

Abordagem Híbrida-Evolutiva para Detecção e Localização de Perdas Não Técnicas em um Sistema de Distribuição de Energia Elétrica em Transição – ID 6099

Autor: Nathan Fernando de Oliveira Fiorini

Coautores: Juan M. Home-Ortiz, Vinicius C. Cunha, Tailan R. Garcia, Felipe Bayma, Ricardo Torquato, Fernanda C. L. Trindade, Walmir Freitas

Empresa: CPFL Paulista

instituto
abradee



CEMIG



XXV Seminário
Nacional de
Distribuição de
Energia Elétrica

SENDI
2025
BELO HORIZONTE

Desafio das distribuidoras

Contextualização

- ❑ Como qualquer distribuidora, a CPFL enfrenta desafios inerentes ao negócio para manter **a alta qualidade da distribuição de energia elétrica** e a **viabilidade econômica**;
- ❑ Dentre os **desafios** está a sua capacidade de **reduzir as perdas**, de modo a alcançar a cobertura regulatória, que garante a sustentabilidade financeira dado problemas que fogem da ação da distribuidora:
 - **Fenômenos físicos**, como o efeito joule;
 - **Criminalidade**, que impossibilitam a ação da distribuidora;
 - **Áreas com Rede clandestina**, que necessitam de anuência para regularização;
 - Etc.

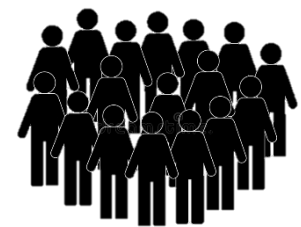
Perdas em distribuição de energia elétrica



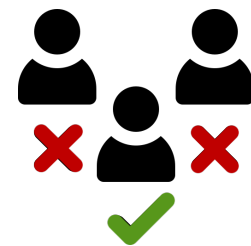
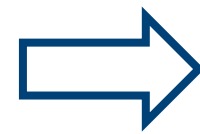
Detecção de irregularidade

Qual o Objetivo?

Eleger quais instalações apresentam comportamento suspeito (fraude ou problema na medição) e deverão ter suas medições inspecionadas



Mais de 10 milhões de clientes



Envio de 600 mil instalações para inspeção por ano (6%)



Maximizar a quantidade de TOI's



Maximizar a Recuperação de Energia + Receita



Premissa básica: “A grande maioria dos clientes não furta”



Prova disso:
em sorteio aleatório a assertividade não ultrapassa **4%**
(400.000/600.000 = 66%)



A assertividade praticada no grupo CPFL gira em torno de **25%**



Outros desafios:
qualidade nos serviços; bases de dados não confiáveis, capacidade de processamento e necessidade celeridade no processo.

Detecção de irregularidades: Variáveis

O consumo é uma das variáveis mais importantes, porém, não pode trabalhar sozinho!

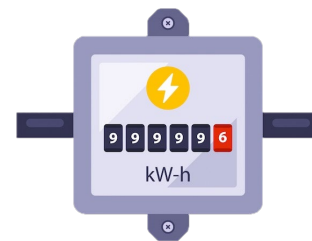
Quando observamos apenas o consumo, comportamentos iguais podem trazer resultados diferentes.

Devido a isso, **mais de 100 variáveis** são utilizadas na modelagem, como por exemplo:



Dados Cadastrais do Cliente

- Idade;
- Município e Bairro;
- Antiguidade da ligação;
- Quantidade de instalações mesmo PN;
- Ramo de atuação;
- Classe;
- CNAE;
- Etc.



Dados de Medição

- Marca e Modelo do Medidor;
- Antiguidade do medidor ;
- Tipo de ligação (M, B ou T);
- Alimentador;
- Transformador;
- Estado de fornecimento;
- Etc.



Histórico de interação com o cliente

- Contato com a empresa;
- Quantidade de inspeções;
- Quantidade Fraudes;
- Antiguidade última inspeção;
- Quantidade de ações de corte;
- Dívida;
- Etc.



Georreferenciamento e Comparação com os vizinhos

- Coordenadas;
- Consumo cliente X Vizinhança;
- Quantidade de fraudes X bairro;
- Quantidade de fraudes no Geohash;
- Quantidade de avarias no Trafo;
- Etc.

Detecção de irregularidade: Ferramentas

Estatística voltada para Seleção de Alvos

A **estatística** é o campo da matemática que **relaciona fatos e números** em que há um conjunto de métodos que nos possibilita coletar dados e analisá-los, assim sendo possível realizar alguma interpretação deles. A estatística é dividida em duas partes: **descritiva e inferencial**.

Para qualquer análise de dados, cabe um estudo para determinar as características e comportamentos que os dados apresentam. Por essa razão, utilizamos técnicas de:

- Análise descritiva/exploratória;
- Padronização e normalização;
- Detecção de outliers;
- Criação de nova variáveis;
- Regressões lineares/logísticas/ GLMs;



Machine Learning para Seleção de Alvos

Com algumas raízes na estatística, **Machine Learning (Aprendizado de Máquina)** é um **campo da inteligência artificial** que desenvolve algoritmos capazes de aprender padrões a partir de dados e fazer previsões ou tomar decisões sem serem explicitamente programados para cada tarefa.

Atualmente utilizamos diversas técnicas de Machine Learning para aumentar a acurácia na detecção de alvos, utilizando o **IBM SPSS Modeler** e a linguagem de **programação Python**. Dentre as técnicas utilizadas temos:

- Árvores aleatórias;
- Floresta aleatória;
- Extreme Gradient Boosting (XGBoost);
- Redes Neurais;
- Classificadores: DBScan, K-Neighbors, etc.



Mas ainda assim, essas técnicas não são suficientes, sendo necessárias ferramentas e variáveis mais avançadas e robustas.

Detecção de irregularidade: Fontes de dados

- ✓ **Baixa integração** entre **métodos de cálculo de PT** e **métodos de localização de PNT** (diferentes bases, insumos, modelos, premissas);
- ✓ **Baixa integração** entre as **bases de dados de engenharia, comercial e de operação**;
- ✓ Baixo nível de pré e pós-processamento de dados pelos **métodos baseados em reconhecimento de padrões**



O projeto tem como objetivo geral **desenvolver metodologias** baseadas em técnicas de **Data Science e Engenharia** para **detecção e localização mais eficiente de perdas não técnicas** nas redes de distribuição, especificamente de furto de energia via consumo irregular, fraude de medidores etc.

Evolução 1: sistemas com nível **baixo** de disponibilidade de medição:

✓ **Dados elétricos (GIS) + dados comerciais + dados de operação**

Evolução 2: sistemas com nível **médio** de disponibilidade de medição (AMI – partially deployed):

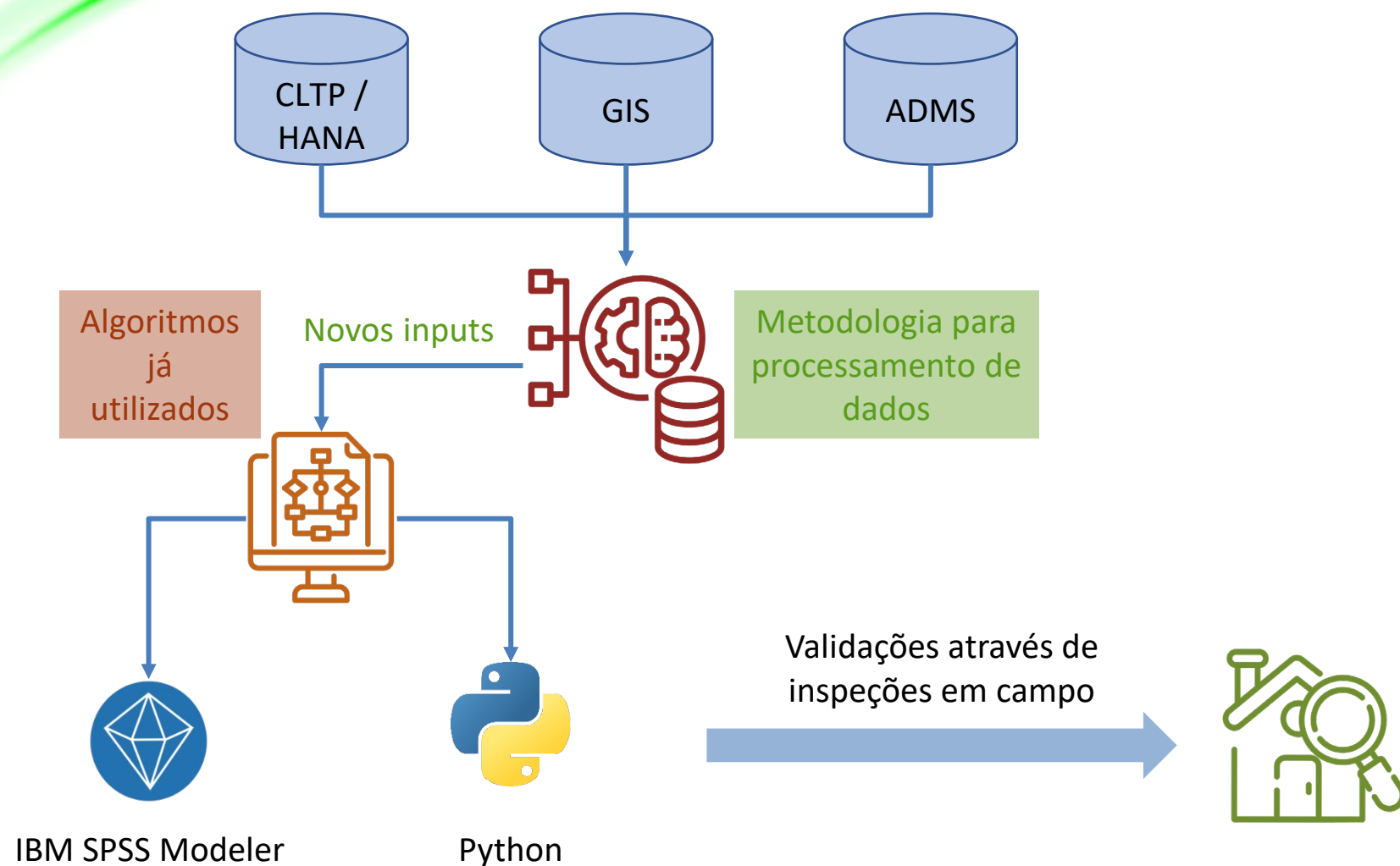
✓ **Dados da Evolução 1 + medições série-temporais esparsas (UCs do grupo A; religadores; reguladores de tensão)**

Evolução 3: sistema com nível **alto** de disponibilidade de medição (AMI – fully deployed):

✓ **Dados da Evolução 2 + medições série-temporais em UCs do grupo B**



Evolução 1: Integração de dados



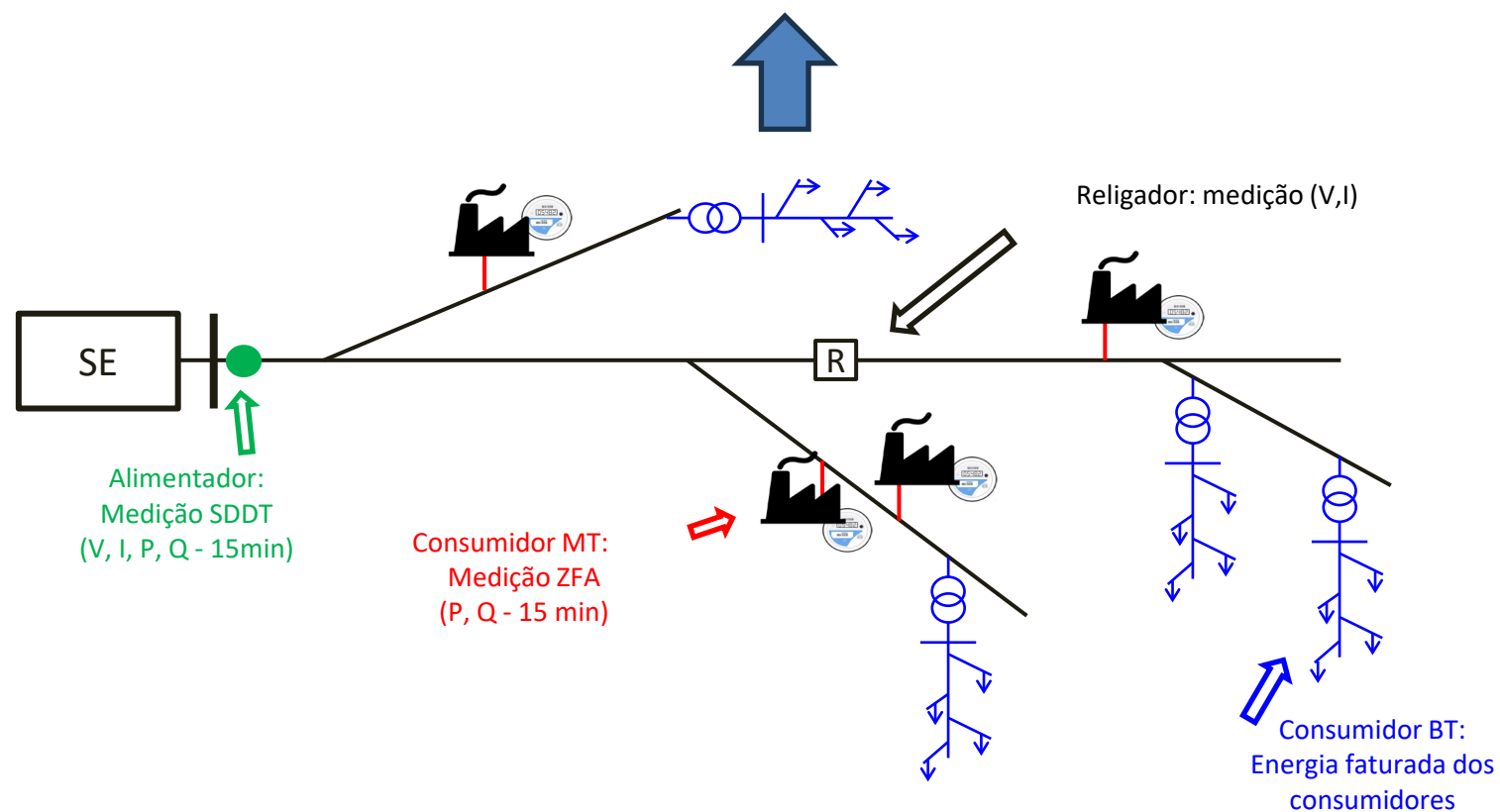
Entregáveis da Evolução 1:

- **Metodologia** para leitura, processamento e **integração** dos novos tipos de dados (de **engenharia e operativos**);
- **Metodologias** para tratamentos dos montantes de **energia não faturados** no mesmo mês (calendário de faturamento do grupo B) e **inconsistências nas bases de dados**, com intuito de **detectar e restringir eletricamente regiões**, locais, consumidores (com ou sem contrato) **com perdas não técnicas**;
- **Ferramenta computacional** que implementa a metodologia de processamento de dados desenvolvida e disponibiliza os dados processados como novas entradas (novos inputs) para os algoritmos e modelos existentes na área comercial da CPFL (e.g., programa IBM SPSS Modeler, scripts dedicados em linguagem de programação Python, entre outros).

Evolução 2: Evolução 1 + medições série temporais esparsas



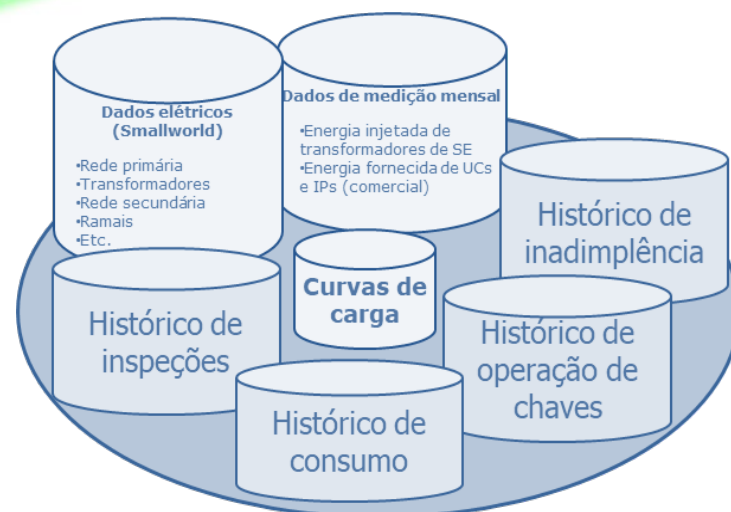
Dados elétricos, comerciais e operacionais



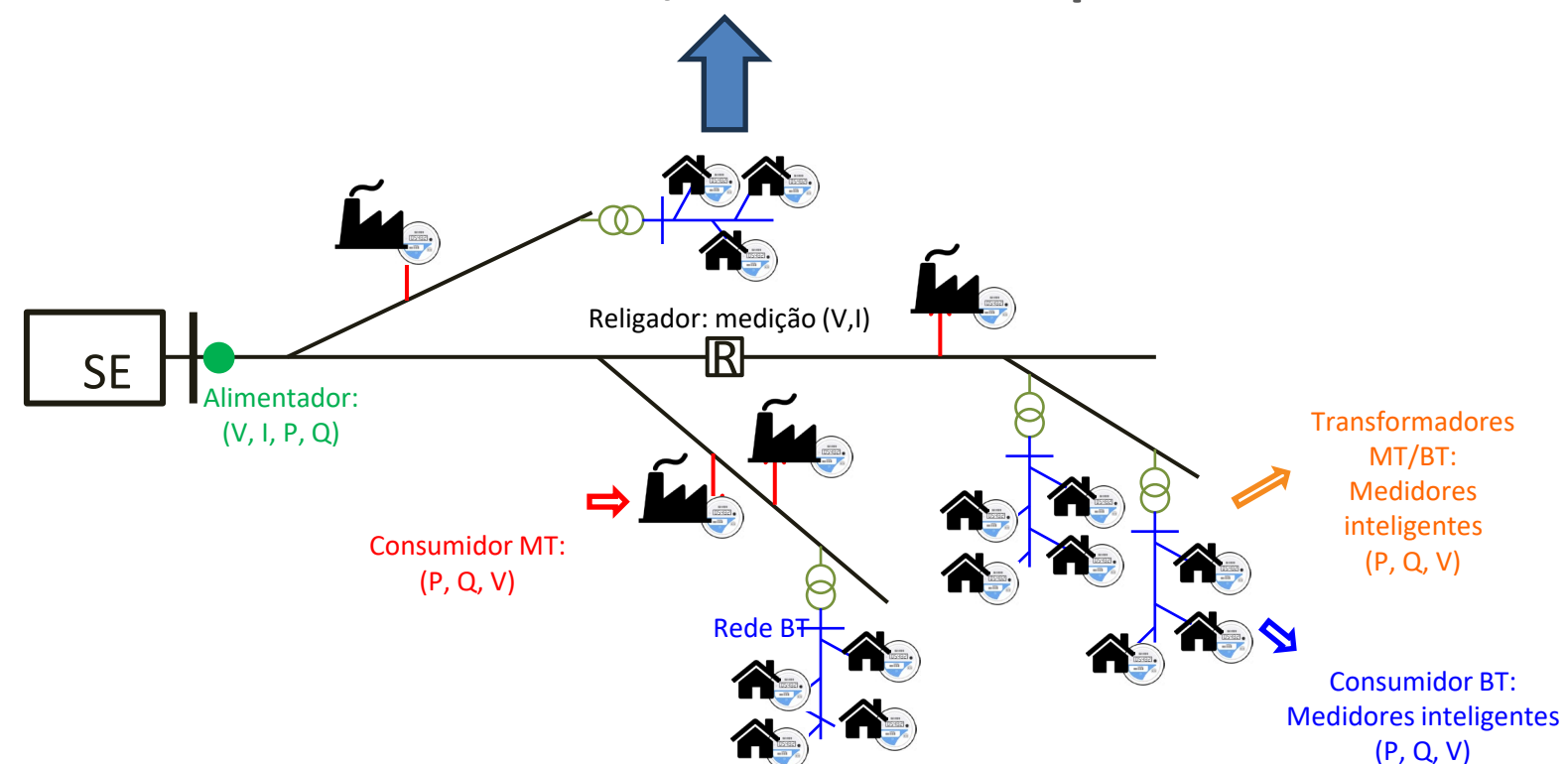
Entregáveis da Evolução 2:

- Integração de dados elétricos + dados comerciais + dados de operação + **medições série-temporais** com resolução de 15 min. (**entrada do alimentador, consumidores do grupo A, religadores etc.**)
- **Aperfeiçoar** a abordagem da **Evolução 1** considerando a subdivisão dos alimentadores (e.g., jusante de um **religador, região com redes de BT específicas** etc.) e menores intervalos de aquisição de dados de medição.

Evolução 3: Evolução 2 + medições série temporais do grupo B



Dados elétricos, comerciais e operacionais



Entregáveis da Evolução 3:

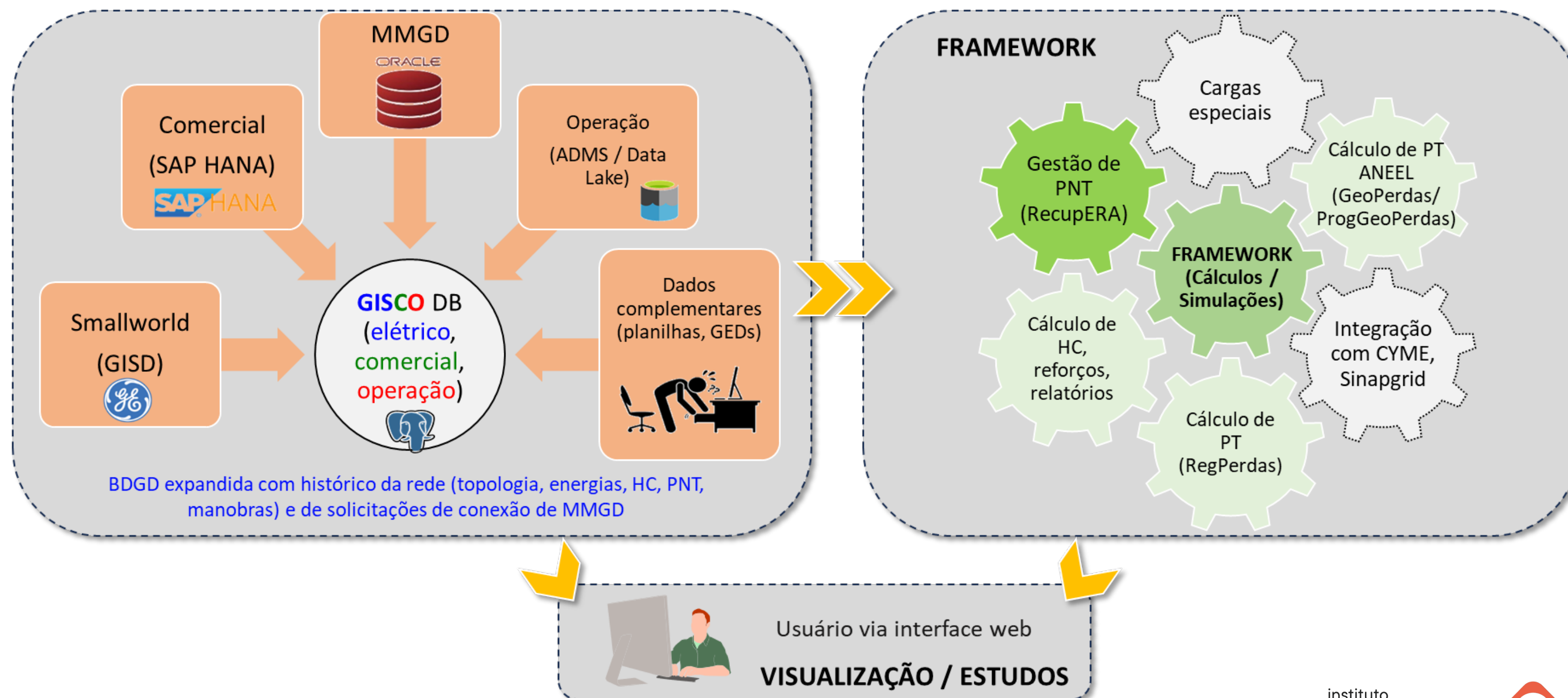
- **Alta cobertura espacial**, permitindo análises, por exemplo, em redes de baixa tensão individualizadas. **Devido à maior disponibilidade de dados** com implementação da **telemedição do Grupo B**, é possível realizar estudos mais precisos:

- ✓ **Correlação de medições de tensão, métodos baseados em fluxo de carga, estimação de estados dedicada.**

PA3092: GESTÃO PNT – Solução estrutural integrada

Tailored in-house solution

A solução será **implementada e integrada** usando o framework de gestão de redes de distribuição iniciado com o **projeto Perdas Técnicas (PA3048)**. Esta **estrutura é expansível e modular** atendendo diversas necessidades do Grupo CPFL Energia.



PA3092: GESTÃO PNT – Ferramenta web

ERAnalytics

Plataforma unificada para análise de redes de distribuição



HC

Visualização e gestão de capacidade de hospedagem de geração distribuída

MAPA DASHBOARD MINIGD



PT

Visualização e gestão de perdas técnicas

SIMULAÇÃO



RecupERA

Visualização e gestão de perdas não técnicas, regiões e consumidores suspeitos

MAPA DASHBOARD



Cargas Especiais

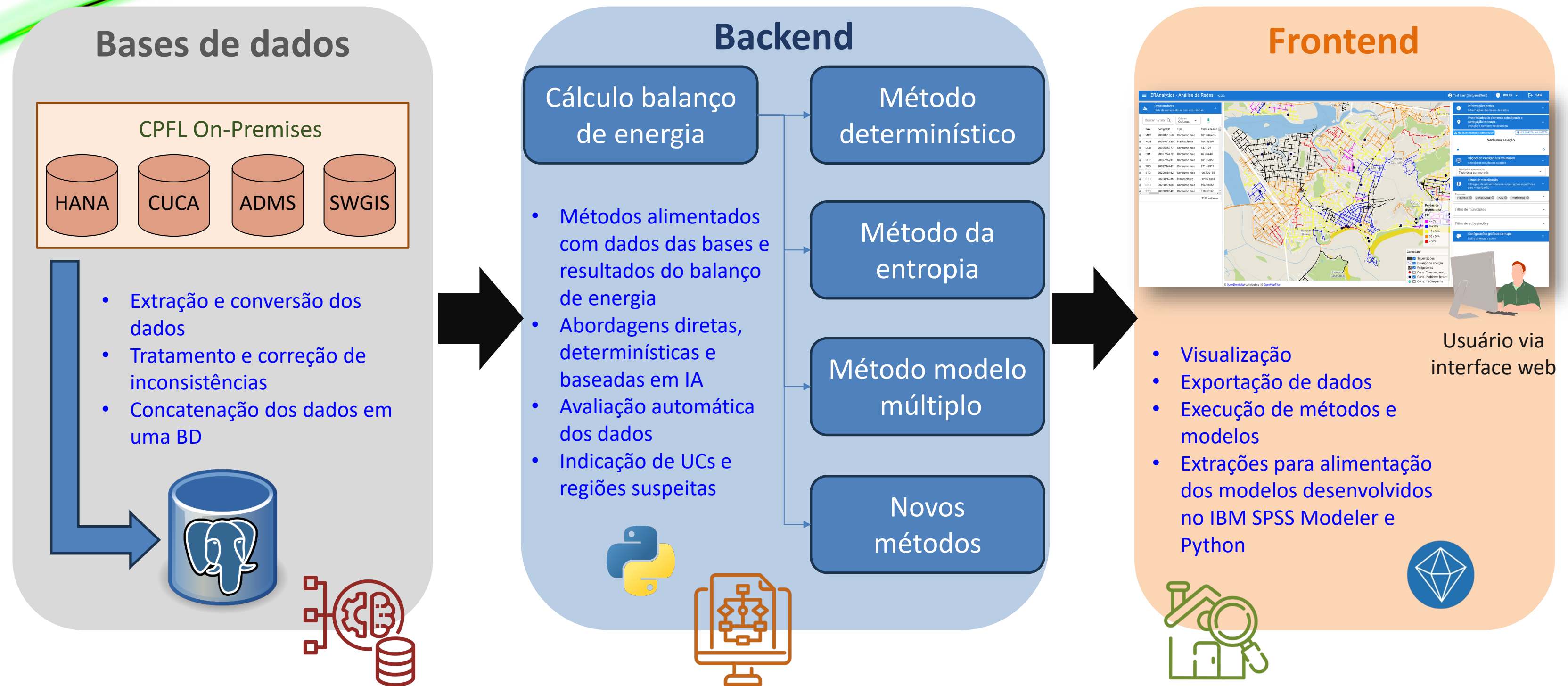
Modelagem e simulação de cargas especiais

SIMULAÇÃO



PA3092: Visão geral da ferramenta

RecupERA

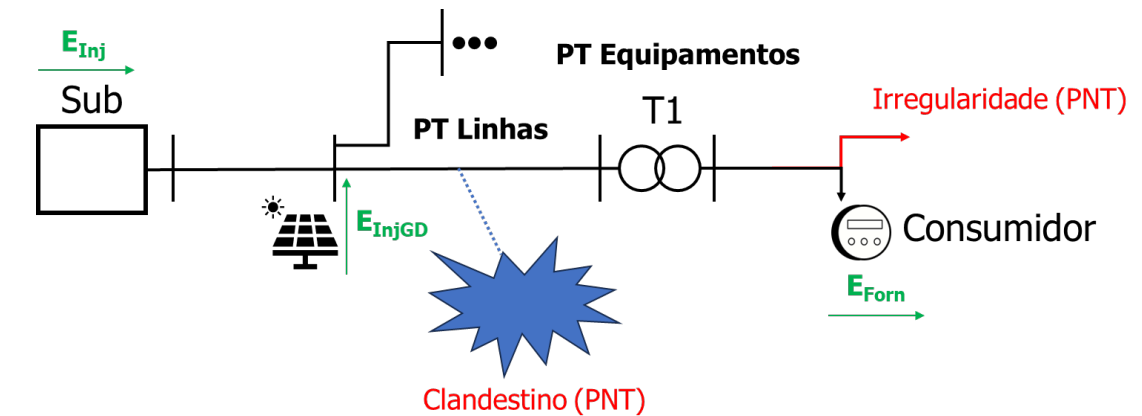


Resumo do cálculo de balanço de energia

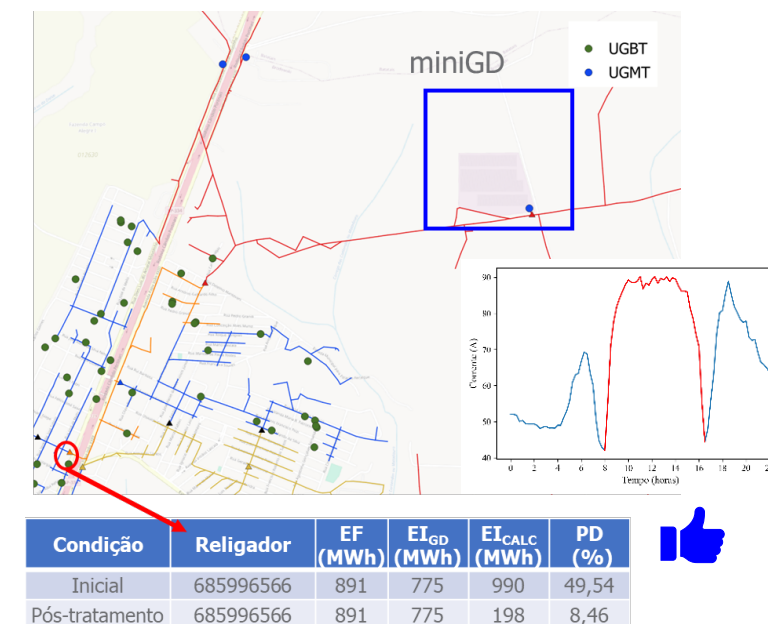
Calculam-se as perdas da distribuição (**PD**) através do balanço de energia, indicando regiões com maiores ou menores tendências de irregularidades (**PNT**). Resumem-se abaixo as principais funcionalidades implementadas para o cálculo do balanço de energia:

- Cálculo do balanço de energia em alimentadores e religadores (melhor filtragem de regiões)
- Utilização da **janela de faturamento** ao invés do mês civil (maior alinhamento com o consumo medido)
- Estimação de **fluxo de potência reverso** para medidas de corrente em alimentadores e religadores (adaptação de medições disponíveis)
- **Dados elétricos atualizados** mensalmente com informações da engenharia – SWGIS (aprimoramento do cálculo de balanço)
- **Tratamento de chaves** utilizando dados da operação – ADMS (aprimoramento do cálculo de balanço)

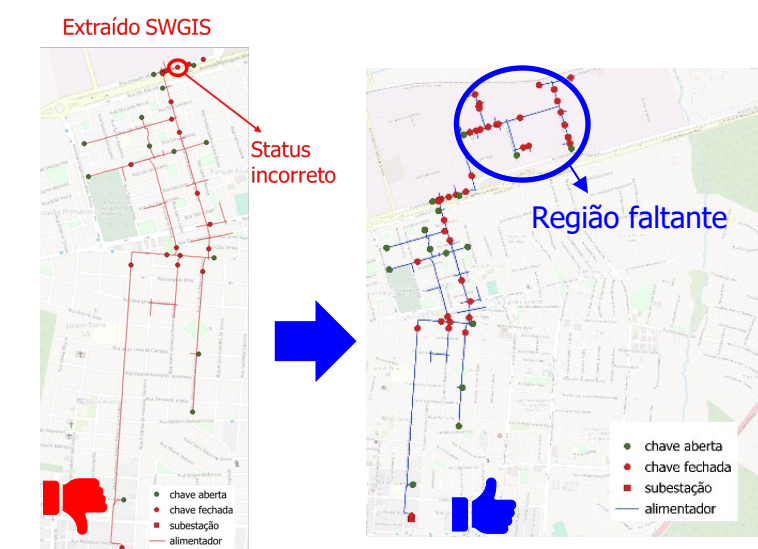
$$PD = PT + PNT = E_{inj} + E_{injGD} - E_{forn}$$



Identificação fluxo de potência reverso (medidas de corrente)



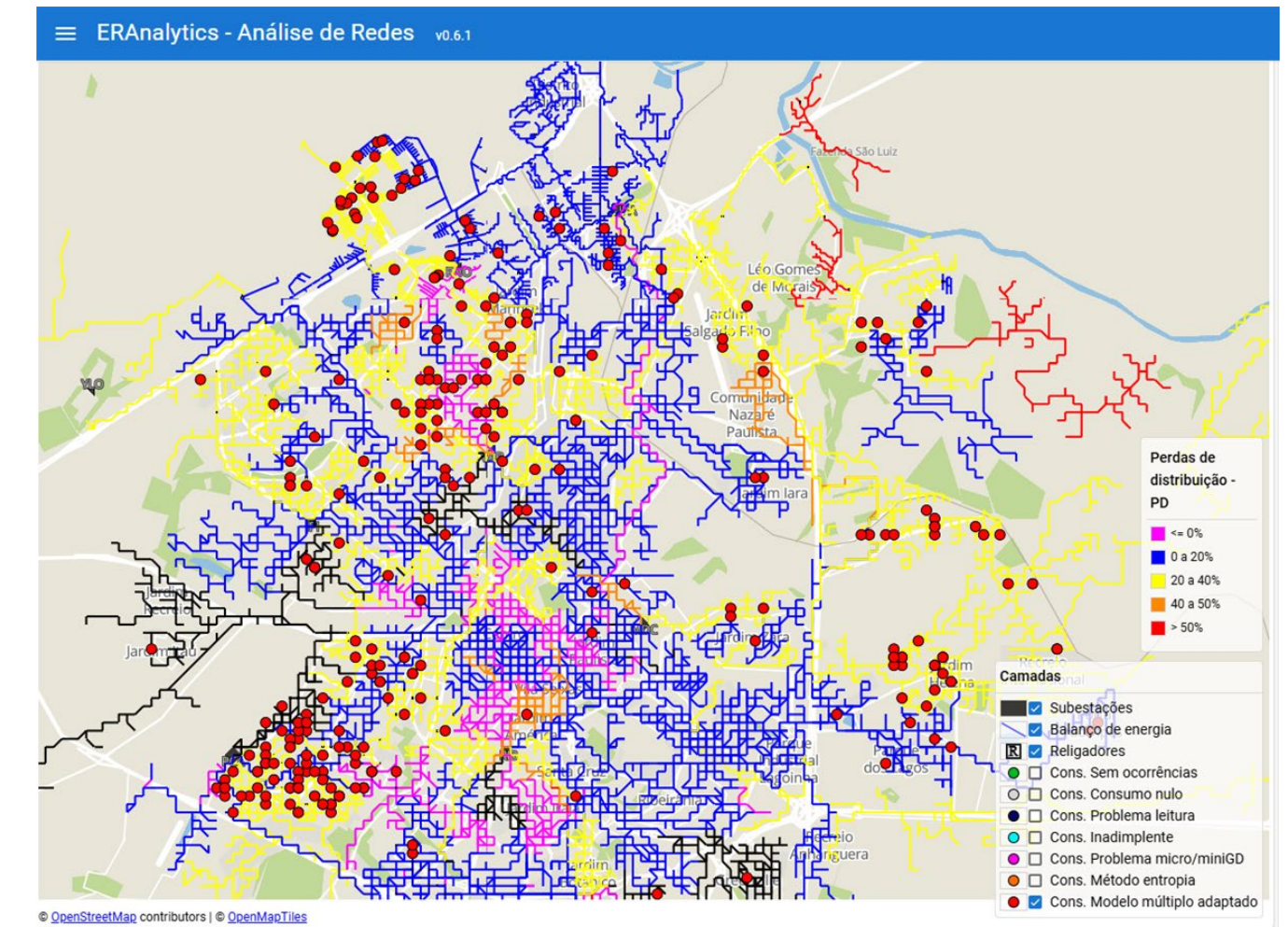
Correção topologia (dados operação)



Resumo do Modelo Múltiplo Modificado

O **modelo múltiplo modificado** é baseado no modelo desenvolvido na ferramenta IBM SPSS Modeler. Resumem-se abaixo as principais funcionalidades incrementadas para este modelo:

- Implementado em ambiente **Python** e integrado à ferramenta (**maior flexibilidade e controle dos modelos**)
- Utilização de **IA/Machine learning** para treinamento e classificação de UCs (e.g., Random Forest e Gradient Boosting)
- Divisão de UCs de acordo com a **classe de consumo** (**melhor representação de subgrupos**)
- Adição de **novos dados de entrada**, tais como valores de perdas da distribuição (balanço de energia) e índice de notas que se correlacionam com irregularidades (**informações que se mostraram relevantes no treinamento do modelo**)
- As regiões das UCs indicadas condizem com regiões de maiores perdas da distribuição (**sinergia entre o modelo e o balanço de energia**)



Alinhamento geográfico
entre UCs indicadas e
valores elevados de perdas

Resumo do método determinístico

Os métodos determinísticos em desenvolvimento apresentam os seguintes objetivos:

1. Determinar inconsistências que estejam impactando o cálculo de balanço de energia, tal como UCs com consumo nulo
2. Criar regras determinísticas para avaliar UCs com consumo nulo não justificável (possíveis PNTs)
3. Indicar regiões para aplicar os métodos de detecção de fraudes

Para atingir tais objetivos, as principais regras são resumidas abaixo:

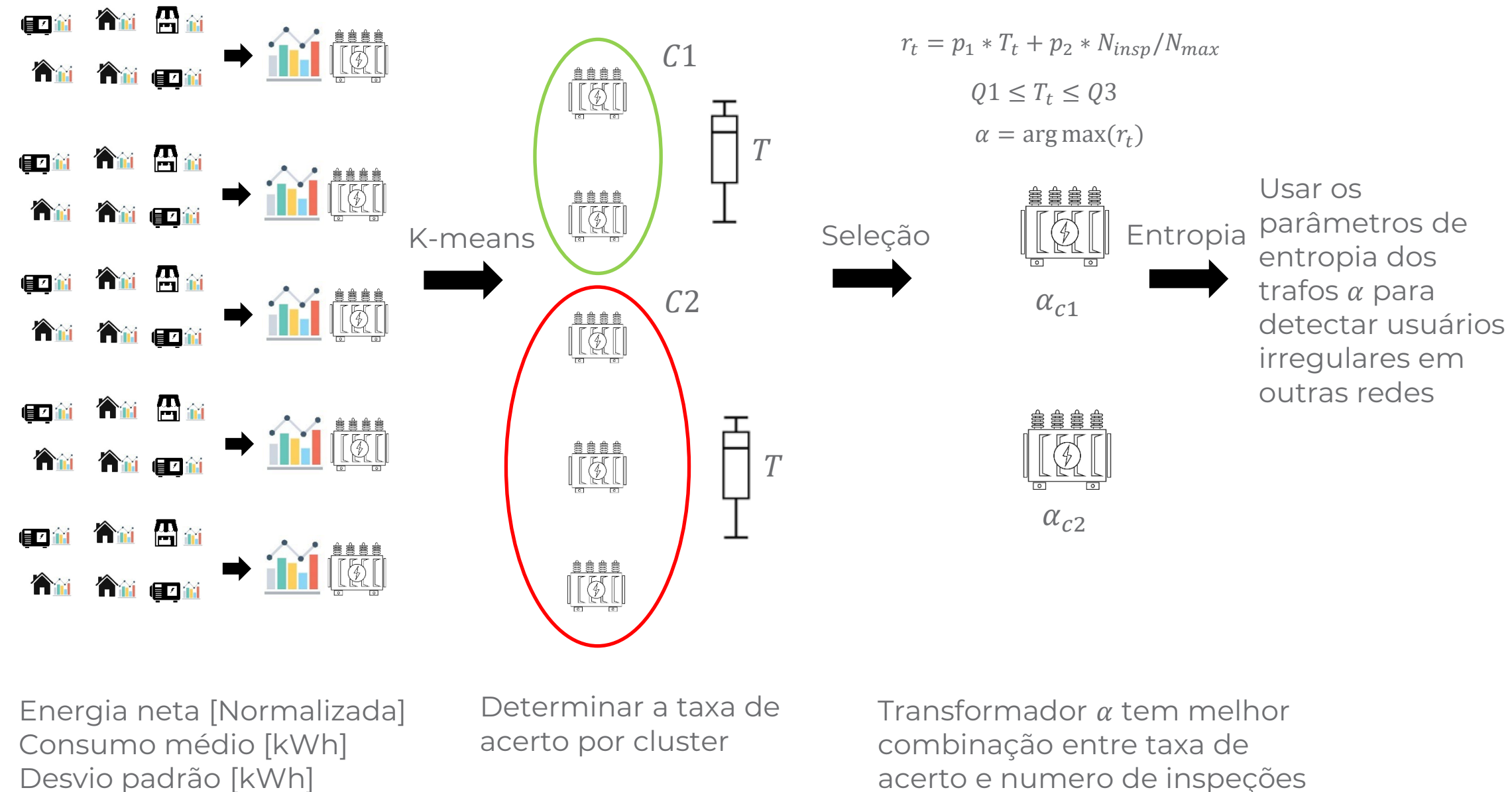
a. Problemas de leitura	b. Inadimplentes	c. Consumo nulo	d. UCs suspensas com consumo	e. Código leiturista	f. Problemas GD
<ul style="list-style-type: none">• Base de consumo nula e base com leituras dos medidores não nulas	<ul style="list-style-type: none">• Base de consumo e leitura não registram valores mesmo antes da atividade de corte de fornecimento em algumas UCs	<ul style="list-style-type: none">• Sem um motivo identificado para o consumo nulo de algumas UCs• Possível necessidade de inclusão de novos critérios de avaliação	<ul style="list-style-type: none">• Potencial problema de dados para definir a UC como "Completamente suspensa"	<ul style="list-style-type: none">• Filtro a partir dos códigos do leiturista	<ul style="list-style-type: none">• Energia registrada na base excede consideravelmente (e.g., 4x) a leitura feita em campo• Energia não condiz com capacidade da GD

Resumo do modelo de entropia

O **modelo da entropia** se baseia na entropia de Shannon, que é um conceito que mede desordem em um conjunto de dados (i.e., **separação por incerteza**). Resumem-se abaixo as principais funcionalidades incrementadas para este modelo:

- Implementado em ambiente **Python** e integrado à ferramenta (**maior flexibilidade e controle dos modelos**)
- **Distinção nos valores de entropia** de inspeções regulares de inspeções com detecção de irregularidades (**potencial indicador de irregularidades**)
- Análise configurável para **diferentes agrupamentos**, tais como transformadores e alimentadores (**alta granularidade e especificidade nos modelos criados**)
- Aplicação de agrupamentos (**K-means**) para **maior abrangência** das regiões sob análise (**modelos específicos podem ser aplicados em regiões com poucas ou nenhuma inspeção – diversificação**)

Agrupamento de transformadores (K-means) para maior abrangência do modelo criado



RecupERA : Frontend – Mapa

ERAnalytics

Plataforma unificada para análise de redes de distribuição



HC

Visualização e gestão de capacidade de hospedagem de geração distribuída

MAPA

DASHBOARD

MINIGD



PT

Visualização e gestão de perdas técnicas

SIMULAÇÃO



RecupERA

Visualização e gestão de perdas não técnicas, regiões e consumidores suspeitos

MAPA

DASHBOARD



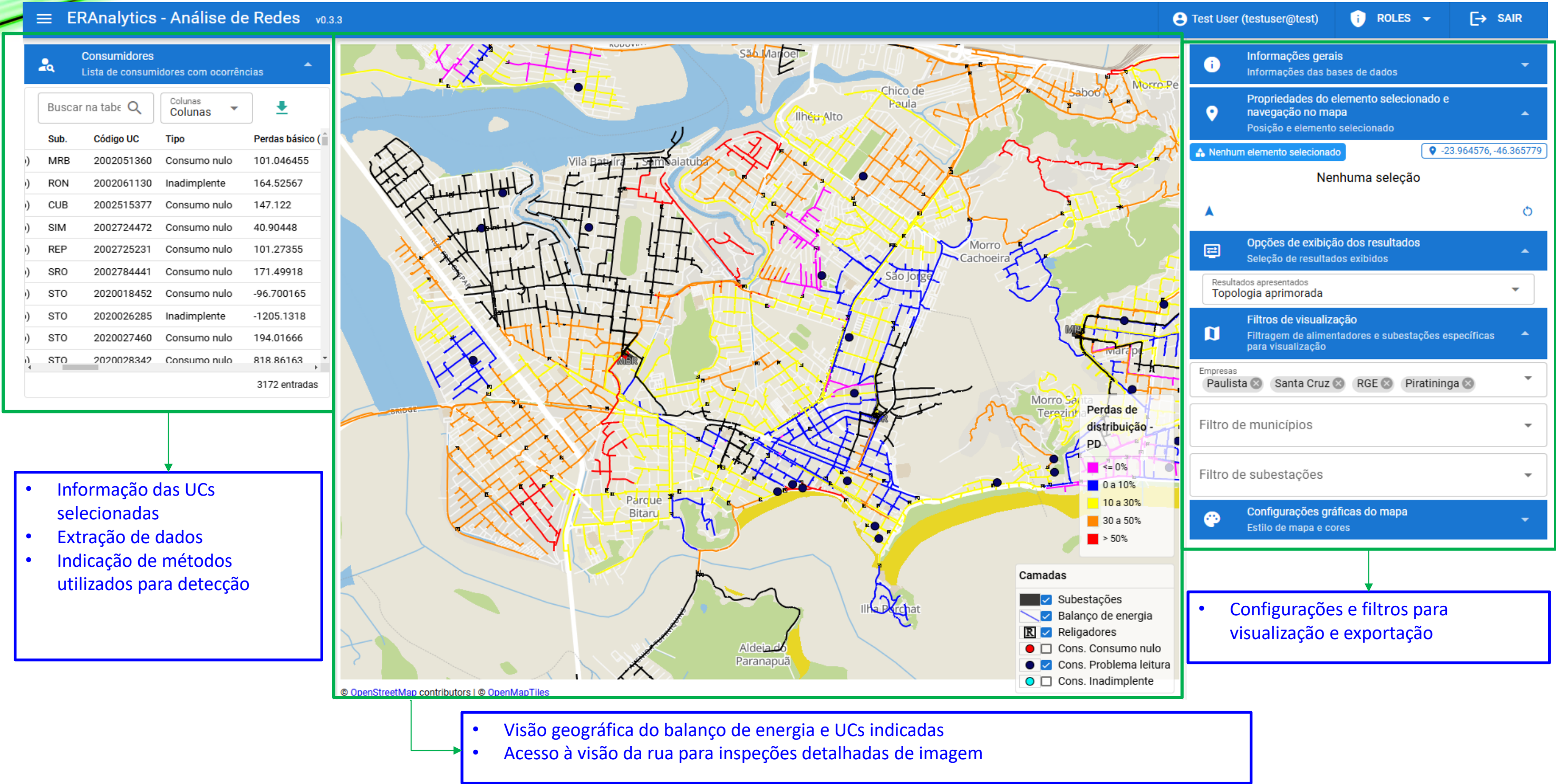
Cargas Especiais

Modelagem e simulação de cargas especiais

SIMULAÇÃO



RecupERA : Frontend – Mapa



RecupERA : Frontend – Dashboard

ERAnalytics

Plataforma unificada para análise de redes de distribuição



HC

Visualização e gestão de capacidade de hospedagem de geração distribuída

MAPA

DASHBOARD

MINIGD



PT

Visualização e gestão de perdas técnicas

SIMULAÇÃO



RecupERA

Visualização e gestão de perdas não técnicas, regiões e consumidores suspeitos

MAPA

DASHBOARD



Cargas Especiais

Modelagem e simulação de cargas especiais

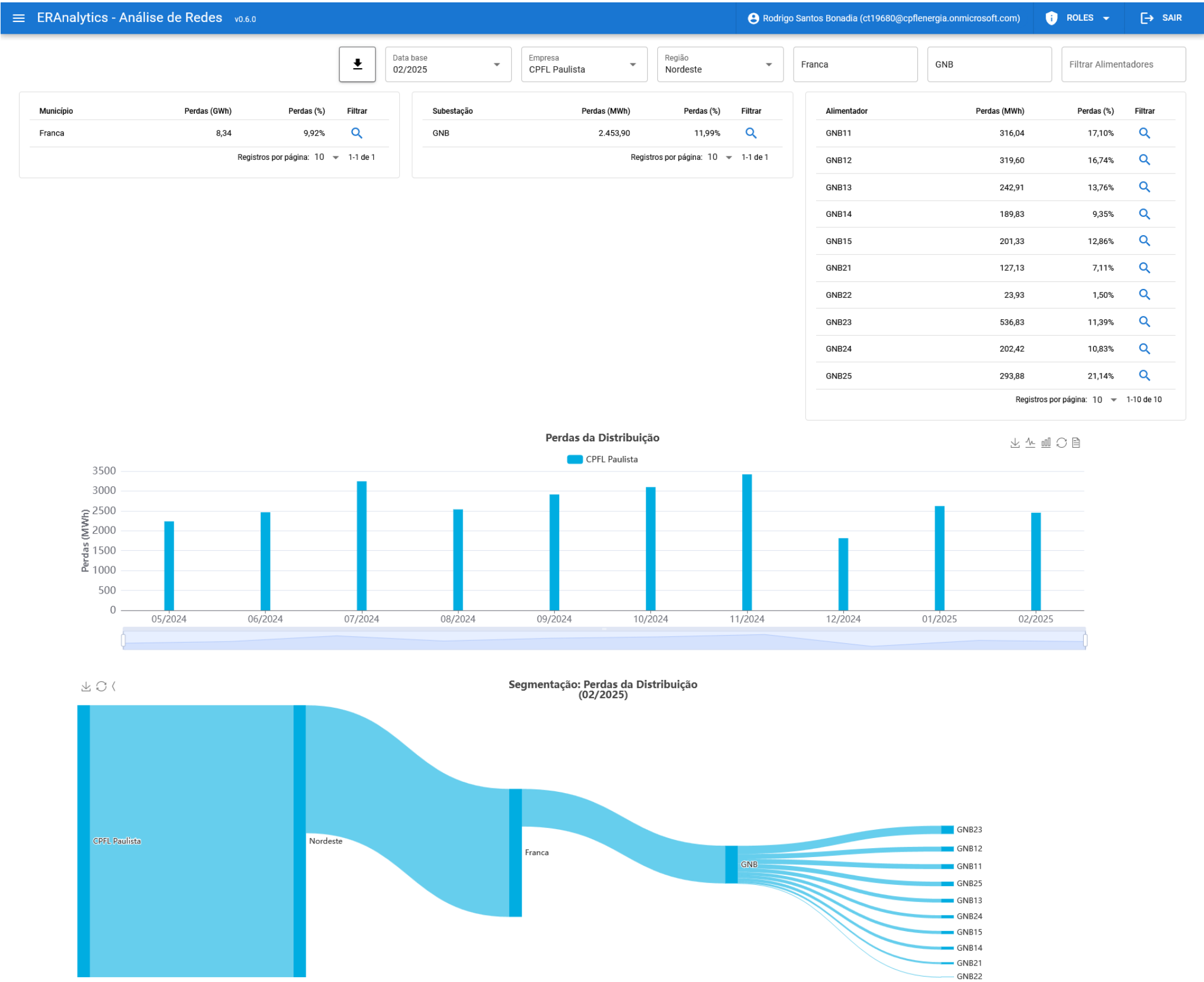
SIMULAÇÃO



RecupERA : Frontend – Dashboard

Dashboard de perdas na distribuição

- **Tabelas:** visualização detalhada das perdas por município, subestação e alimentador
- **Gráfico temporal:** visualização da evolução das perdas ao longo do tempo
- **Gráfico Sankey:** visualização da separação das perdas por município, subestação e alimentador



RecupERA : Frontend – Dashboard

- Filtragem por **município**
- Filtragem por **subestação**
- Filtragem por **alimentador**
- Todos os dados (tabelas, gráfico temporal e *Sankey*) são filtrados de acordo com a seleção



Comentários finais

- ✓ A ferramenta web foi desenvolvida em 12 meses de projeto e está em processo de testes e aperfeiçoamentos
- ✓ Até o momento contabilizam-se os seguintes ganhos no processo de gestão de PNT:
 - Maior integração e melhor visualização dos dados
 - Correção automática de bases de dados
 - Aprimoramentos no cálculo de balanço de energia
 - Maior flexibilidade na adaptação de métodos baseados em IA (modelo múltiplo modificado, desenvolvido em Python)
- ✓ Não necessita da instalação de dispositivos adicionais na rede



OBRIGADO!

