ELLIIT

Hybrid AI – Where data-driven and model-based methods meet

Focus Period Linköping 2022

Multilayer Hybrid Models for Power Outage Analysis of a Brazilian Distribution Company

Marcelo Azevedo Costa
Federal University of Minas Gerais

LiU LINKÖPINGS UNIVERSITET

LUND UNIVERSITY

UFMG

Introducing myself

- ✓ Full Professor (2019) of Applied Statistical Methods in the Department of Production Engineering at the Federal University of Minas Gerais/Brazil.

- ✓ I really enjoy data modelling.

- ✓ Research interests are: **statistical models** applied to the electrical sector, applied statistics, network analysis, **spatial statistics**, **time series analysis**, **artificial neural network** theory and applications.

Research Motivation

Statistical/computational modeling of **CEMIG**-**D** business model using databases and technical knowledge (2019-2021)

Generation

Transmission

Distribution

Hybrid AI – Where data-driven and model-based methods meet

The distribution company with the largest concession área in Brasil (**562.760 km²**)

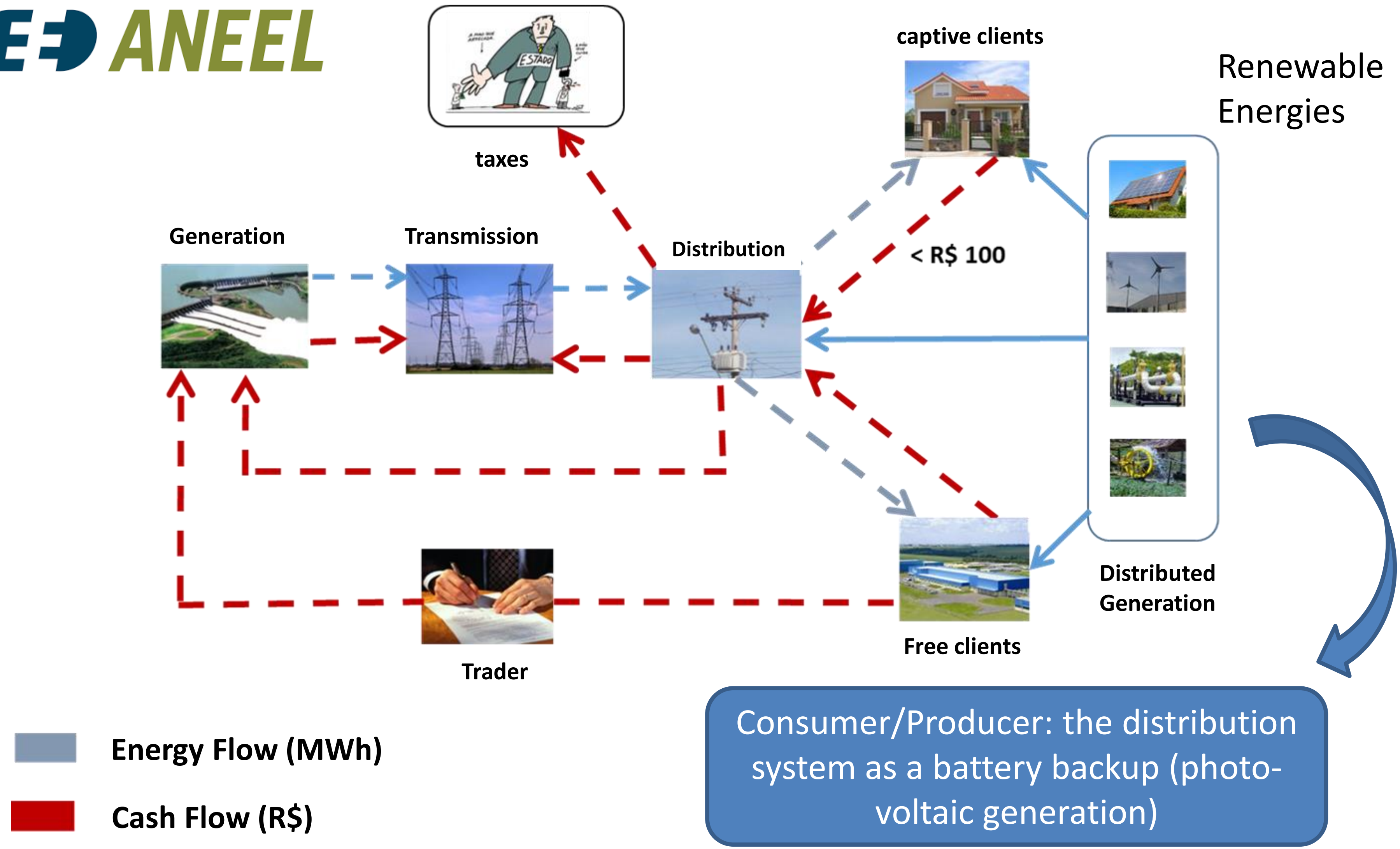
Statistical/computational modeling of CEMIG-D business model using databases and technical knowledge (2019-2021)

R&D Objective:

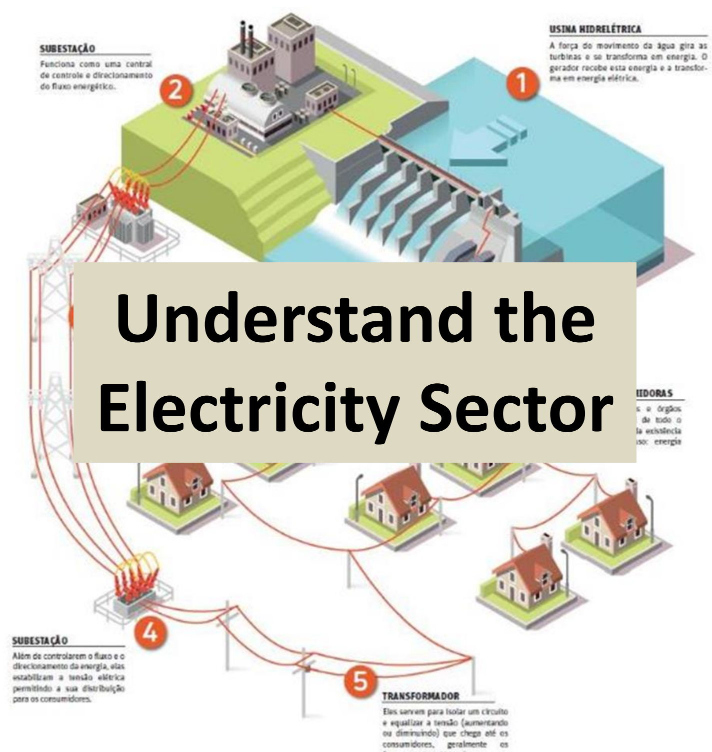
- To develop a computational system to simulate CEMIG-D business model using structural equation models, correlation matrices, regression models, machine learning models and artificial neural networks.

- The project was successfully finished in August 2021. The project received two awards in Brazilian Energy Conferences.

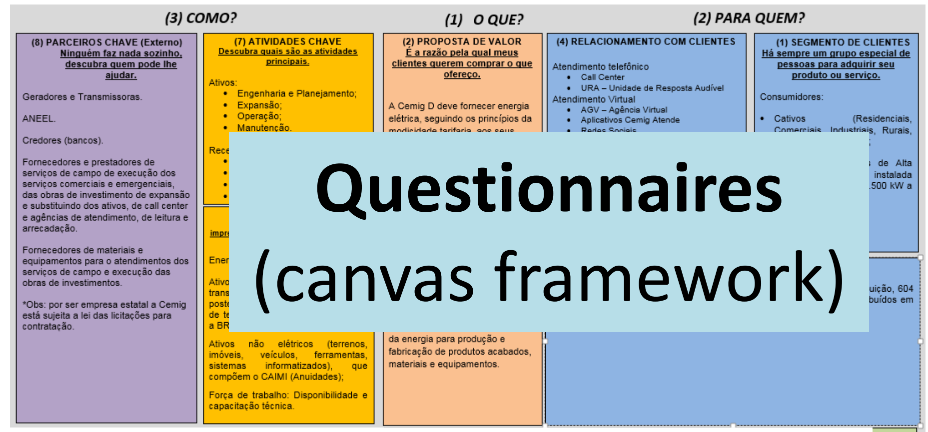
The Brazilian Energy Distribution Business Model



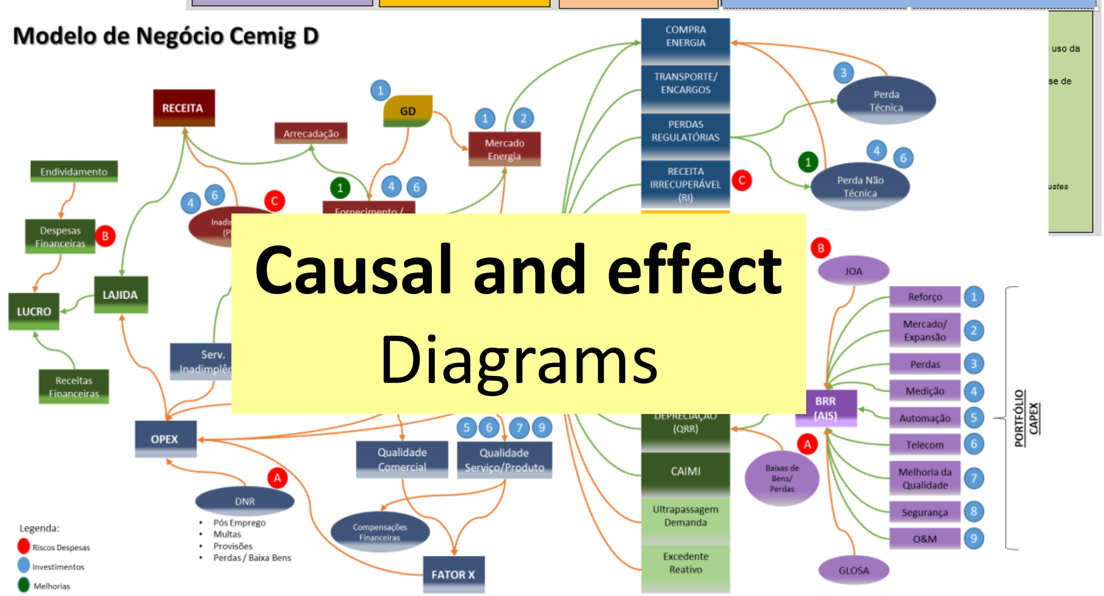
Extensive qualitative analysis using frameworks



Understand the Electricity Sector



Questionnaires (canvas framework)

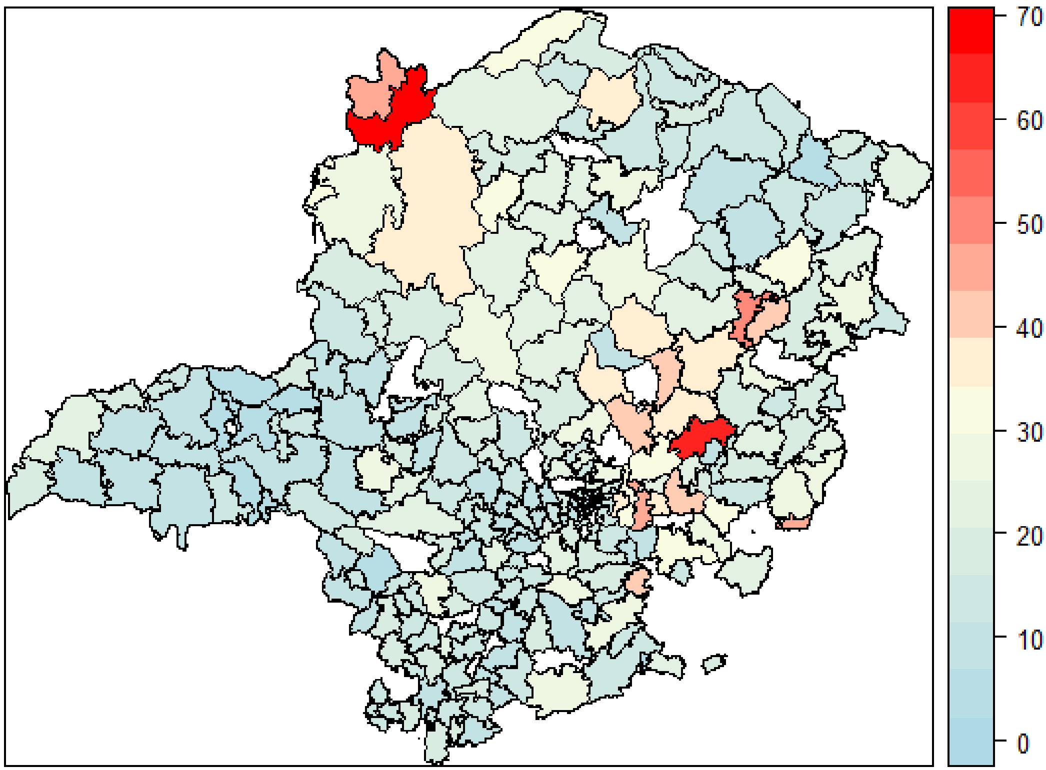


Causal and effect Diagrams

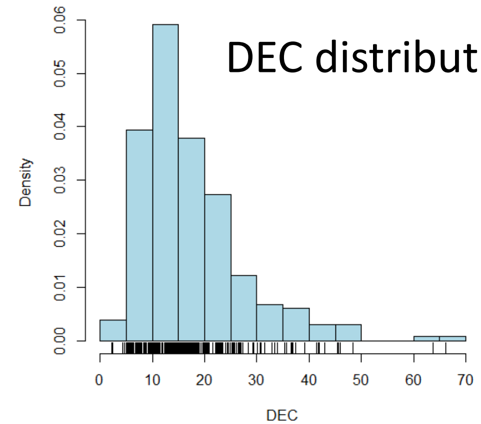
The Quality Model (First Case Study)

Duration of Interrupted Energy (DEC)

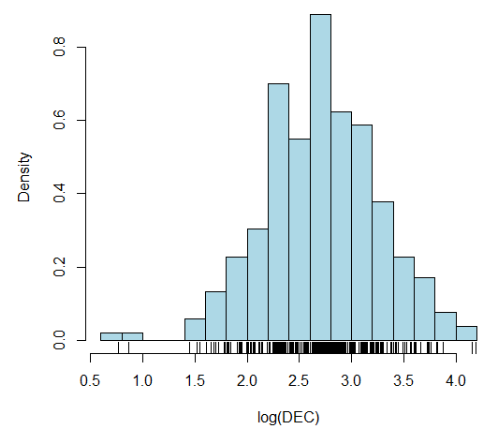
DEC indicator (2017)



DEC distribution



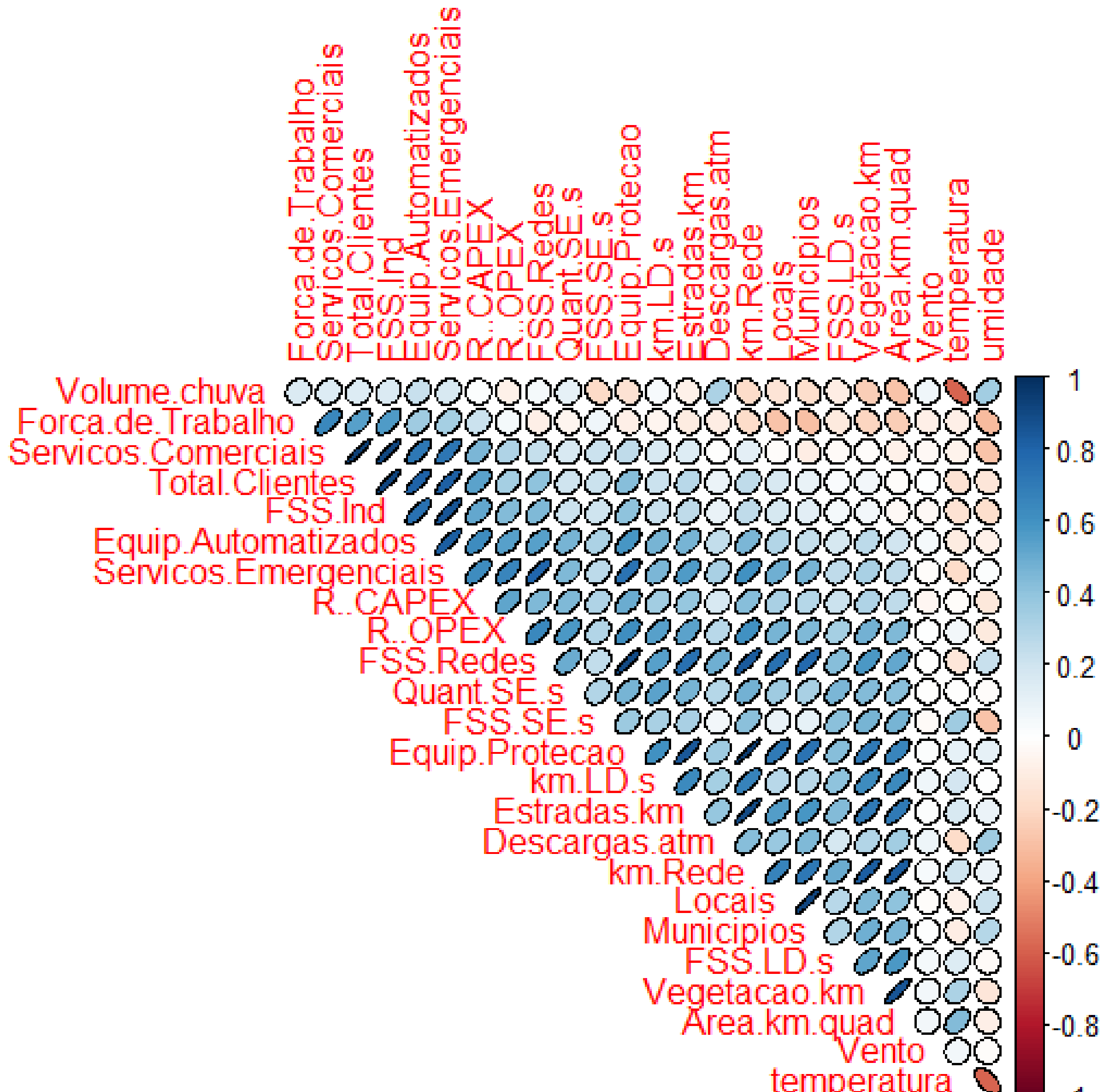
log(DEC) distribution



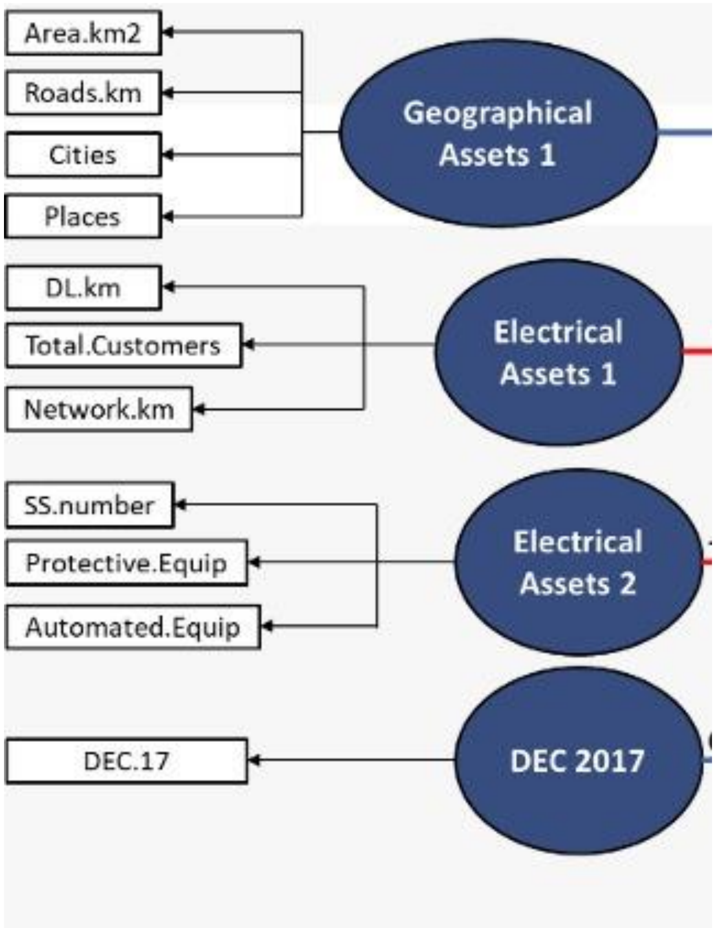
Drivers of the DEC based on technical evidence

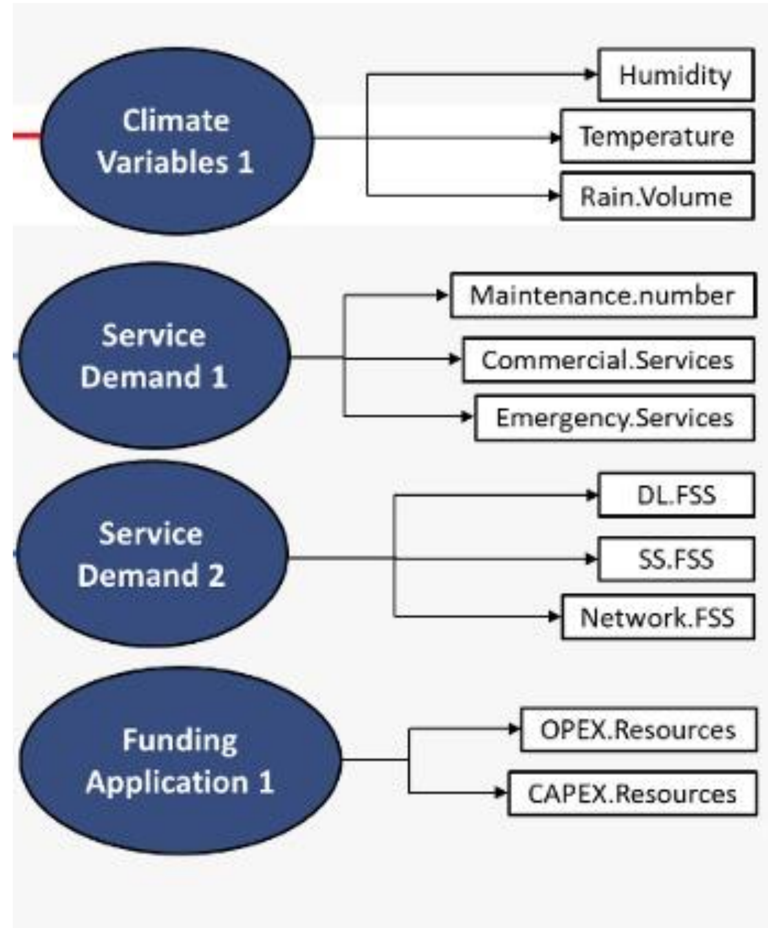
idx	Drivers	Expected correlation with DEC
1	Service area (km2)	**positive**
2	Extension of roads in the service area (km)	**positive**
3	Number of municipalities in the service area	**positive**
4	Number of locations served according to the electrical company definition	**positive**
5	Extension of distribution lines (km)	**positive**
6	Extension of distribution network (km)	**positive**
7	Number of consumers	**positive**
8	Number of substations	**positive**
9	Number of electrical protective equipment	**negative**
10	Number of automated equipment	**negative**
11	Humidity index (%)	**negative**
12	Average temperature (oC)	**negative**
13	Average precipitation (mm)	**positive**
14	Number of working (maintenance) teams	**negative**
15	Number of commercial services	**negative**
16	Number of emergency services	**negative**
17	Number of interruptions due to falling trees on the distribution lines	**positive**
18	Number of interruptions due to falling trees at substations	**positive**
19	Number of interruptions due to falling trees in the distribution network	**positive**
20	Operational expenditures (OPEX/R$)	**negative**
21	Capital expenditures (CAPEX/R$)	**negative**

Correlation between features



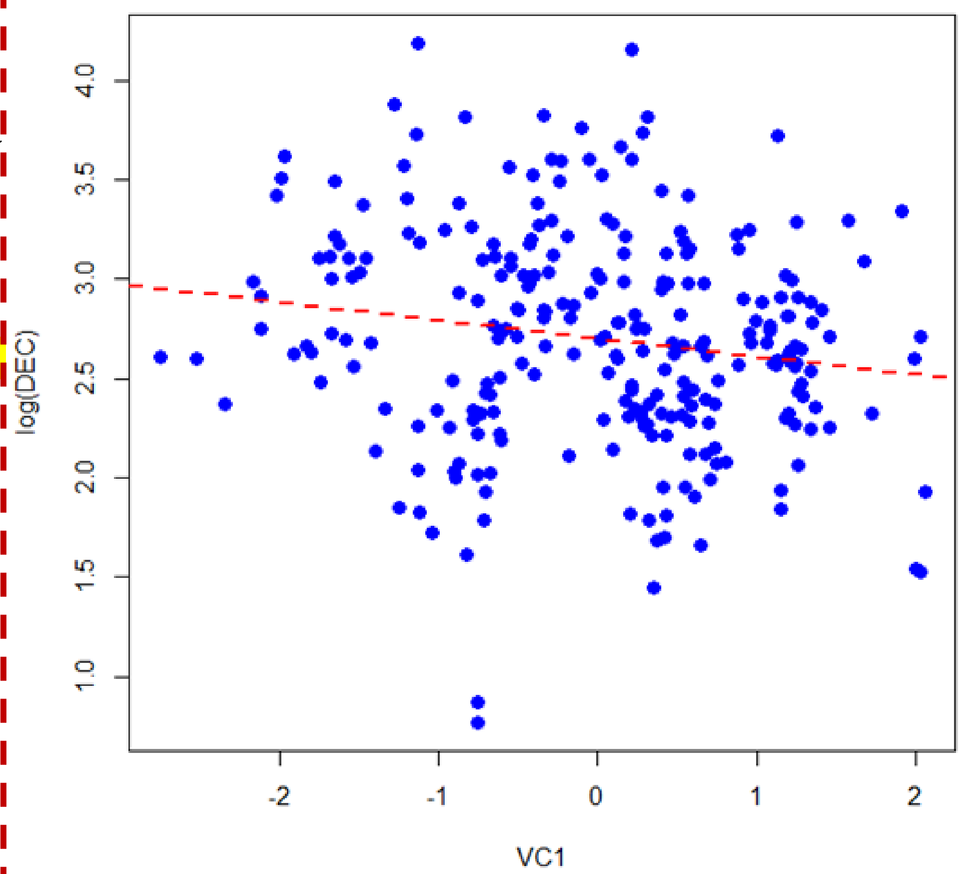
Structural Equation Modelling (SEM)





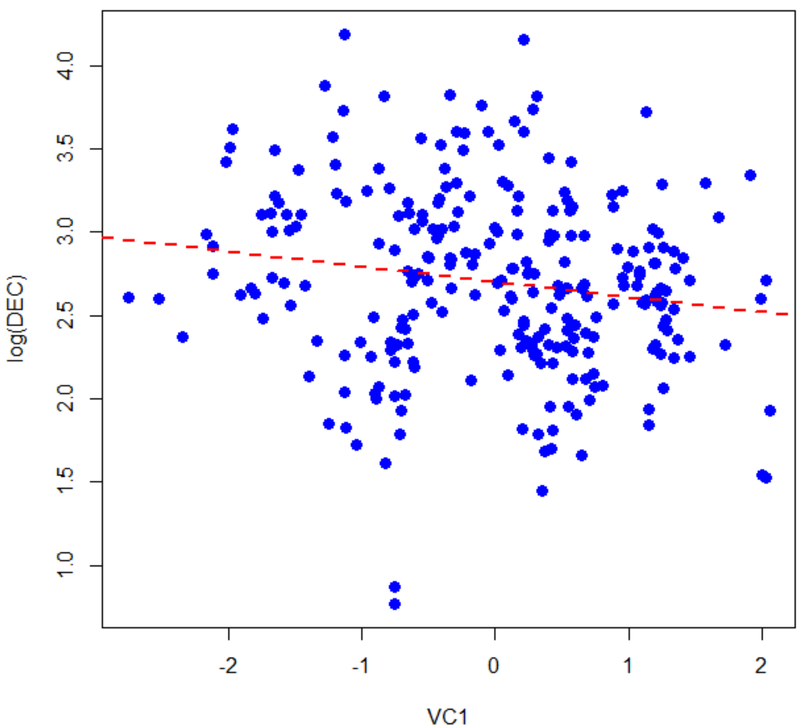
Then, the unexpected!!!

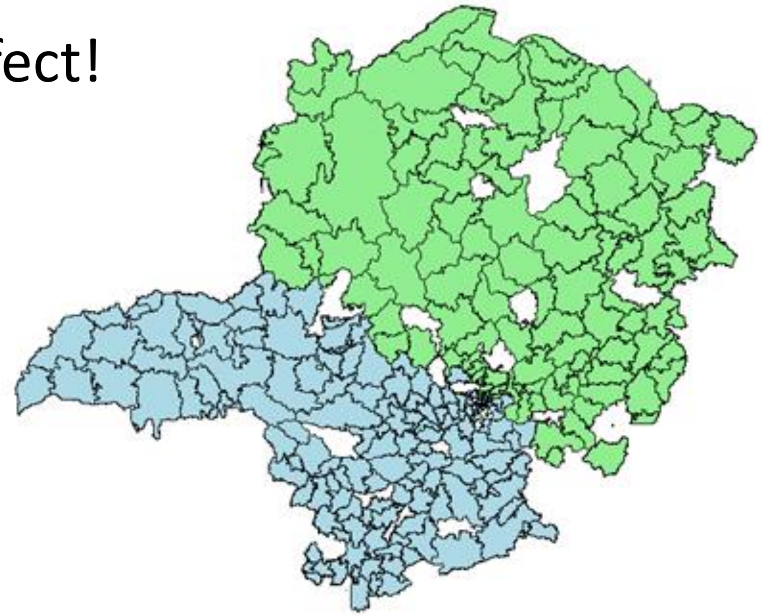
```
lm(formula = DQ ~ AG1 + AE1 + AE2 +

Coefficients:
              Estimate Std. Error t
(Intercept)  6.103e-17  4.066e-02
AG1          3.645e-01  8.198e-02
AE1         -5.087e-01  1.095e-01
AE2         -6.563e-01  1.132e-01
VC1         -1.831e-01  4.265e-02
DS1         -1.297e-01  7.985e-02
DS2          8.606e-01  8.234e-02
AR1          2.451e-01  6.649e-02
---
Signif. codes:  0 '***' 0.001 '**'
```

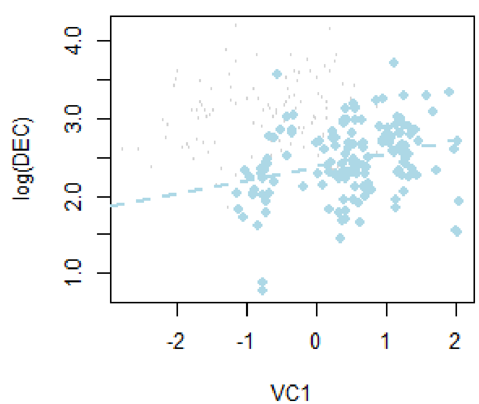


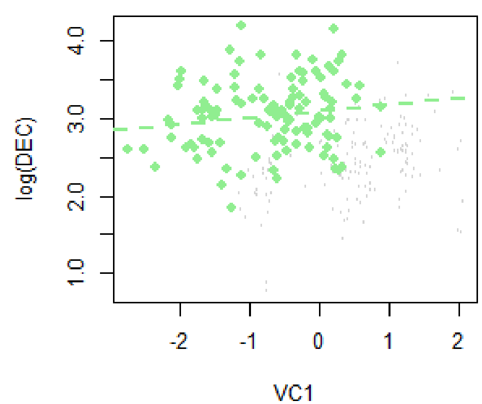
Explanation...

Correlation paradox! Confounding effect!
you name it!

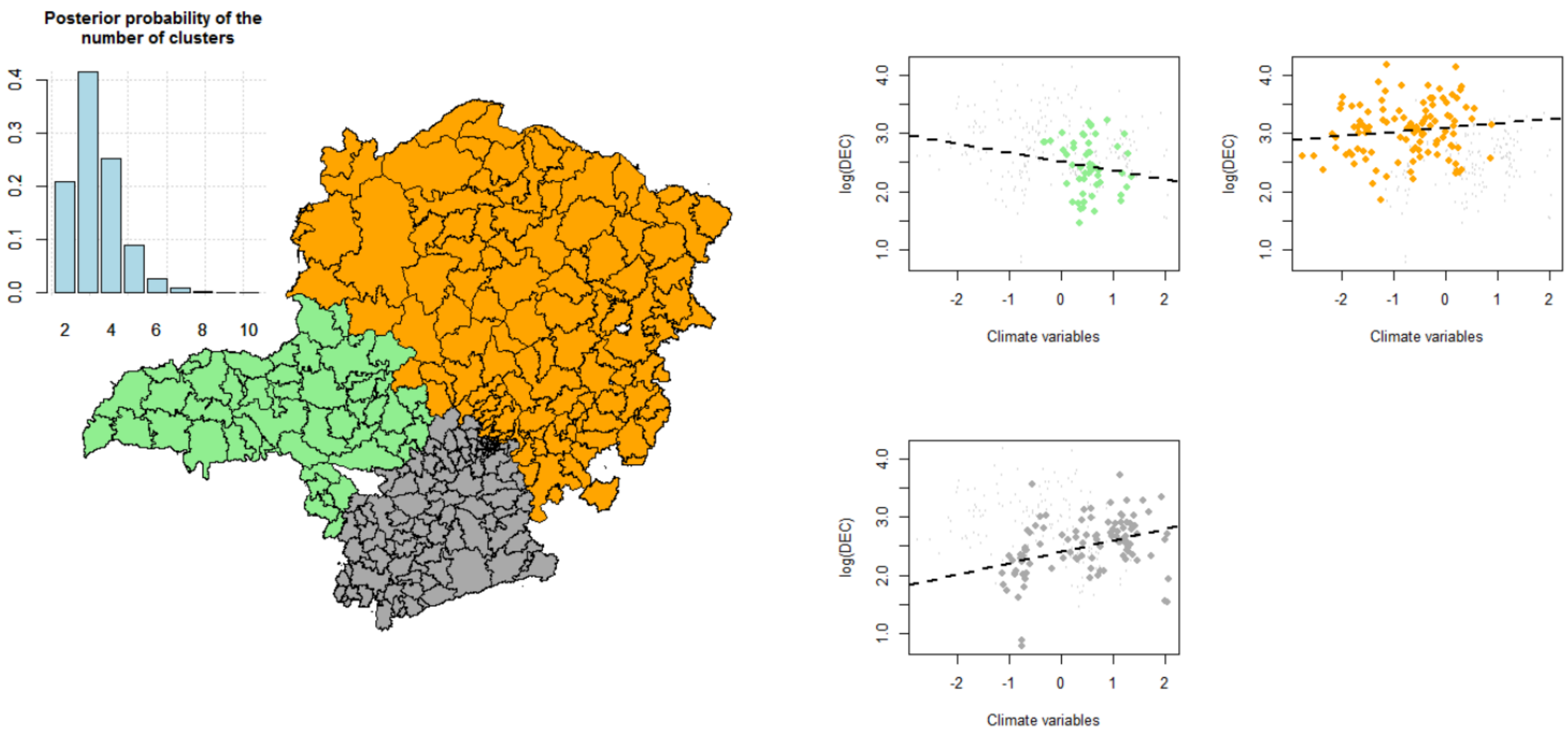








Solution:

- How many geographical clusters?
- 

A novel clustering-based spatial regression model applied to consumer power outage indicator
Marcelo Azevedo Costa, Leandro Brioschi Mineti, Álvaro Lédo Ferreira (working paper)

The proposed Bayesian model

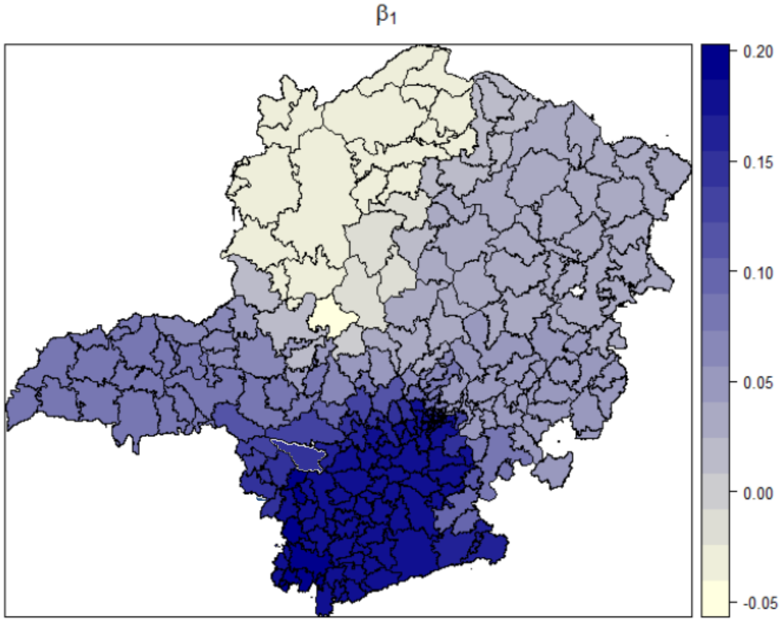
$$Y_i \sim Normal(\beta_{j0} + \beta_{j1} x_i;\ \sigma^2)$$

Priors:

$$k \sim (1-c)^k$$

$$\boldsymbol{\beta}_j | \sigma^2 \sim \mathcal{N}_{p+1}(\mathbf{0}; \sigma^2(\lambda_0 \mathbf{I})^{-1})$$

$$\sigma^2 \sim IG(a_0, b_0)$$



A Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm is proposed to estimate the parameters of the model

Costa, M. A., Mineti, L. B., Mayrink, V. D., & Lopes, A. L. M. (2019). Bayesian detection of clusters in efficiency score maps: An application to Brazilian energy regulation. Applied Mathematical Modelling, 68, 66-81.

What if we are missing something?

- Geographical heterogeneity was adjusted and linear regression models were adjusted for each geographical partition.

Predictor variable	Expected sign	Cluster 1			
		univariate	coefficient	P-value	VIF
Climate variable	positive	0.0671	-0.0186	0.6840	1.49
Demand for electrical services I	positive	-0.0068	-0.0027	0.9764	6.21
Demand for electrical services II	positive	0.1118	0.3720	0.0000	3.93
Electrical assets I	positive	-0.0265	-0.3589	0.0002	9.04
Electrical assets II	negative	0.0059	-0.1601	0.1654	13.38
Geographical assets	positive	0.0531	0.0993	0.1090	4.40
Operational and capital costs	negative	0.0359	0.0980	0.0778	2.70
Sample size: 115 electrical areas					

What if we are missing something?

- Geographical heterogeneity was adjusted and linear regression models were adjusted for each geographical partition.

Cluster 1:
Demand for electrical services II
Eletrical Assets I

Cluster 2:
Demand for electrical services I
Demand for electrical services II
Electrical Assets I
Electrical Assets II
Geographical Assets

Cluster 3:
Demand for electrical services II
Eletrical Assets II
Geographical Assets

Hybrid Models for the DEC indicator

Measurement 146 (2019) 425–436

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

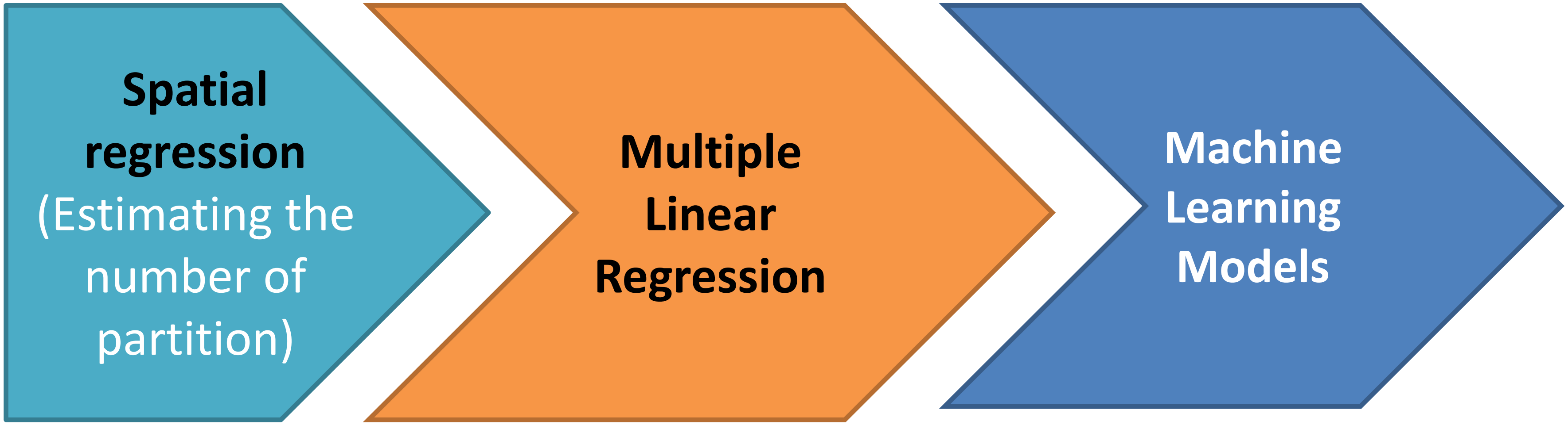
Failure detection in robotic arms using statistical modeling, machine learning and hybrid gradient boosting

Marcelo Azevedo Costa [a,*], Bernhard Wullt [c], Mikael Norrlöf [c,b], Svante Gunnarsson [b]

[a] *Department of Production Engineering, Universidade Federal de Minas Gerais, Brazil*
[b] *Department of Electrical Engineering, Linköping University, Sweden*
[c] *Robotics and Discrete Automation, ABB AB, Sweden*

Failure detection in robotic arms using statistical modeling, machine learning and hybrid gradient boosting

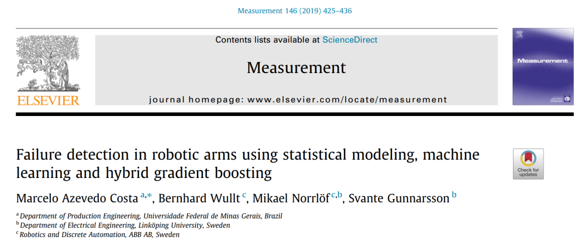
Hybrid Models for the DEC indicator



CART
Random Forest
xgboost
Neural Networks
...

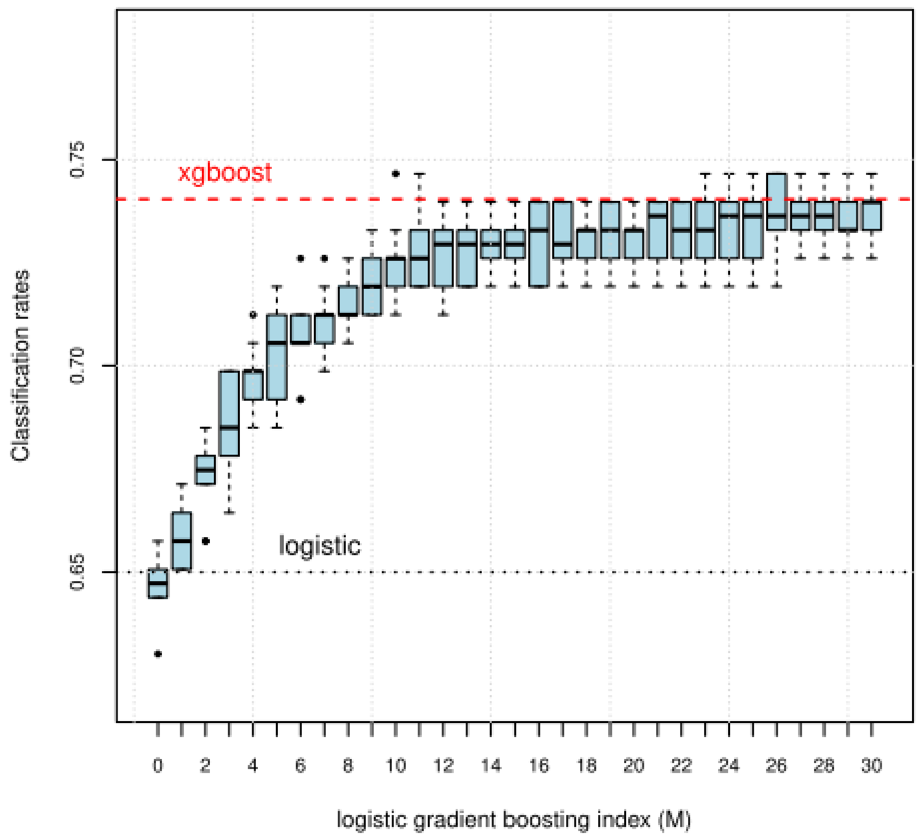
Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Boosting a baseline model


Measurement 146 (2019) 425–436
Contents lists available at ScienceDirect
Measurement
journal homepage: www.elsevier.com/locate/measurement

Failure detection in robotic arms using statistical modeling, machine learning and hybrid gradient boosting

Marcelo Azevedo Costa [a,*], Bernhard Wullt [c], Mikael Norrlöf [c,b], Svante Gunnarsson [b]
[a] Department of Production Engineering, Universidade Federal de Minas Gerais, Brazil
[b] Department of Electrical Engineering, Linköping University, Sweden
[c] Robotics and Discrete Automation, ABB AB, Sweden



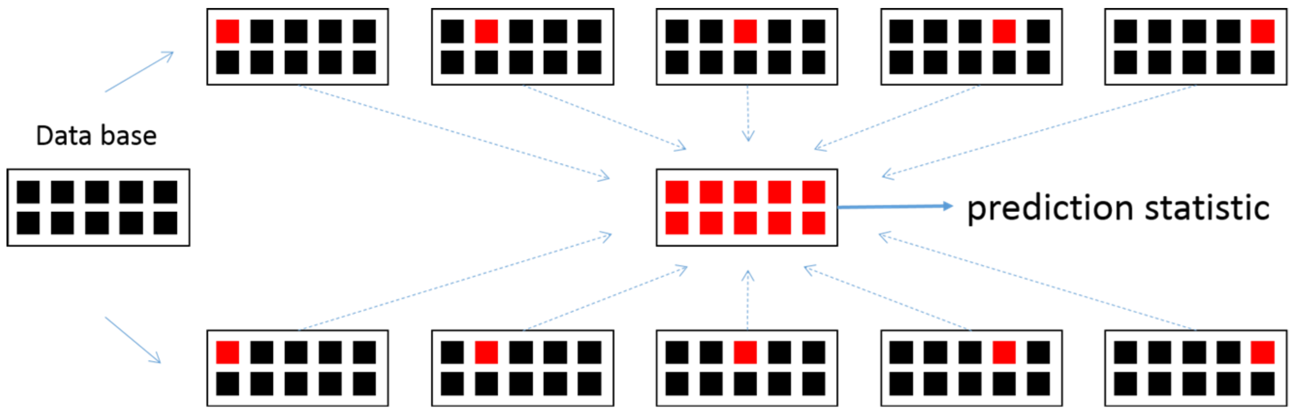
(b) Hybrid gradient boosting with Logistic and Random Forests.

Best model

Base line model

Performance analysis

leave-one-out cross-validation



$$R^2_{pred}$$

Predicted Coefficient of Determination

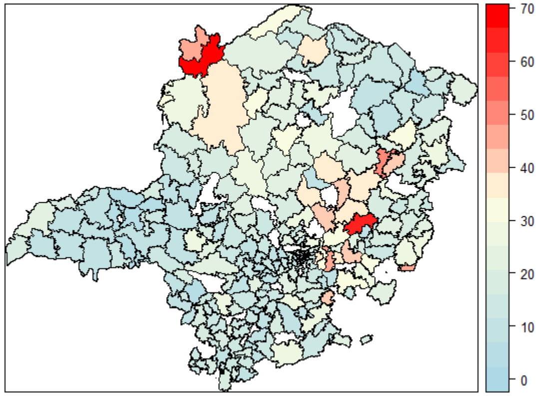
Standard Multiple regression
R^2_{pred} = 53,6%

Spatial Regression
R^2_{pred} = 61,79%

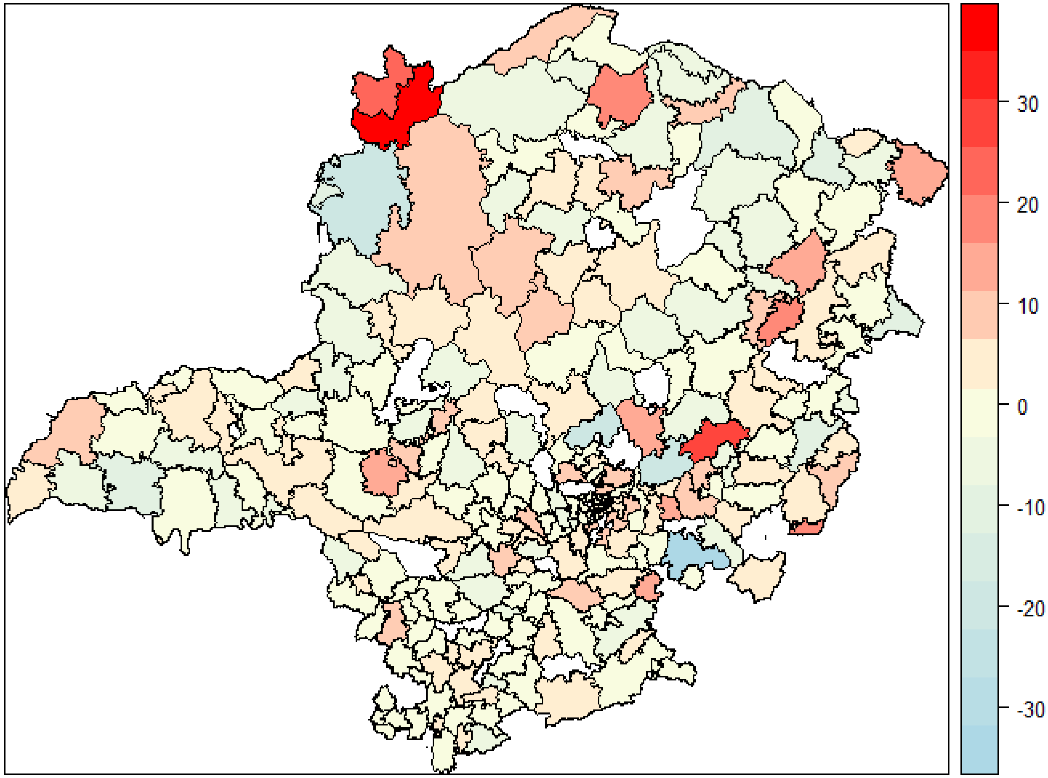
Machine Learning Models
R^2_{pred} = 67,62%

Analysis of the residuals

Log(DEC)



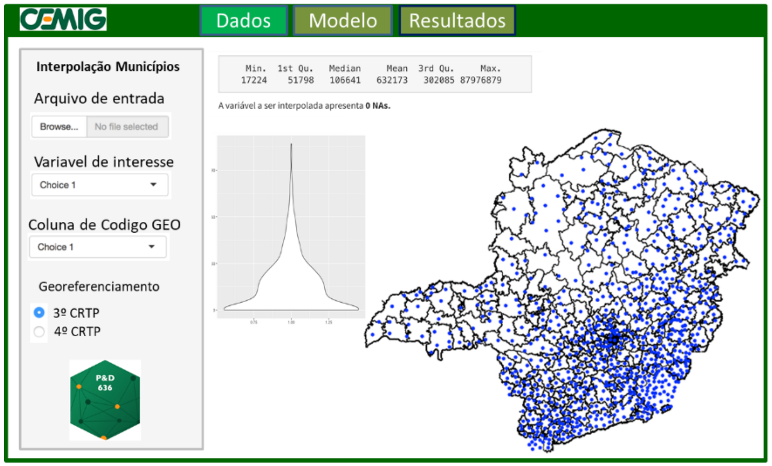
Residuals

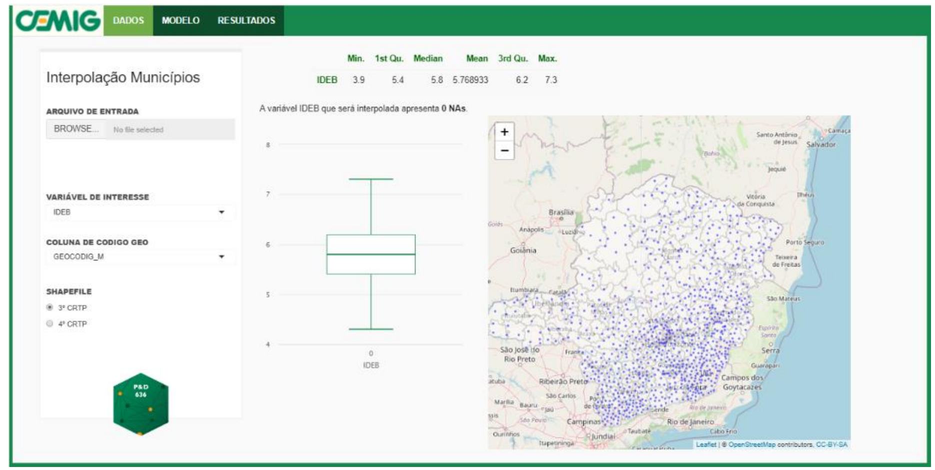


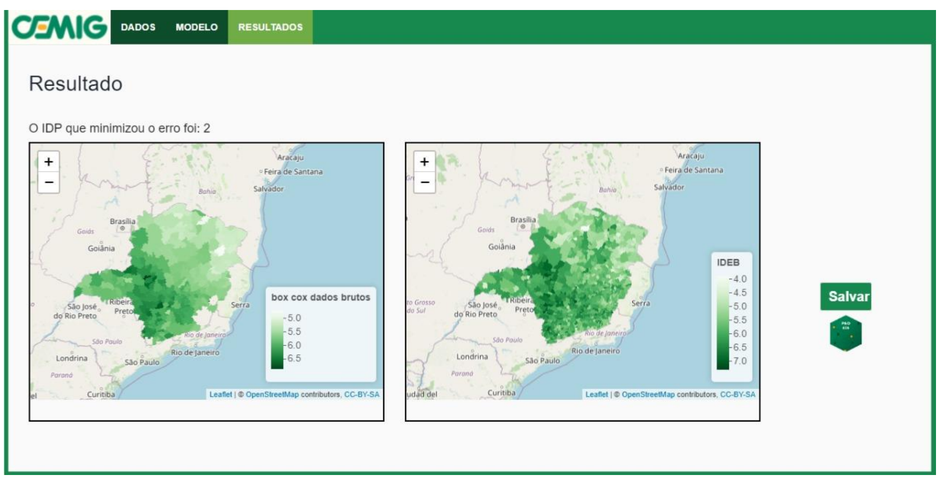
Discussion and conclusion

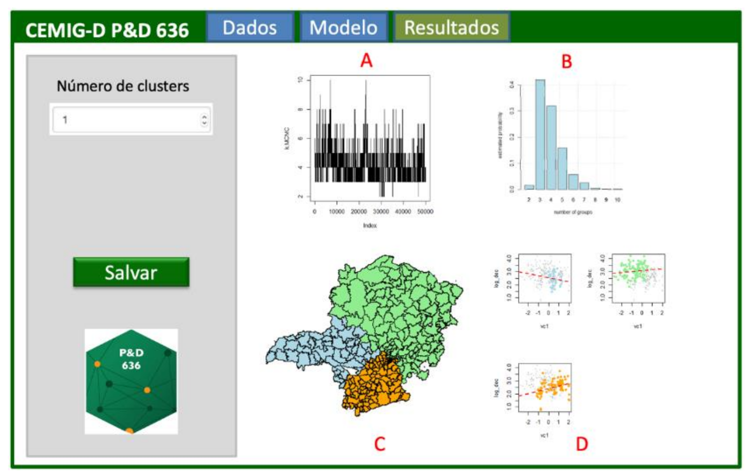
- Hybrid AI – Where data-driven and model-based methods meet
- Hybrid AI – Where data-drive, **technically-based** and model-based methods meet.

Computational tool (screenshots)









Thanks!

Questions?

Transmission Lines in Brazil

https://gisepeprd2.epe.gov.br/WebMapEPE/

ELLIIT

Hybrid AI – Where data-driven and model-based methods meet

Focus Period Linköping 2022

Multilayer Hybrid Models for Power Outage Analysis of a Brazilian Distribution Company

Marcelo Azevedo Costa

Federal University of Minas Gerais

LiU LINKÖPINGS UNIVERSITET

LUND UNIVERSITY

UFMG