PROGRAMA DE PESQUISA E DESENVOLVIMENTO CEMIG / ANEEL

WORKSHOP P&D D0636

Modelagem Estatístico-Computacional do Modelo de Negócio da Cemig Distribuição Utilizando Bases de Dados e Conhecimento Técnico

Modelos Multi Camadas aplicados ao estudo do índice DEC

Prof. Dr. Marcelo Azevedo Costa
Equipe UFMG
Equipe CEMIG-D

UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS — ANEEL Agência Nacional de Energia Elétrica | Programa de Pesquisa e Desenvolvimento — CEMIG NOSSA ENERGIA, SUA FORÇA — MINAS GERAIS GOVERNO DIFERENTE. ESTADO EFICIENTE.

Agenda

I. Análise Exploratória do índice DEC

II. Modelos paramétricos para o índice DEC

III. Modelos multi-camadas para o índice DEC

Statistical Modeling: The Two Cultures

Leo Breiman

Statistical Science, Vol. 16, No. 3 (Aug., 2001), pp. 199-215

a) Focus on finding a good solution – that's what consultants get paid for;

b) Live with the data before you plunge into modeling;

c) search for a model that gives a good solutions, either algorithmic or data;

d) predictive accuracy on test sets is the criterion for how good the model is;

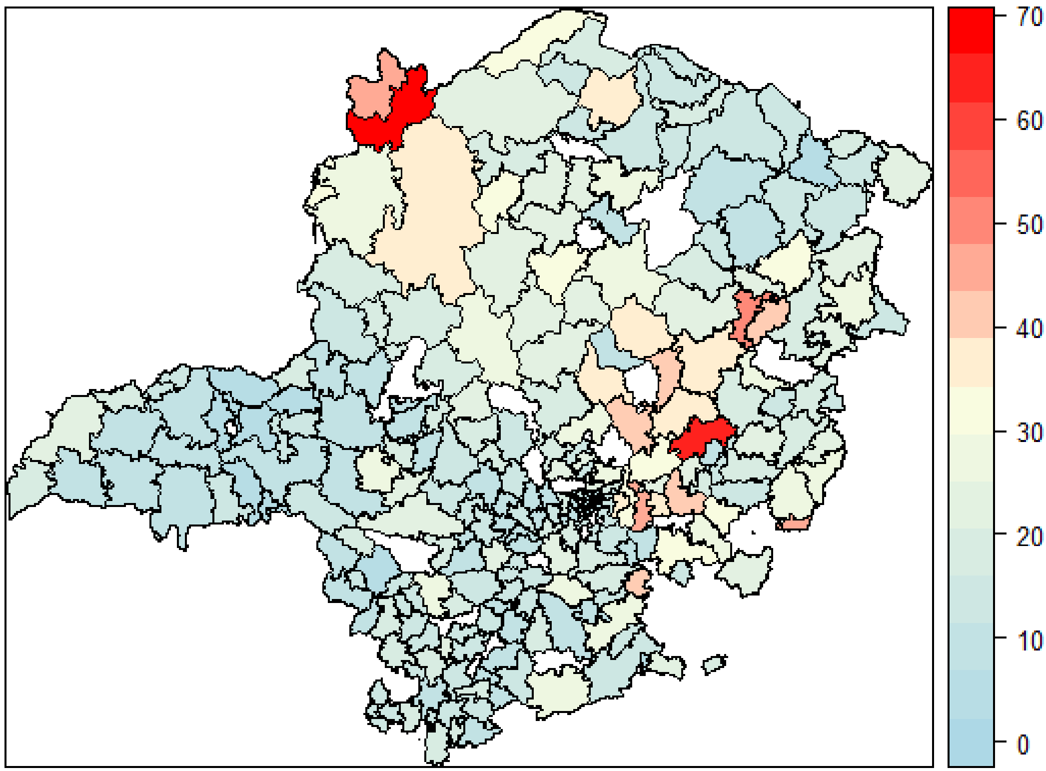
O índice DEC

DEC : **D**uração **E**quivalente de interrupção por unidade **C**onsumidora - Intervalo de tempo que, em média, no período de apuração, em cada unidade consumidora do conjunto considerado ocorreu descontinuidade da distribuição de energia elétrica.

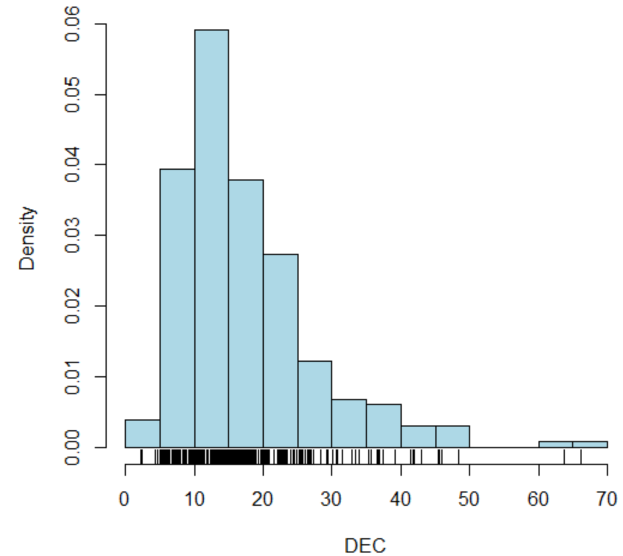
- A continuidade do fornecimento é avaliada pela ANEEL através de subdivisões das distribuidoras, denominadas Conjuntos Elétricos.

https://www2.aneel.gov.br/aplicacoes/srd/indqual/default.cfm

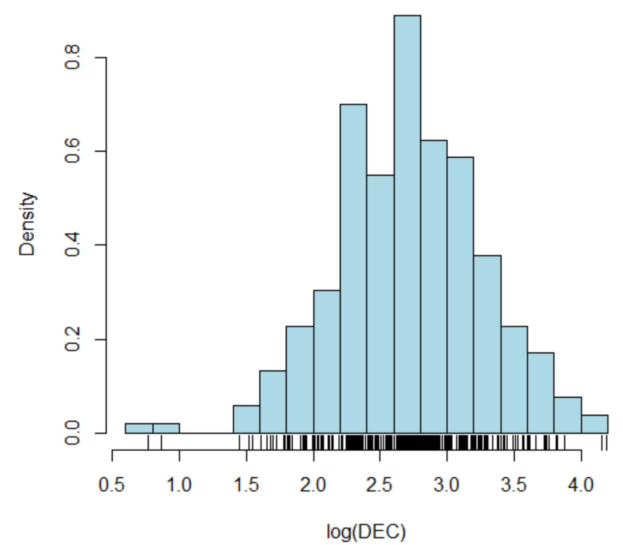
O índice DEC



70
60
50
40
30
20
10
0



Density
DEC
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
10
20
30
40
50
60
70



Density
log(DEC)
0.0
0.2
0.4
0.6
0.8
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Variáveis preditoras selecionadas

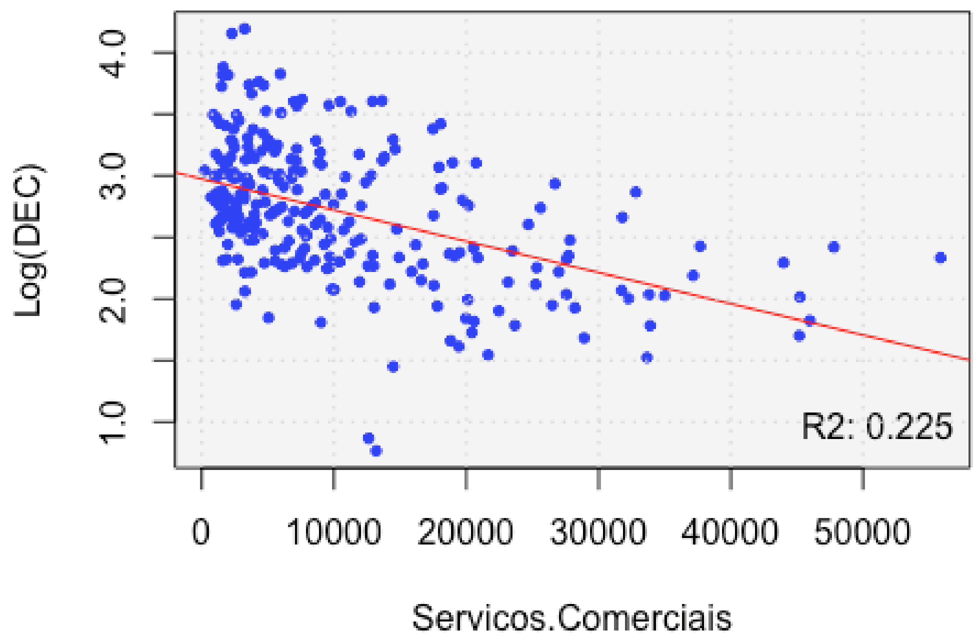
Fatores Latentes	Variáveis que compõem o fator latente
Ativos Geográficos	Área de atendimento (km^2)
	Extensão de estradas na área de atendimento (km)
	Número de municípios atendidos
	Número de locais atendidos (por definição da CEMIG)
Ativos Elétricos 1	Extensão das linhas de distribuição (km)
	Extensão da rede de distribuição (km)
	Número total de clientes atendidos
Ativos Elétricos 2	Número de subestações
	Número de equipamentos de proteção
	Número de equipamentos automatizados
Variáveis Climáticas	Índice de umidade no período (%)
	Temperatura média no período (^oC)
	Quantidade de chuva (mm)
Demanda de Serviços 1	Número de equipes de trabalho utilizada
	Número de serviços comerciais realizados
	Número de serviços emergenciais realizados
Demanda de Serviços 2	Número de interrupções devido a queda de árvores nas linhas de distribuição
	Número de interrupções devido a queda de árvores nas subestações
	Número de interrupções devido a queda de árvores nas redes de distribuição
Aplicação de Recursos	Capital gasto com OPEX (Operational Expenditures - R$)
	Capital gasto com CAPEX (Capital Expenditures - R$)

Variáveis preditoras (25)

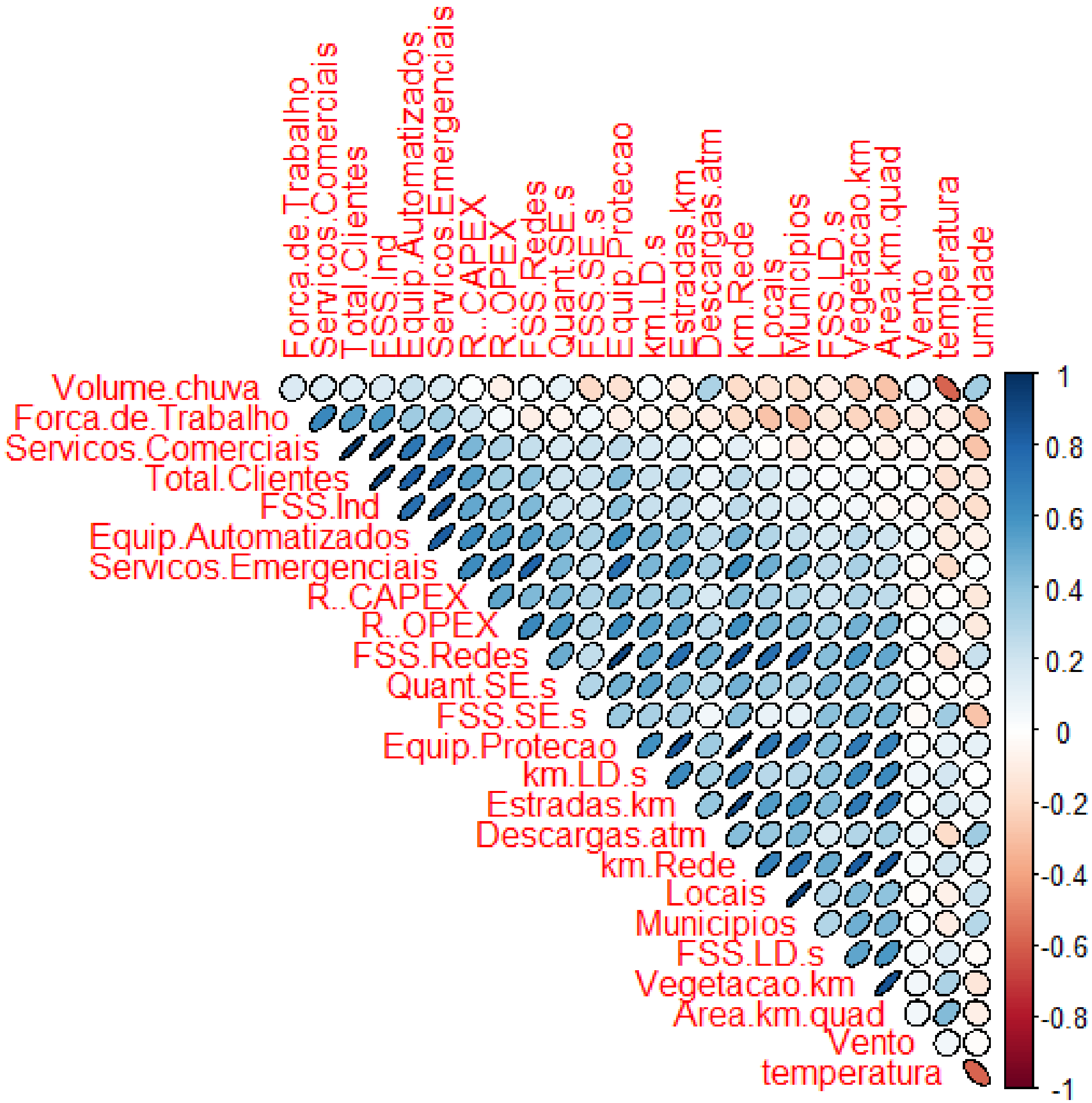
Variável preditora	Coeficiente	valor-P	R^2
Servicos.Comerciais	-0,000025	0,0000	0,2254
Total.Clientes	-0,000010	0,0000	0,2090
Equip.Automatizados	-0,007731	0,0000	0,1468
Municipios	0,070730	0,0000	0,1267
Forca.de.Trabalho	-0,000166	0,0000	0,1217
FSS.LD.s	0,083660	0,0000	0,1192
Volume.chuva	-0,001959	0,0000	0,1182
FSS.Ind	-0,000237	0,0000	0,1078
Area.km.quad	0,000071	0,0000	0,0987
Locais	0,056310	0,0000	0,0888
Vegetacao.km	0,000645	0,0000	0,0836
FSS.Redes	0,000142	0,0001	0,0597
km.Rede	0,000093	0,0001	0,0559
R..OPEX	0,000000	0,0011	0,0397
temperatura	0,094450	0,0017	0,0365
Equip.Protecao	0,000179	0,0113	0,0240
Estradas.km	0,000015	0,0363	0,0164
R..CAPEX	0,000000	0,0411	0,0156
Servicos.Emergenciais	-0,000018	0,0477	0,0147
Vento	-0,139500	0,1236	0,0089
FSS.SE.s	0,016840	0,2686	0,0046
Quant.SE.s	0,025150	0,4115	0,0025
Descargas.atm	0,000005	0,6313	0,0009
km.LD.s	0,000062	0,9086	0,0000
umidade	0,000598	0,9379	0,0000

$$\log(DEC) = \beta_0 + \beta_1 x$$

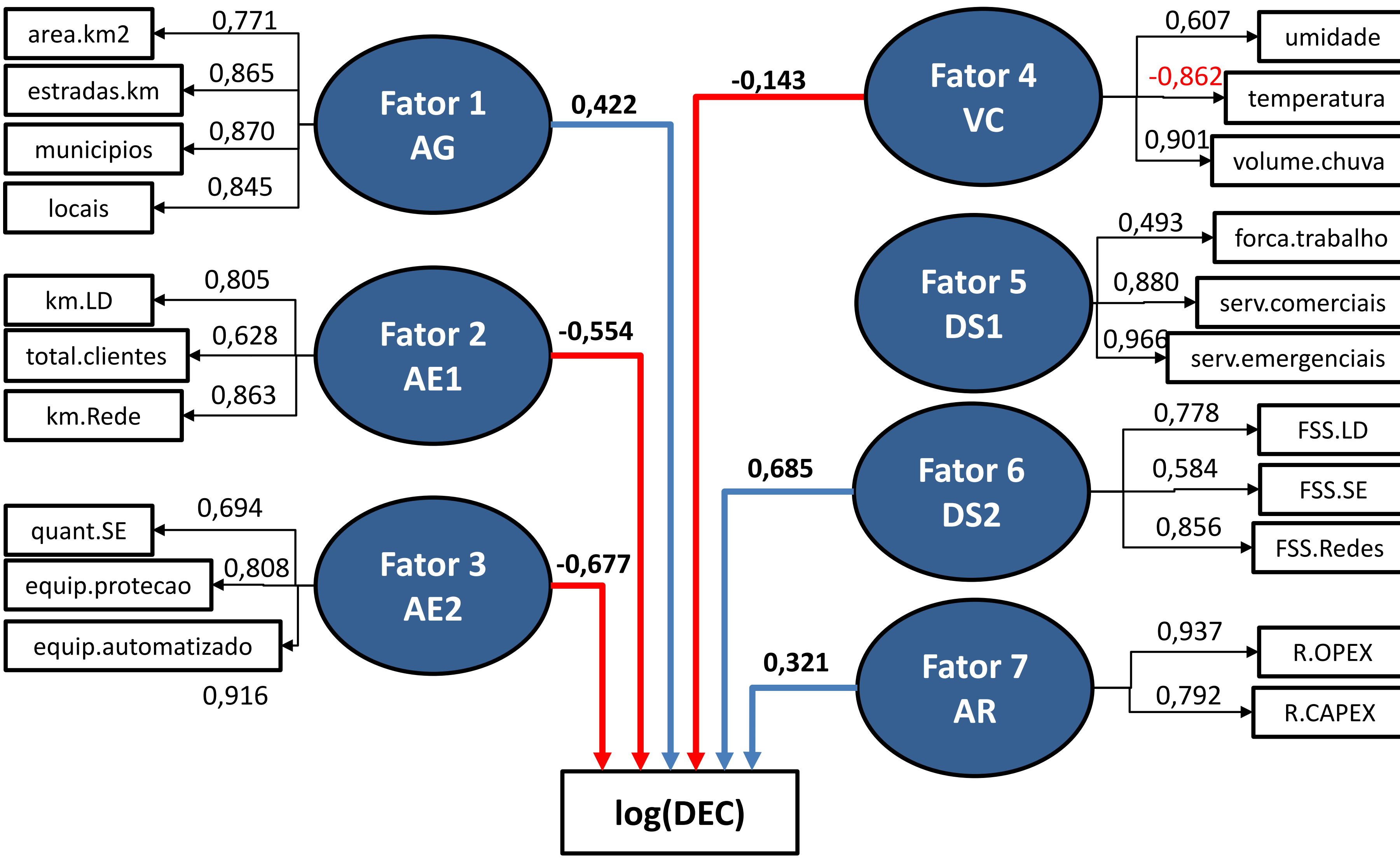
$$DEC = e^{\beta_0 + \beta_1 x}$$



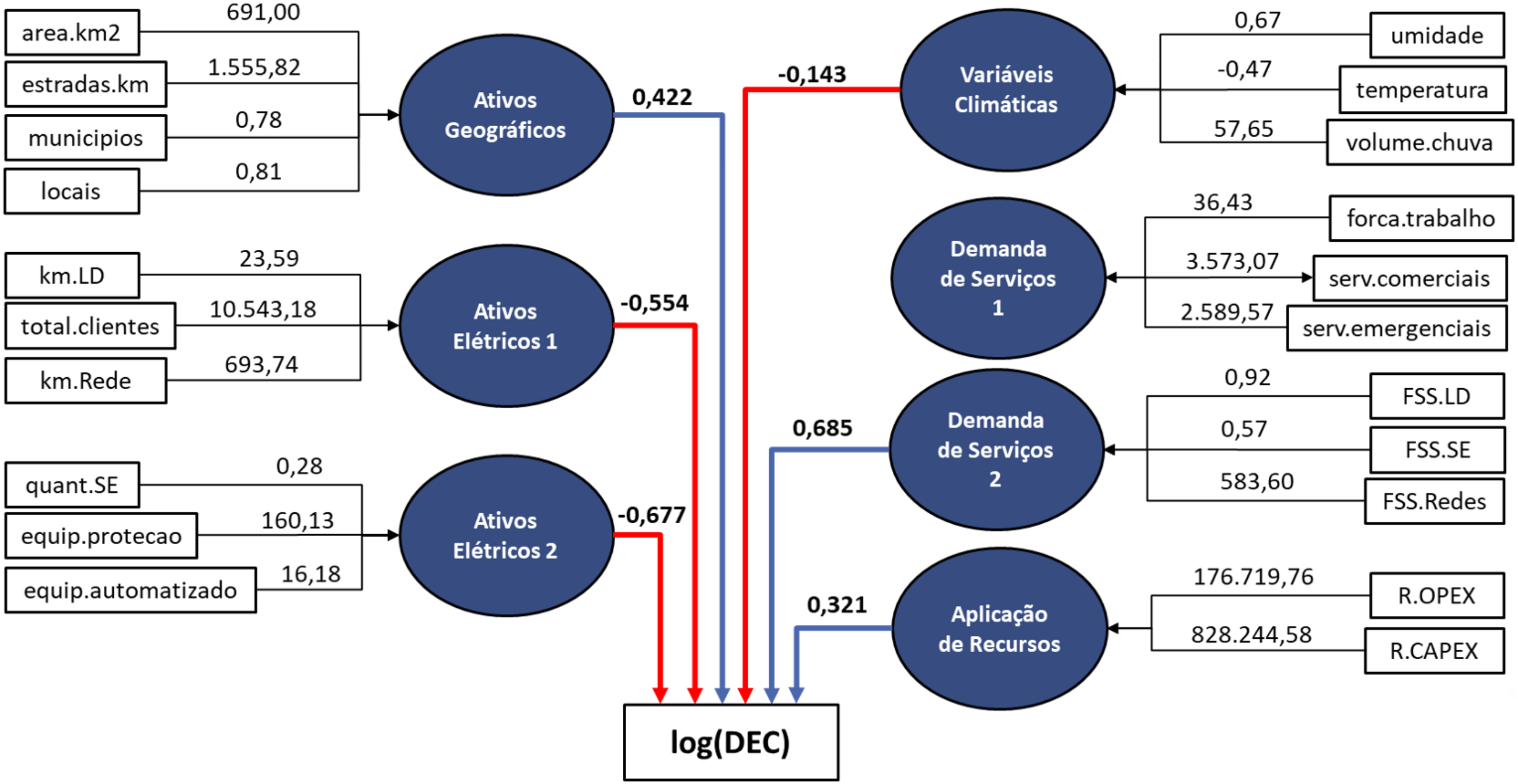
Variáveis preditoras (25)



O Modelo de Equações Estruturais (MEE)



O Modelo de Equações Estruturais (MEE)



Modelo de Regressão - MEE

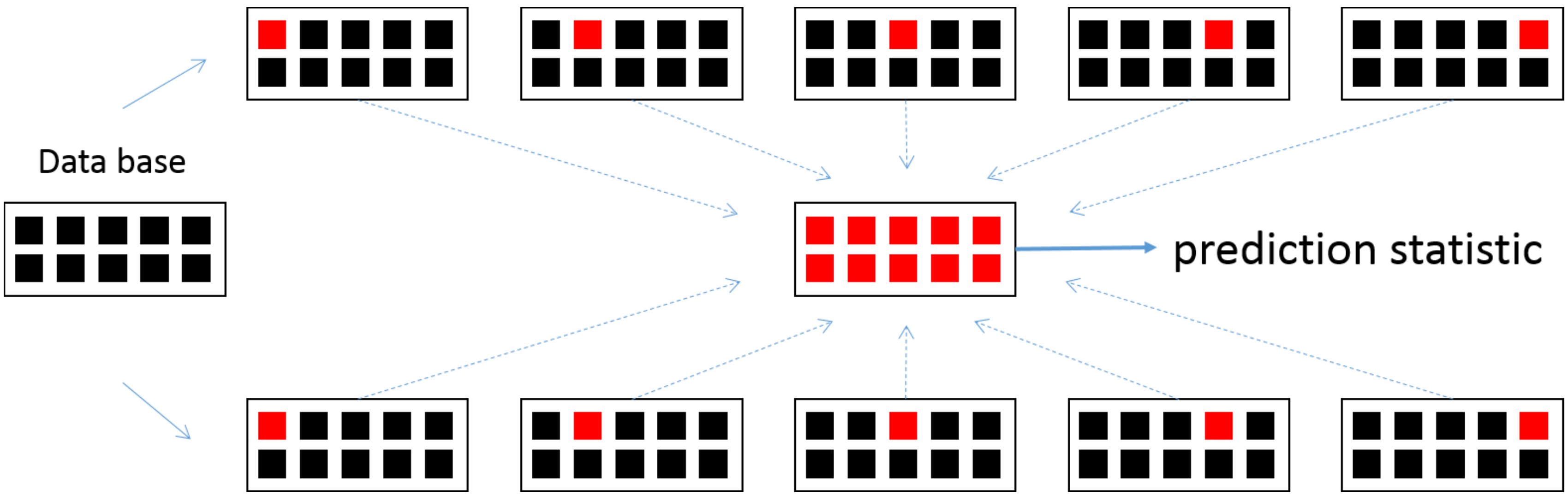
```
lm(formula = DQ ~ AG1 + AE1 + AE2 + VC1 + DS1 + DS2 + AR1, data = dados)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.103e-17  4.066e-02   0.000 1.000000
AG1          3.645e-01  8.198e-02   4.446 1.30e-05 ***
AE1         -5.087e-01  1.095e-01  -4.647 5.36e-06 ***
AE2         -6.563e-01  1.132e-01  -5.796 1.96e-08 ***
VC1         -1.831e-01  4.265e-02  -4.293 2.49e-05 ***
DS1         -1.297e-01  7.985e-02  -1.624 0.105596
DS2          8.606e-01  8.234e-02  10.451  < 2e-16 ***
AR1          2.451e-01  6.649e-02   3.686 0.000277 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Multiple R-squared: 0.5702

Adjusted R-squared: 0.5586

Validação cruzada *k*-fold



predictive accuracy on test sets is the criterion for how good the model is

No caso do DEC, estamos utilizando validação cruzada do tipo **leave-one-out** (n=264)

R^2 Preditivo para o modelo MEE

```
## Calculando o R2Preditivo
yhat <- rep(NA, nrow(dados))
for(cont in 1:nrow(dados)){
  modelo <- lm(DQ ~ AG1 + AE1 + AE2 + VC1 + DS1 + DS2 + AR1,
               data=dados[-cont,])
  yhat[cont] <- predict(modelo, newdata=dados[cont,])
}
R2pred(dados$DQ, yhat)
```

```
>  R2pred(dados$DQ, yhat)
[1] 0.5346226
```

```
>  R2pred(dados$DQ, yhat)
[1] 0.5641286
```

Modelo com TODAS as 25 variáveis

R^2 Preditivo para o modelo Ridge

- Existem técnicas estatísticas para regularização de um modelo com várias variáveis de forma a melhorar o R^2 Preditivo

```
>  R2pred(dados$DQ, yhat)
[1] 0.5673525
```

Modelo com TODAS as
25 variáveis e regularização

Modelos Híbridos Multi-Camadas

Measurement 146 (2019) 425–436

Contents lists available at ScienceDirect

Measurement

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

ELSEVIER

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

Check for updates

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

LINKÖPING UNIVERSITY


ABB ROBOTICS


LINK-SIC

UFMG

CISB CENTRO DE PESQUISA E INOVAÇÃO SUECO-BRASILEIRO

SAAB

Hybrid Gradient Boosting
Modelos Multi-Camadas

Regionalização Espacial

Random Forests/xgboost

CART – Classification and Regression Trees

Modelos de Equações Estruturais

Layer 1

Layer 2

Layer 3

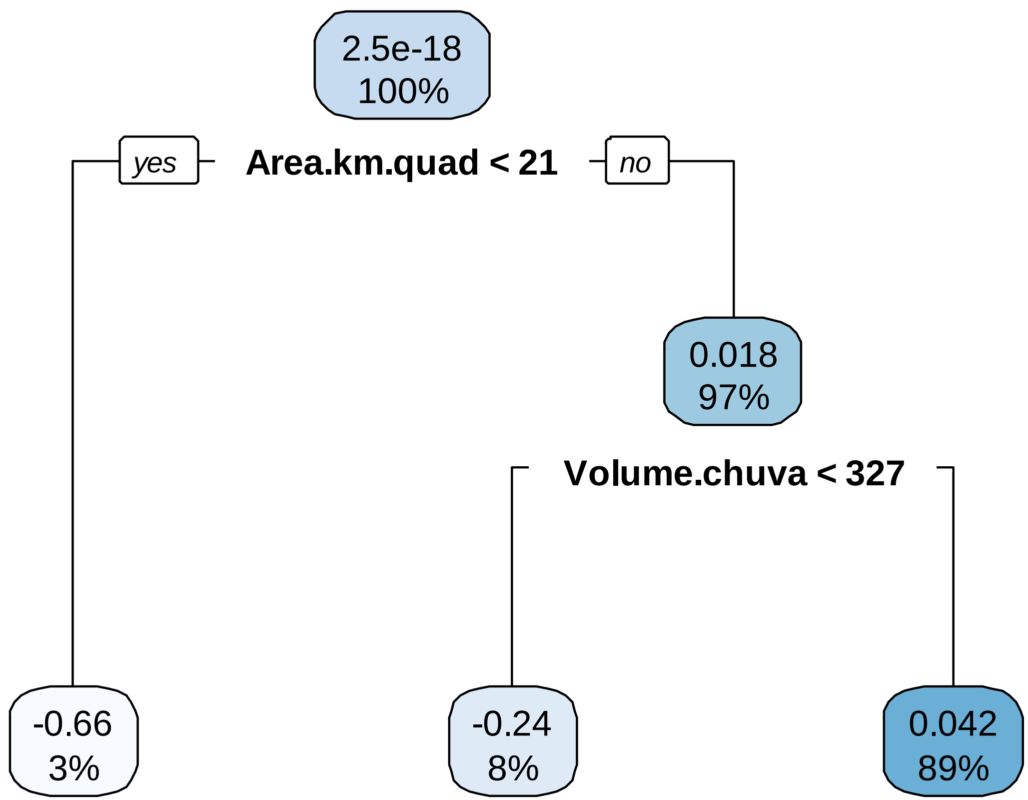
... Layer M

Modelo Multi-Camadas: Regressão + CART (Árvore de Regressão)


```
>  R2pred(dados$DQ, yhat)
[1] 0.5713538
```

Modelo Multi-Camadas
25 variáveis e regularização

Modelo Multi-Camadas: MEE + CART (Árvore de Regressão)




```
>  R2pred(dados$DQ, yhat)
[1] 0.5713538
```

Modelo MEE + CART
(CART com 25 variáveis)

Modelos Multi-Camadas

- **MEE + xgboost**

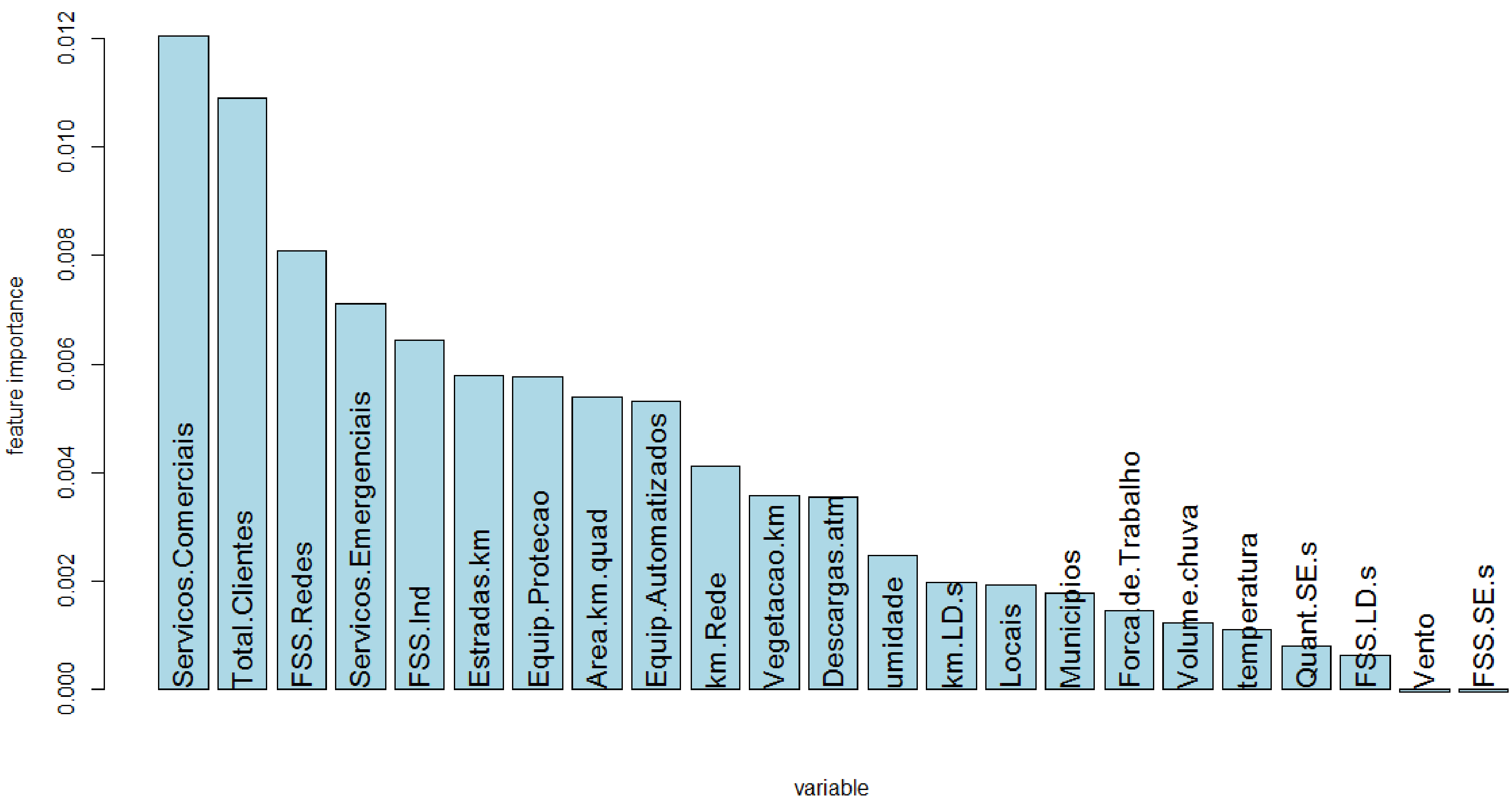
```
>  R2pred(dados$DQ, yhat)
[1] 0.6431925
```

- **MEE + Random Forest**

```
>  R2pred(dados$DQ, yhat)
[1] 0.6494157
```

Ajuste do Modelo de Random Forest

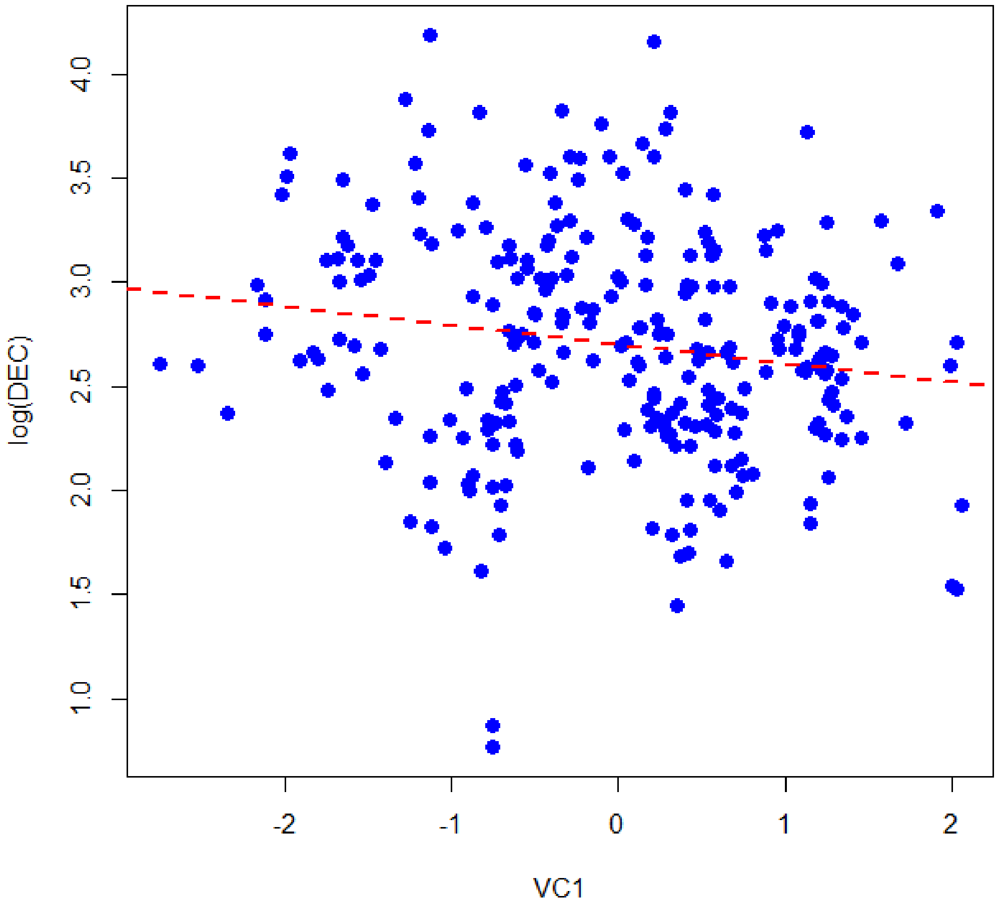
Feature Importance



Modelos Univariados versus
Feature Importance

Variável preditora	Coeficiente	R^2
Servicos.Comerciais	-0,000025	0,2254
Total.Clientes	-0,000010	0,2090
Equip.Automatizados	-0,007731	0,1468
Municipios	0,070730	0,1267
Forca.de.Trabalho	-0,000166	0,1217
FSS.LD.s	0,083660	0,1192
Volume.chuva	-0,001959	0,1182
FSS.Ind	-0,000237	0,1078
Area.km.quad	0,000071	0,0987
Locais	0,056310	0,0888
Vegetacao.km	0,000645	0,0836
FSS.Redes	0,000142	0,0597
km.Rede	0,000093	0,0559
R..OPEX	0,000000	0,0397
temperatura	0,094450	0,0365
Equip.Protecao	0,000179	0,0240
Estradas.km	0,000015	0,0164
R..CAPEX	0,000000	0,0156
Servicos.Emergenciais	-0,000018	0,0147
Vento	-0,139500	0,0089
FSS.SE.s	0,016840	0,0046
Quant.SE.s	0,025150	0,0025
Descargas.atm	0,000005	0,0009
km.LD.s	0,000062	0,0000
umidade	0,000598	0,0000

Variável Climática (VC)



Variável Climática (VC)

```
Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.00781 -0.34300  0.01288  0.32557  1.46836 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.70449    0.03331  81.179  < 2e-16 ***
VC          -0.09090    0.03358  -2.707  0.00724 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5403 on 261 degrees of freedom
Multiple R-squared:  0.02731,    Adjusted R-squared:  0.02358 
F-statistic: 7.327 on 1 and 261 DF,  p-value: 0.007241
```

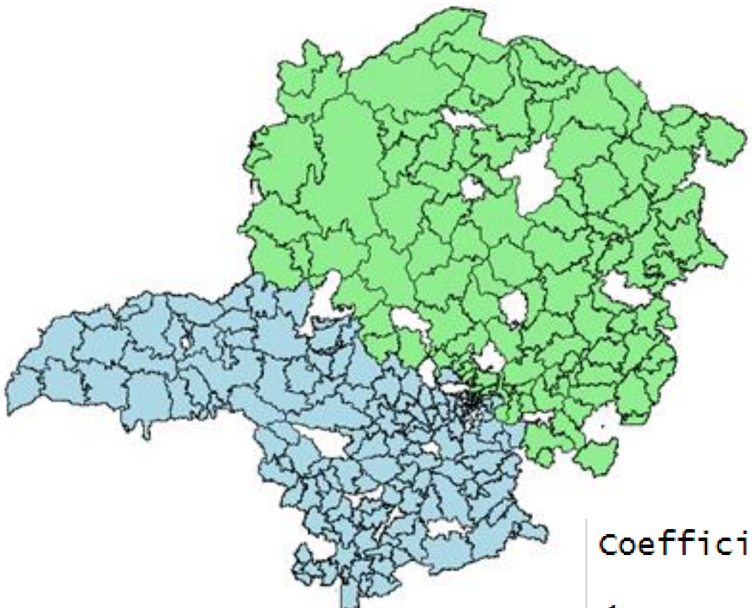
working paper

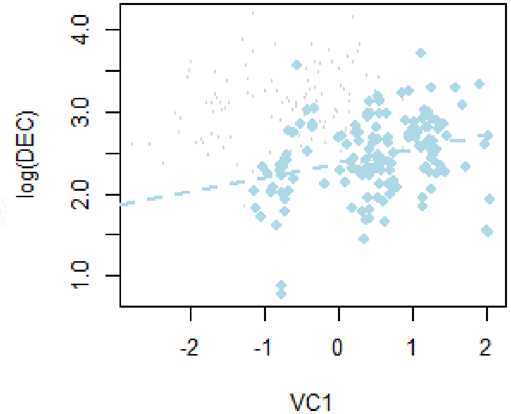
A novel clustering-based spatial regression model applied to consumer power outage indicator

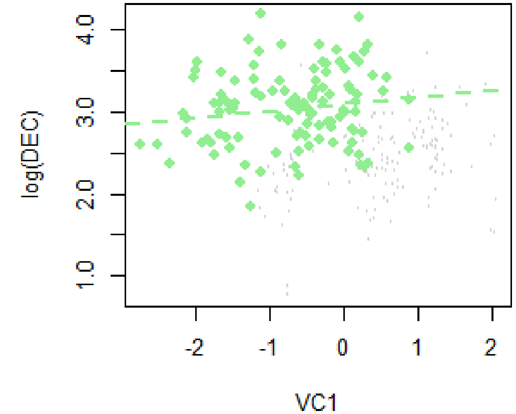
Marcelo Azevedo Costa[a], Leandro Brioschi Mineti[a], Álvaro Lédo Ferreira[a]

[a]*Department of Industrial Engineering, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil*

Regressão Espacial
Variáveis Climáticas







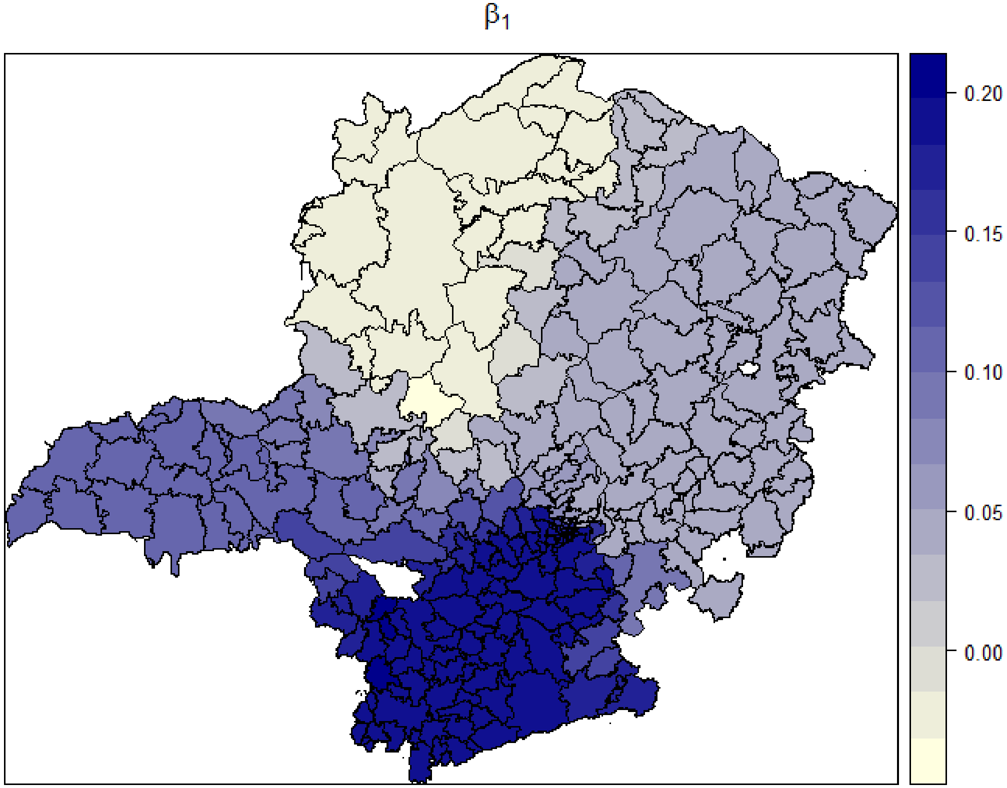

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.37211    0.04420  53.664  < 2e-16 ***
x.aux        0.17277    0.04764   3.627 0.000393 ***
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.09959    0.05703  54.349   <2e-16 ***
x.aux        0.08275    0.05312   1.558    0.122
```

Paradoxo de Simpson

Regressão Espacial
Variáveis Climáticas

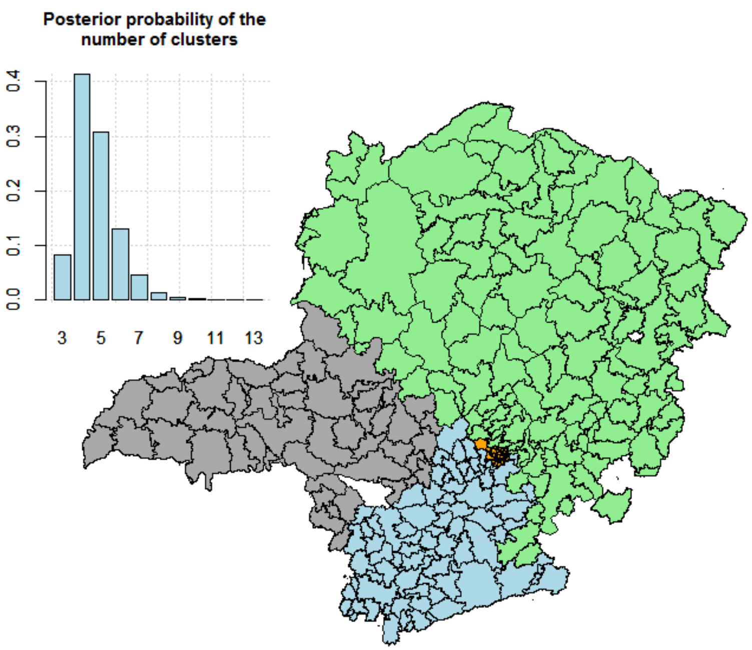
$$Y_i \sim Normal\left(\beta_{j0} + \beta_{j1} x_i;\ \sigma^2\right)$$

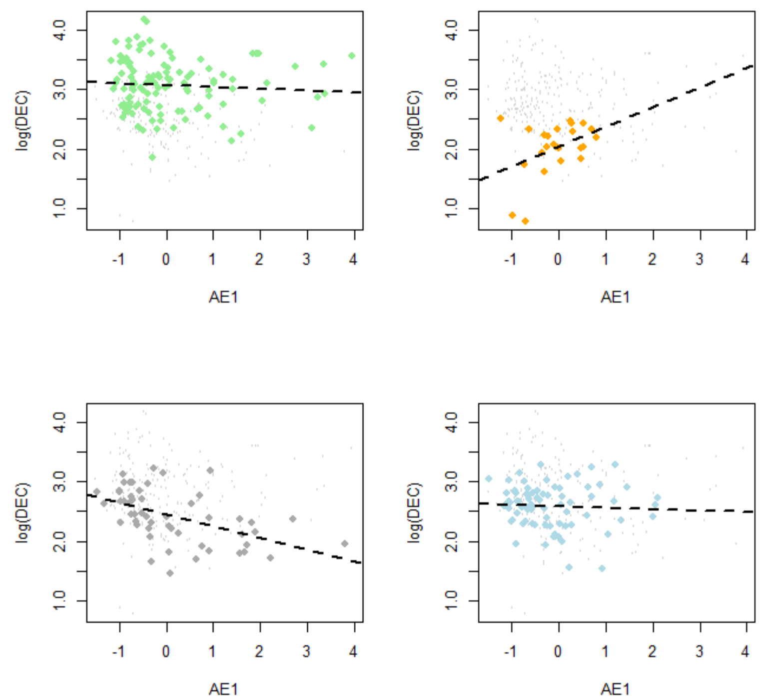


Relatório 6 - Modelos Espaciais de Regressão Univariada

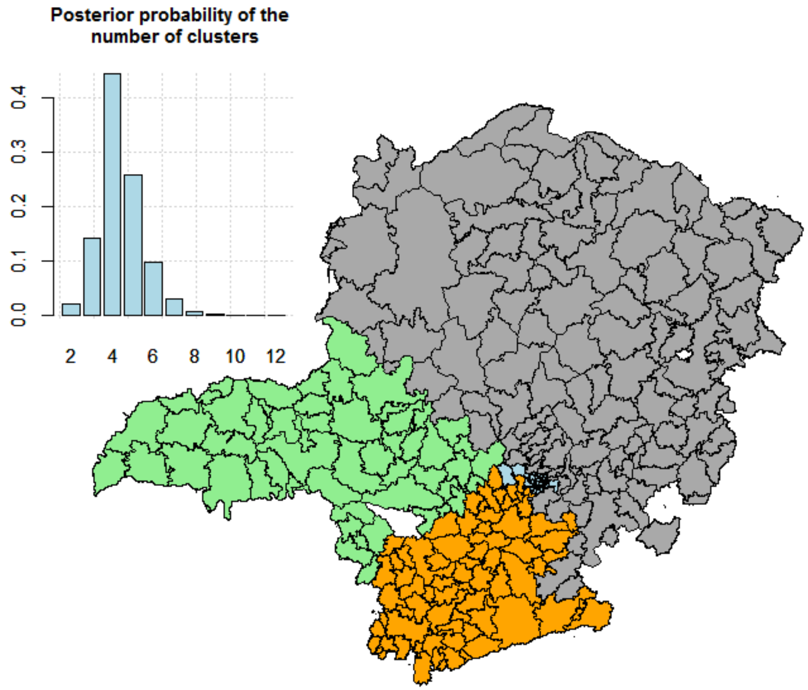
$$Y_i = \beta_{j0} + \beta_{j1} x_i + \epsilon_i$$

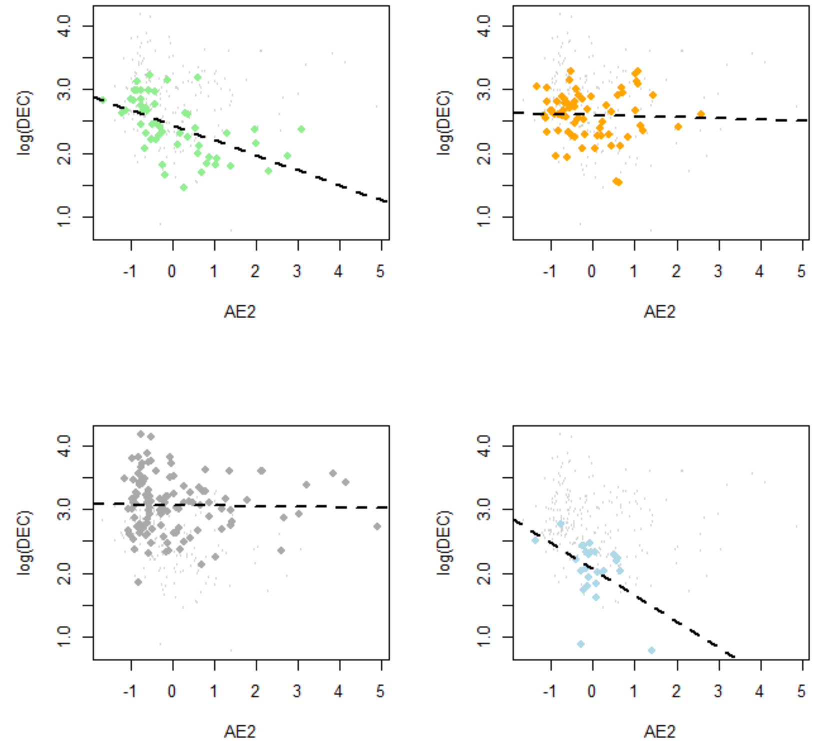
Ativos Elétricos 1 (AE1)



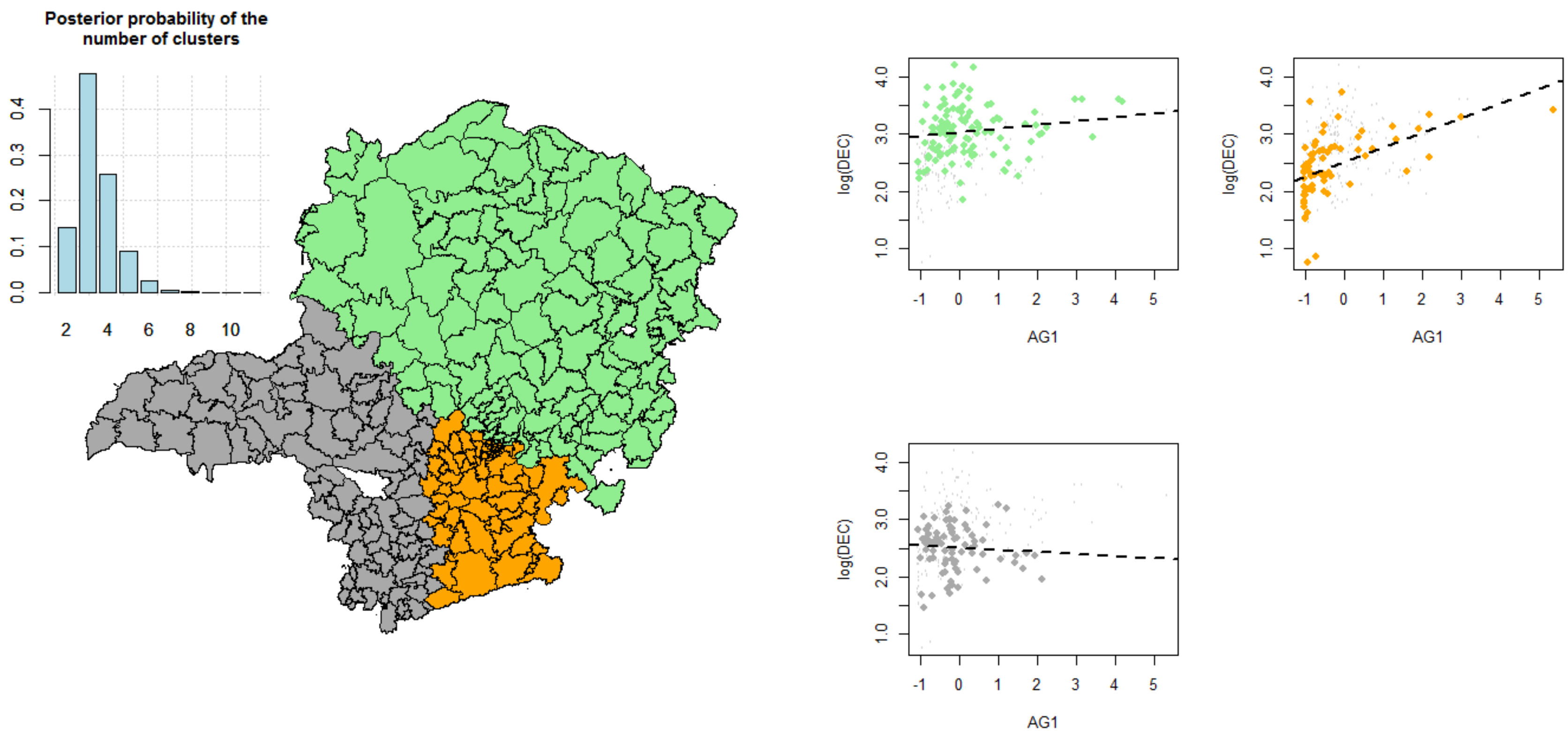


Ativos Elétricos 2 (AE2)

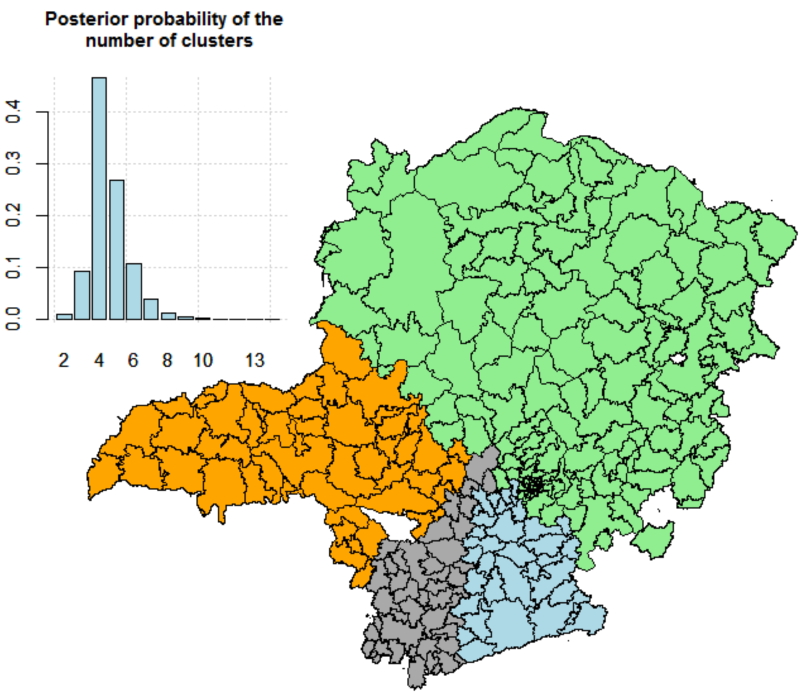


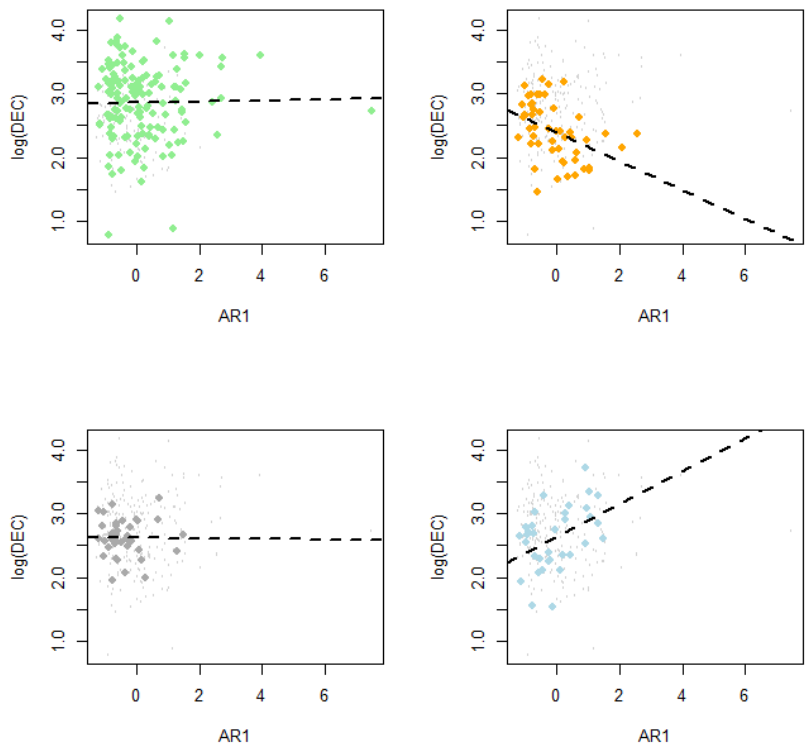


Ativos Geográficos (AG)

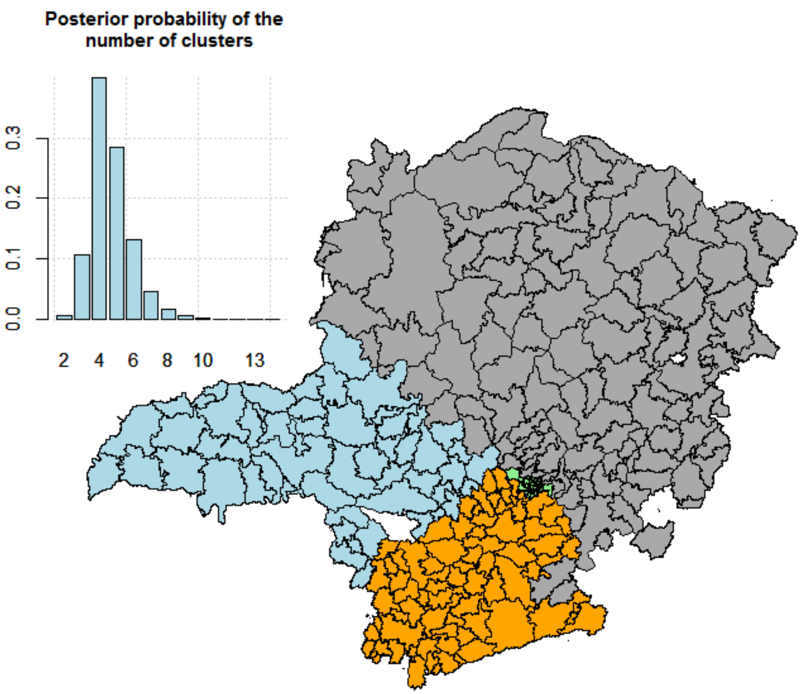


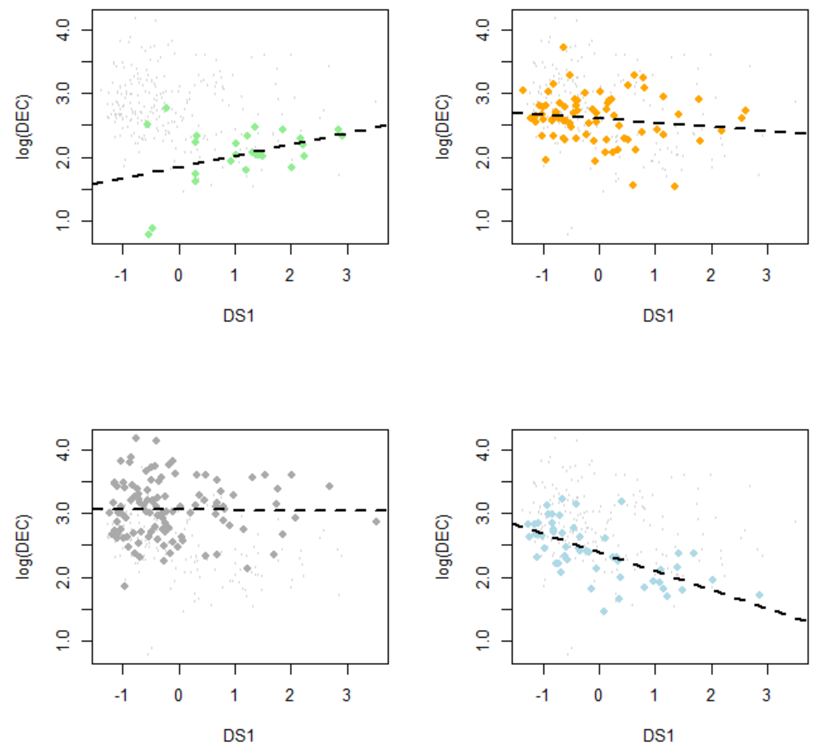
Aplicação de Recursos (AR)



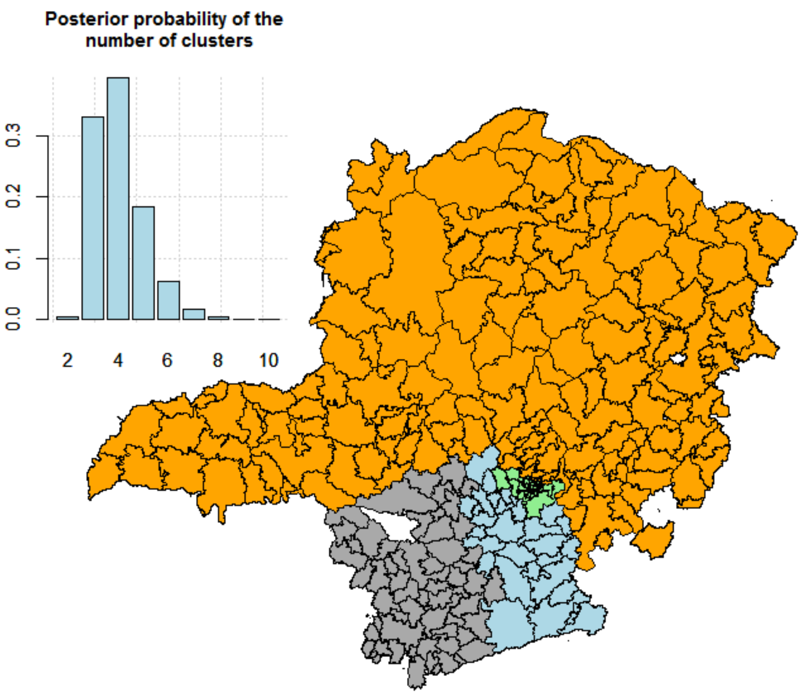


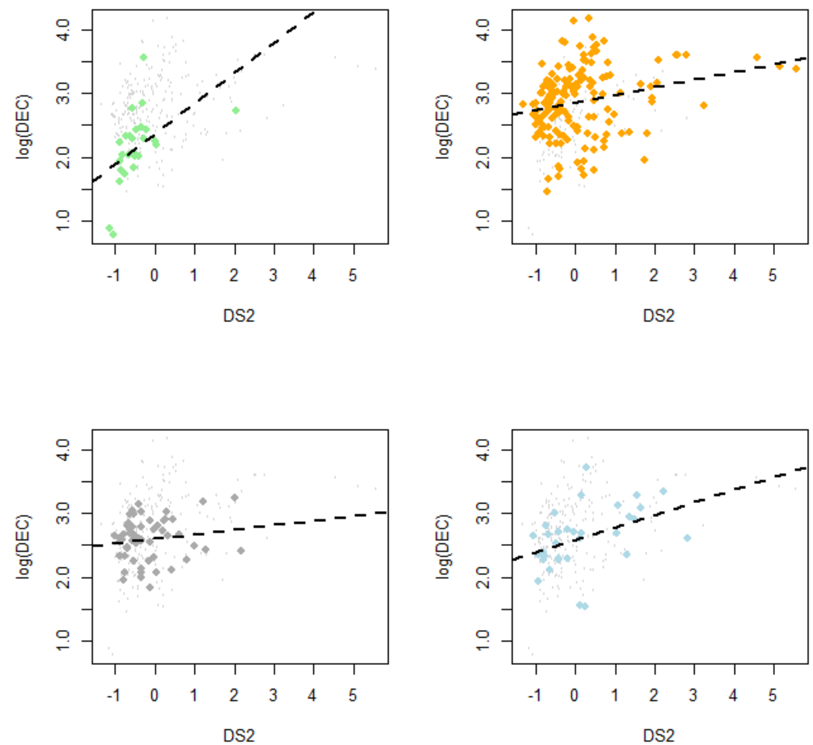
Demanda de Serviços (DS1)



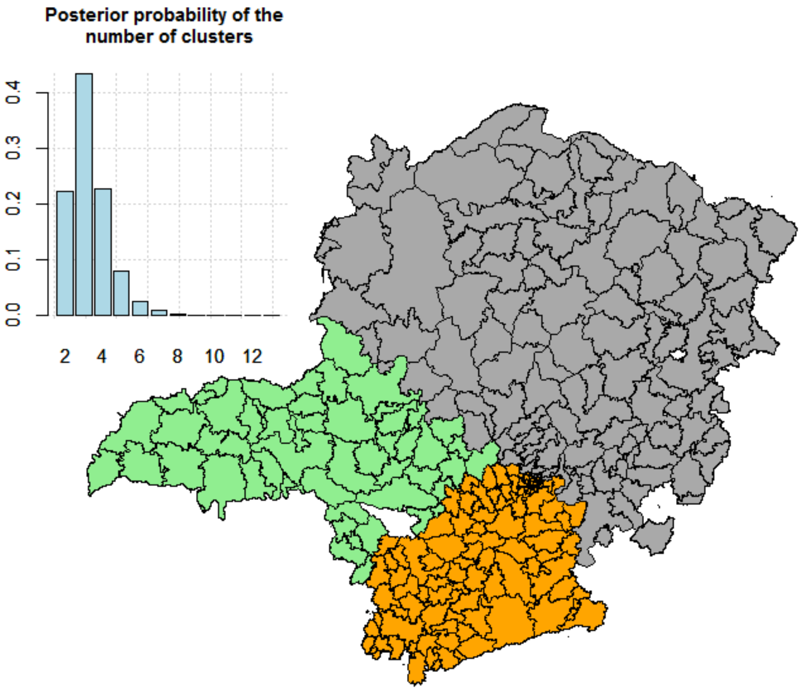


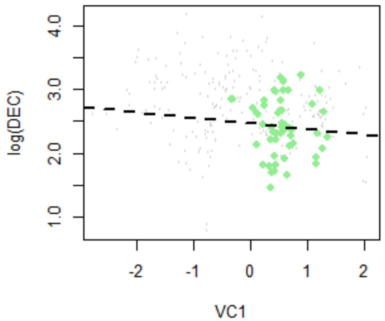
Demanda de Serviços 2 (DS2)

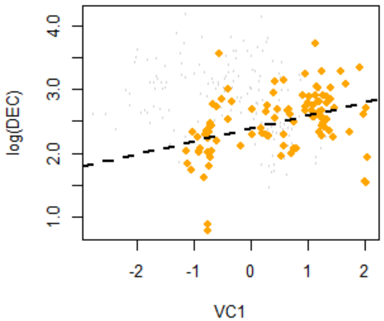


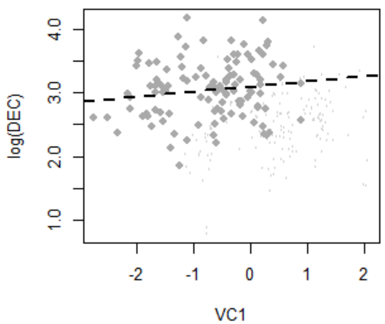


Variáveis Climáticas (VC)

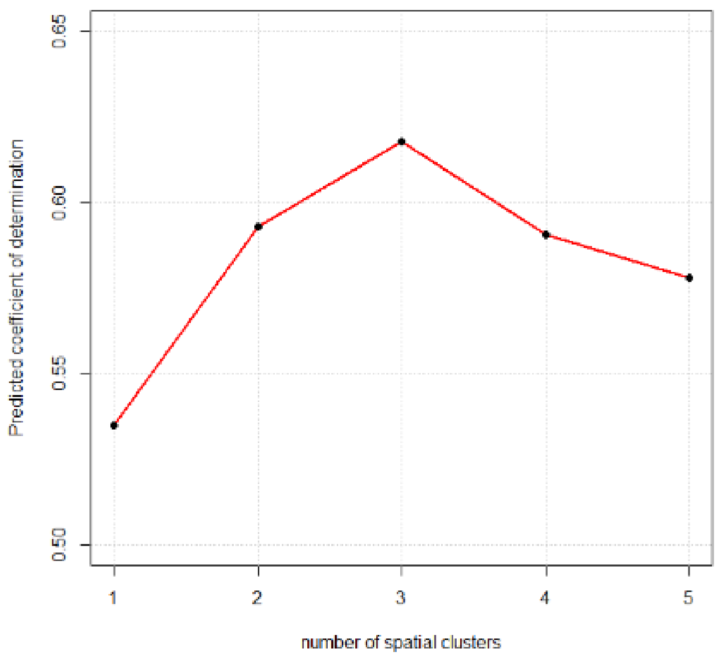




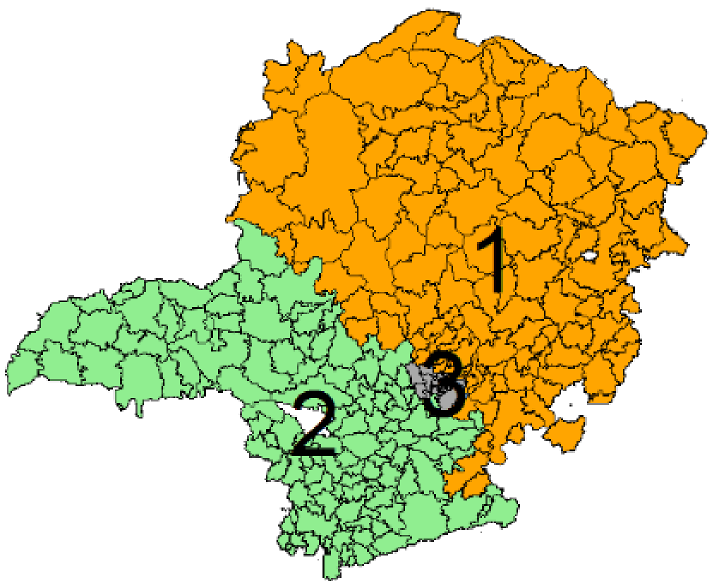




Partições estimadas + MEE



(a) Predictice coefficient of determination (R^2_{pred}) for different number of clusters.



(b) Electrical areas divided into 3 spatial clusters.

Modelo Multi-Camadas: Regressão Espacial Univariada + MEE


```
>  R2pred(y, yhat)
[1] 0.6179093
```

Grupo	R2pred
1	0.2980404
2	0.5065615
3	0.3598554

Modelo Multi-Camadas: Regressão Espacial Univariada + MEE + Random Forests


```
>  R2pred(y, yhat)
[1] 0.6761937
```

Grupo	R2pred (MEE)	R2pred (MEE + RF)
1	0.2980404	0.4517246
2	0.5065615	0.5502779
3	0.3598554	0.3931501

0.6744018 (MEE + Random Forests + GBDCD)