iterative user-centered design approach. In this setting, we will regularly engage with CESER stakeholders during the projects’ first two quarters. This will enable us to capture and map all desired use-cases, focus areas, and associated workflows with the potential features and components that can be developed – as these areas will directly motivate the corresponding data service architectures, visualization components, design of the operations platform, and overall functionality.

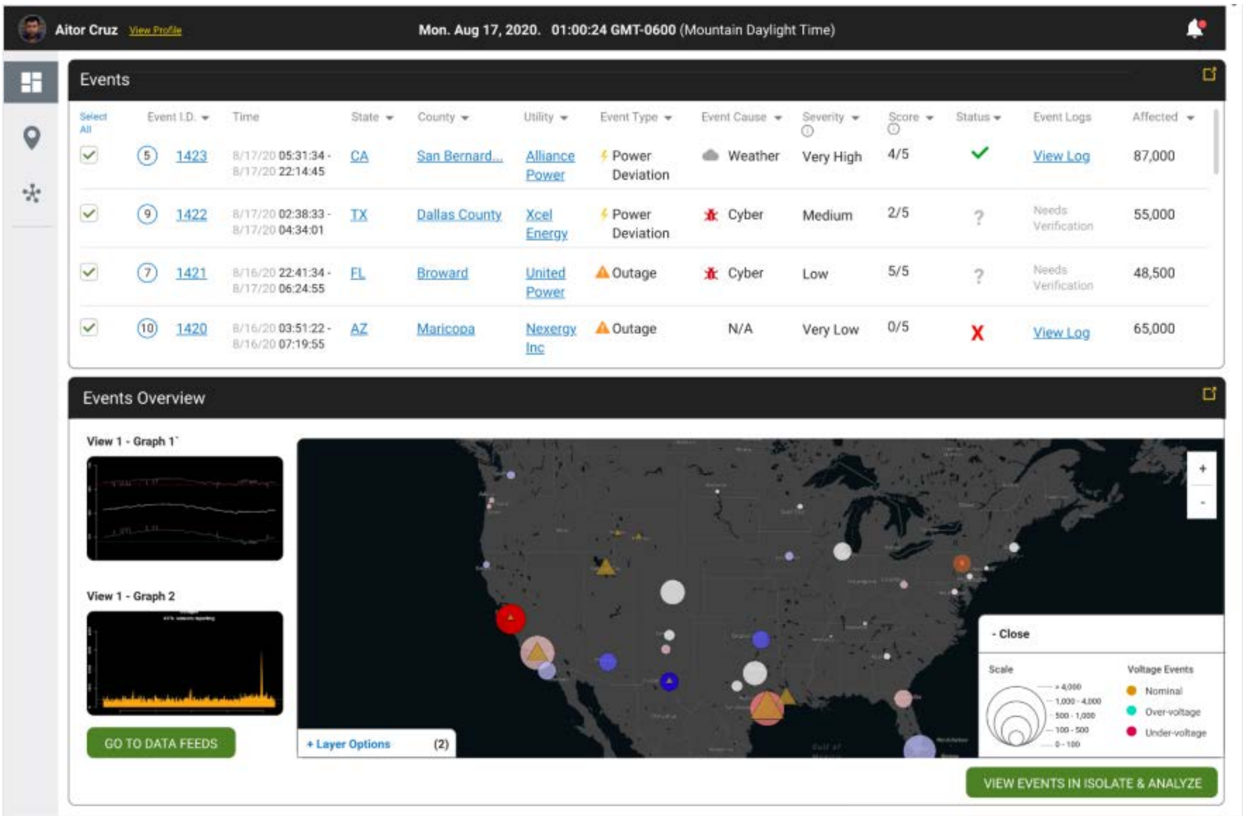
From all captured potential use-cases, we will evaluate the desired priorities and determine an initial focus within the proposed budget and scope. The additional stakeholder use-cases, features, and focuses that are not an immediate priority, or out of the scope of the current budget, will be captured and planned to be addressed during software extension efforts. For example, the project's core data functions can ultimately support full-scaled analysis efforts across Distribution, Service Operator, Interconnect, and National grid scales. However, to provide an extensible and effective product that can be continually upgraded, an initial focus will need to be adequately defined. As a default, our focus will position the initial version of the operational platform's functionality to support Distribution system operations.

As the informed designs are developed and implemented, we will confirm the solution's effectiveness and directions with facilitated feedback from stakeholders. This will enable us to provide the flexibility to capture expanding feature requests with program evolution while maintaining an extensible core capability.

To this end, the software components developed in the initial SAGA project will be extended and scaled beyond prototypes and integrated into an operational architecture, with any new required component designs or extensions requested being created through stakeholder engagement. To this end, we will focus on four areas of initial software development – Core Data Services, Event Identification Platform, Visual Analytics components, and an authenticated Operational Platform – followed by stages of integration testing and final deployment monitoring. These areas of effort are outlined below. As with the current SAGA project, NREL will continue its partnership with Gridmetrics on SAGA Ops.

Gridmetrics' Efforts:

- Increase number of available sensors to the include entire partner base.
- Negotiate with members who currently do not contribute data.
- Increase all available sensor reporting intervals to 1-minute rates.
- Develop sensor data metrics/indices for inclusion within distributed data packages.
- Enhance sensor data access and interfaces with the NREL Core Data Services.



Aitor Cruz View Profile

Mon. Aug 17, 2020. 01:00:24 GMT-0600 (Mountain Daylight Time)

Events

Select All	Event I.D.	Time	State	County	Utility	Event Type	Event Cause	Severity	Score	Status	Event Logs	Affected
☑	(5) 1423	8/17/20 05:31:34 - 8/17/20 22:14:45	CA	San Bernard...	Alliance Power	Power Deviation	Weather	Very High	4/5	✓	View Log	87,000
☑	(9) 1422	8/17/20 02:38:33 - 8/17/20 04:34:01	TX	Dallas County	Xcel Energy	Power Deviation	Cyber	Medium	2/5	?	Needs Verification	55,000
☑	(7) 1421	8/16/20 22:41:34 - 8/17/20 06:24:55	FL	Broward	United Power	Outage	Cyber	Low	5/5	?	Needs Verification	48,500
☑	(10) 1420	8/16/20 03:51:22 - 8/16/20 07:19:55	AZ	Maricopa	Nexergy Inc	Outage	N/A	Very Low	0/5	X	View Log	65,000

Events Overview

Figure 1. Example design of an operational interface from current SAGA project. Designs can include abilities for users to navigate geospatial overviews of real-time sensor distributions, investigate current and historical timeseries, interactively explore data sets with visual analytics capabilities, and create reports of periods of interest.

NREL Efforts:

Core Data Services

We will refine and expand three of the existing containerized data services and add an additional events data service. The services will be composed into a well-defined core data system of containers that interacts with one another, for deployment and orchestration in a scalable environment. This effort is supported by active collaboration and integrated developments with Gridmetrics' API development team. Each containerized service will be tested along with integration testing the composition.

Metadata Service

API to query, store, and expose subscribed Gridmetrics' sensor metadata. We will expand the service to include an updated set of available Gridmetrics' sensor, site, and group metadata. Metadata includes sensor IDs, location data, group inclusion, and group sites' sensor lists.

- We will update the service API's filtering and table transformation capabilities.

Streaming Data Service

Kafka stream processing platform that includes topics for subscription and publishing of Gridmetrics' 1-minute sensor data.

- We will update and stress-test this service to handle Gridmetrics' full-set of available sensor data.
- We will grow the service to provide dynamic creation of Kafka topics to enable users to subscribe to raw data feeds of selected sensors of interest.
  - For example, this service will interface with the event identification platform which will subscribe to custom topics/sensor groups and process timeseries with custom logic, which in turn can emit data to other Kafka topics.
- We will expand the service's interface to the Historical Data Service (described below), to provide desired pre-processing before storage in Druid.

Historical Data Service

API to the Druid data store containing Gridmetrics' sensor timeseries, site and group indices, and derived metrics that are piped directly from the Streaming Data Service.

- We will extend the API's functionality to enable users to create custom advanced Druid-SQL queries across sensor timeseries data.
- We will test and design Druid's data segmenting to be optimized for analytic querying over large batches of timeseries data.

Events Data Service

We will define and develop this new service API to handle storage and retrieval of events that were flagged by prototypical event identification models running on the event identification platform.

- Using feedback from stakeholder engagement, we will design, develop, and deploy this service.
  - The API schema, routing, and filtering will be defined according to support cases.

Event Identification Platform

Based on iterations and engagement with stakeholders, we will design, develop, and deploy an Event Identification Platform that will host custom models that connect to topics within the Streaming Data Service, process sensor data, and produces conditional events that are subsequently stored in the Events Data Service.

- Platform development will include the creation and deployment of an additional web service that will host production-ready event identification models (e.g., outages, out of range values, or future machine learning event predictions).
- Initial event identification models will be designed, developed, and included to allow users to define sensor sets of interest, nominal operating ranges, and various outage conditions from which events can be created and logged.

Visual Analytics

We will refine and expand the prototype developed SAGA visualization components to further support analysis and pattern discovery interactions with the geospatial timeseries data, supporting the canonical visual analytics tasks of:

- *Overview*: Provide an overview of the complete sensor collection and desired reference infrastructure.
- *Zoom*: Zoom into spatial and temporal regions of interest.

- *Filter*: Filter out individual or groups of sensors/timeseries.
- *Details-on-demand*: Select individual or groups sensors and time periods to investigate the timeseries traces.
- *Relate*: View and compare timeseries behaviors between multiple groups of sensors.
- *History*: Keep a history of actions to support replay and progressive refinement.
- *Extract*: Allow extraction of sub-collections and the query of parameters.

Operational Interface

Based on iterative engagement with CESER stakeholders, we will design, develop, and deploy an authenticated front-end web client – to enable users to explore data sources, observe overviews of system states, analyze events, and create reports / records of findings. This work will be supported by the following tasks:

- Design of user interface (UI) and user experience (UX) pairings, tailored to stakeholders use-cases.
- User access control and authentication capabilities.
  - Definition of user types and roles (e.g., Operator, Analyst, Admin etc.), associated views, interactions, and capabilities.
- Design and integration of all visual analytics components and geospatial visualizations.
- Integration with the Core Data Service APIs and creation of corresponding data views.
- Creation of data views and user interactions with logs from identified events.
- Development of capability for users to create reports of session details (e.g., sensor sets, timeseries overviews, included events, and associated visualizations).

Unit and Integration Testing

As the software systems that are being developed are ultimately intended for operational environments, we will rigorously validate software components via well-defined software testing steps. Our teams follow software engineering best-practices throughout the development life cycle. This includes a focused phase of thorough integration testing, to ensure all interdependent components (e.g., back-end services and front-end clients) behave accordingly in nominal settings, while providing robust error-handling.

These efforts will directly support the project's ability to successfully function under required operating conditions, establishing trustworthiness in component behavior and end-to-end functionality. Moreover, the findings during the respective testing phases will provide deep insights into any potential logic issues and performance bottlenecks that can be addressed early and support the system's continued extensibility.

To that end, the following testing strategies are outline below.

- *Unit Testing Strategy:*
  - Proper definition of software units (e.g., functions)
  - Arrangement of independent tests
  - Tracking of pass/fail assessments and resolutions
- *Integration Testing Strategy:*
  - Identification of interdependent software components
  - Creation of representative scenarios for nominal and erroneous situations
  - Documentation of scenario and component performances

Deployment and Operational Monitoring

The outcome of this project are software systems that are ready for operational environments. For the scope of this project, the deployment of developed software systems will reside on NREL infrastructure. We will document the environment architecture, necessary constraints, and any required support systems – along with build and deployment processes – to support the availability of this capability for future porting to stakeholder ecosystems. The following outline the task summaries for this period:

- Deployment of Core Data Services and Event Identification Platform onto NREL systems
  - Includes NREL's existing data systems: Druid and Kafka Clusters
- Deployment of Operational Interface and initializing user access
- Monitoring of service and platform health and resilience
  - Creation of reports of system utilization and data accumulation rates. This will inform continued operations costs, environment considerations, and overall system extension.
- Management of user access and tracking any identified issues

NREL and Gridmetrics joint efforts include:

- Refining the Gridmetrics' API and planning its evolution with the Core Data Services.
- Providing ability for users to explore available sensor data sets/groups and subscribe to chosen sensors, up to the agreed upon maximum.
- Advance existing standards and create new standards, as needed.

4 Tasks

To successfully deliver an extensible and robust SAGA Ops software platform, the project's components and phases will be managed in terms of following Tasks. Foundationally, the initial thrust of the project is focused on CESER stakeholder Design Iterations, as these crucially determine the project's focus and scope around a particular perspective of the Gridmetrics' data. The subsequent objectives, design details, and development tasks will be tracked and planned accordingly throughout the 18-month period of performance.

Task 1. CESER design iterations

This task is focused on capturing and mapping all desired stakeholder use-cases, focus areas, and associated workflows with the potential features and components that can be developed. At the end of the design iterations, we will have successfully captured focused use-cases that support CESER stakeholders. These will feed into the design of the data service architectures, visualization components, operations platform, and overall functionality. The focus will be refined iteratively, with NREL presenting updated options and designs based on CESER feedback through several review cycles. This process will continue until DOE and NREL reach a mutually satisfactory focus and initial design.

Task 2. Identify sensor subscription sets for initial operational investigation

Based on the set of available sensors from Gridmetrics', this task will identify the sets of sensors that CESER stakeholders wish to subscribe to – up to the maximum agreed upon amount (i.e., 300,000 at the initialization of this proposal). This selection process will initially take place during the CESER design iterations. These sensors can be chosen from geographical distributions with the option of identifying availability of sensor assets in proximity to identified areas of interest. Additionally, this task will identify a selection process that supports updating and/or modifying the subscription set to swap-out sensor subscription selections across the available catalog.

Task 3. Refinement and expansion of SAGA Core Data Services

Refinement of the three existing containerized data services (i.e., Metadata, Historical, and Streaming Data Services), and addition of the Events Data Service. This Task is supported by active collaboration with the Gridmetrics' API development team.

Task 4. Development of Event Identification Platform

Based on iterations with CESER stakeholders, we will design and develop a platform that will host custom models that connect to the Streaming Data Service to process sensor data, produce conditional events from model logic, and subsequently store event data in the Events Data Service.

Task 5. Design and development of visual analytics components

Expansion of the prototypical visual analytic components develop in SAGA to further support desired interactions, analyses, and pattern discovery capabilities with sensor geospatial and timeseries data.

Task 6. Design and development of web-based Operations Interface

Based on the iterative engagements and feedback from CESER stakeholders, we will design and develop an authenticated front-end web client, that integrates with the Core Data Services, Event Identification Platform, and visual analytics components – to enable users to explore data sources, observe overviews of system states, analyze events, and create reports / records of findings.

Task 7. Integration testing of software components (i.e., Core Data Services, Event Identification Platform, Visual Analytics components, and Operations Platform)

Rigorous validation of software components, definitions of associated interfaces, and overall system integration testing via well-defined processes. These efforts will directly support the project's ability to successfully function under required operating conditions, establishing trustworthiness in component behavior and end-to-end functionality.

Task 8. Deployment and monitoring of SAGA Ops software components

Deployment of developed software systems onto NREL infrastructure. Documentation of the necessary environment, and any required support systems – along with build and deployment processes – to support the availability of this capability for any stakeholder ecosystems. We will provide active monitoring of deployed systems to manage user access, document system utilization and data storage metrics, and address any potential system issues.

Task 9. Implementation of Software and Data Management Plans

Descriptive documentation of the data being collected, analyzed, visualized, and stored – In conjunction with documentation of the software systems produced. Additionally, this will include detailed processes for maintenance, upgrades/feature requests, and future deployment considerations.

Task 10. Advance applicable standards

As part of a related project funded by the U.S. Department of Energy, Office of Electricity, Technology Commercialization Fund, the SAGA team recently created the new American National Standard, ANSI/SCTE 271 2021, Requirements for Power Sensing in Cable and Utility Networks. The standard specifies precision, sampling rate, and configuration requirements if vendors choose to measure and report voltage and/or current in hardware and software to enable advanced power sensing in cable and utility networks. The standard is generating much interest with utilities and broadband providers. Hundreds of SCTE 267 standard-compliant sensors have been built, dozens are in service, and deployment is accelerating in many areas across the U.S.

The opportunity to advance applicable standards includes further defining best practices for deploying, testing, and operationally integrating of the new SCTE 267-compliant sensors used by CableLabs member companies, Gridmetrics, and utilities. The team will continue to iterate on this standard as needed to achieve the goal of SAGA Ops. To that end, NREL renewed SCTE Standards membership and started discussion to extend SCTE 271 with "dot" standard(s), e.g., to provide an applications guide, testing/verification methodology, best practices, additional functionality, calibration, etc.

5 Deliverables

Aligned with the efforts and tasks presented in the previous section, the following items will be delivered.

D1. SAGA Ops design document.

This will summarize the content and results of Task 1. This will include definitions of terminology and desired stakeholder areas of interest and features. Additionally, this will include the agreed upon initial focus area, user interface/user-experience mockups, and supporting workflows for defined use-cases. This document will inform corresponding development effort details for the Core Data Services, Event Identification Platform, Visual Analytics components, and Operations Platform functionalities.

Delivery: End of Q3, FY2023

D2. Gridmetrics' Data Subscription Details

This deliverable is based on the results of Task 2 and provides documentation describing current sensor subscription details (e.g., geospatial coverage, counts, and any predefined sensor group distributions), subscription update process, and corresponding data terms.

Delivery: End of Q3, FY2023

D3. Core Data Services Software

Based on the results of Task 3 and Task 7, supporting code repositories and associated documentation describing the Core Data Services (e.g., Metadata Service, Streaming Service, Historical Service, Events Service) will be delivered.

Delivery: End of Q2, FY2024

D4. Event Identification Platform Software

Based on the results of Task 4 and Task 7 supporting code repositories and associated documentation describing the Event Identification Platform, and the initial model(s) deployed with it, will be delivered.

Delivery: End of Q2, FY2024

D5. Operations Platform Software

Based on the results of Task 6 and Task 7 supporting code repositories and associated documentation describing the web-based Operations Platform will be delivered.

Delivery: End of Q3, FY2024

D6. SAGA Ops review version (80% functionality).

This will include a demonstration of the system at 80% functionality, as a result of completing Tasks 3, 45, 5, 6, and initiating Task7. This demonstration will focus on the Operations Platform functionality, in conjunction with its interfaces to the Core Data Services and Event Platform. This version will be used to solicit feedback from CESER for the final version, before fully completing Task 7 and beginning the final project deployment with Task 8.

Delivery: End of Q2, FY2024

D7. SAGA Ops final version (100% functionality).

This will include a demonstration of the full system functionality, as a result of completing Task 7 and Task 8. Additionally, documentation of the deployment process and associated environment will be provided. All accompanying final software and associated documentation will be delivered with this version.

Delivery: End of Q3, FY2024

D8. Software Maintenance Plan.

This will identify dependencies that will drive future software updates to maintain functionality, as well as listing available improvements, extensibility/inclusion into other systems, and additional features that can be added in the future.

Delivery: End of Q3, FY2024

Appendix C: Letters from Utilities Supporting Continued SAGA Development

Begins on the next page.

DATE: June 28, 2022

Principal Investigator Michael Ingram
National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401

SUBJECT: Continued commercialization of NREL Situational Awareness of Grid Anomalies

Dear Mr. Ingram:

Holy Cross Energy supports continued commercialization NREL's DOE CEDS project "Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics: Near Real-Time Cyber-Physical Resiliency through Machine Learning". With the completion of the DOE TCF-funded ANSI/SCTE 271 standard and the availability of the next-generation sensors, the project is proceeding well. We recognize the value in developing and field-validating visual analytics that integrate cyber-physical data from Cable TV broadband power supplies with utility information systems to enhance electric distribution grid visibility and operational situational awareness, detecting patterns of operation indicative of cyber incidents. This work accelerates our efforts that align with the DOE goal to advance cyber resilient energy delivery systems that are designed, installed, operated and maintained to survive a cyber-incident while sustaining critical functions.

Holy Cross Energy is a member-owned electric cooperative serving 59,000 meters in rural Colorado.

As we transition SAGA and TCF project responsibilities within Holy Cross Energy, we continue supporting thess exciting efforts by:

- Participating as a member of the Technical Review Committee (TRC) for the projects,
- Providing industry experience-based guidance, directional support, and strategic direction to the projects.

Holy Cross Energy understands the value of this work and related follow-on projects, and looks forward to continued collaboration with the project team.

Sincerely,

Bob Farmer
Vice President, Information Technology
Holy Cross Energy

Utilities
electric · stormwater · wastewater · water
700 Wood Street
PO Box 580
Fort Collins, CO 80522

970.221.6700
970.221.6619 – fax
970.224.6003 – TDD
utilities@fcgov.com
fcgov.com/utilities

Date: 1/4/2023

ATTN: PI Michael Ingram
National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401

Re: Letter of Support to continue NREL's SAGA Research & Development

---

Dear Mr. Ingram:

The City of Fort Collins Utilities Light & Power recommends the U.S. Department of Energy (DOE) fund continued commercialization of NREL's Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics: Near Real-Time Cyber-Physical Resiliency through Machine Learning". We are founding members of the SAGA Technical Review Committee and have guided SAGA research, development, and deployment efforts dating back to 2017.

The City of Fort Collins Utilities Light & Power is a municipally owned electric utility that serves around 80,000 electric meters within the city limits of Fort Collins, Colorado. Our Vision is to sustainably provide for the energy needs of our community now and in the future with safe, renewable, reliable, resilient, and affordable electricity through a culture of innovation and operational excellence. We accomplish this through our employees' dedication to excellence, our environmental, economic, and social stewardship, building future flexible infrastructure, research, and innovation, as well as our data-driven decision making.

We believe SAGA and derivative efforts can deliver value by integrating cyber-physical data from Cable broadband networks with utility information systems to enhance electric distribution grid visibility and operational situational awareness, detecting patterns of operation indicative of cyber incidents. SAGA derivative efforts such as the Power Event Notification Systems (PENS), the GridCON Report, and the ANSI/SCTE 271 Standard, Requirements for Power Sensing in Cable and Utility Networks, all have the potential to accelerate the DOE goal to advance cyber resilient energy delivery systems that are designed, installed, operated, and maintained to survive a cyber-incident while sustaining critical functions.

The City of Fort Collins Utilities Light & Power understands the value of SAGA and derivative projects and recommends the U.S. Department of Energy continue supporting SAGA research, development, and commercialization.

Sincerely,

Adam Bromley, P.E.
Director of Operations & Technology
City of Fort Collins Utilities Light & Power

NLI
NORTHERN LIGHTS, INC.
The power of local service
A Touchstone Energy® Cooperative

January 26, 2023

Mr. Michael Ingram
National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401

RE: Letter of Support to continue NREL's SAGA Research & Development

Dear Mr. Ingram,

Northern Lights, Inc., a member owned electric cooperative serving northern Idaho and western Montana, recommends the U.S. Department of Energy (DOE) fund continued commercialization of NREL's "Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics: Near Real-Time Cyber-Physical Resiliency through Machine Learning". We are founding members of the SAGA Technical Review Committee and have guided SAGA research, development, and deployment efforts dating back to 2017.

We believe SAGA can deliver value by integrating cyber-physical data from Cable broadband networks with utility information systems to enhance electric distribution grid visibility and operational situational awareness, detecting patterns of operation indicative of cyber incidents. SAGA derivative efforts such as the Power Event Notification Systems (PENS), the GridCON Report, and the ANSI/SCTE 271 Standard, *Requirements for Power Sensing in Cable and Utility Networks,* all have the potential to accelerate the DOE goal to advance cyber resilient energy delivery systems that are designed, installed, operated, and maintained to survive a cyber-incident while sustaining critical functions.

Northern Lights understands the value of SAGA and derivative projects and recommends the U.S. Department of Energy continue supporting SAGA research, development, and commercialization.

Sincerely

Steve Elgar

Dr. Steve Elgar
President
Northern Lights, Inc.