

ANOVA1.0-TPNC

Yuri Lavinas

July 14, 2016

Summary

O objeto é descobrir se existem variações ente os métodos e quais são as variáveis mais influentes.

Os métodos utilizados para comparação são o gaModel, a versão com listas, os métodos híbridos com e sem clusterização. Para cada um dos métodos temos algumas variações nas variáveis utilizadas. Variamos os anos (2005-2010), as regiões (Kanto, EastJapan), a profundidade (<100km) e finalmente o catálogo utilizado (SC da Yen-san).

Statistical Analysis

ANOVA test and HSD Tukey

Vou utilizar o ANOVA para nos dados obtidos para verificar qual composição de variáveis e métodos mais influenciam no resultado final.

Após as execuções vou aplicar o ANOVA em uma data.frame composto pelos dados das **médias dos melhores indivíduos da última geração** para cada cenário de execução.

Caso uma variável esteja fora do intervalo de confiança ($P < 0.05$), vou aplicar novamente o ANOVA retirando essa variável do teste.

Aplico um teste post hoc nos resultados do ANOVA para especificar quais são os grupos que diferem. O teste utilizado foi o Tukey teste.

É importante resaltar que para todos os casos, aplico uma função de limite, que altera os valores dos bins com mais que 12 ocorrências para 12.

Começo a análise carregando o data.frame com os dados, seguindo para a aplicação do teste ANOVA e finalizando com o uso do Tukey teste.

Filtering

Seleciono os modelos com terremotos com profundidade ≤ 100 km.

```
subTabela = finalData[finalData$depths==100,]
summary(subTabela)
```

```
## loglikeValues          model      depths      years
## Min.      :-5221.9   GAModel        : 240   100:3360   2005:560
## 1st Qu.   :-2369.3   ReducedGAModel    : 240   25 :    0   2006:560
## Median    :-2112.1   EMP-GAModel        : 240   60 :    0   2007:560
## Mean      :-2074.5   EMP-ReducedGAModel : 240                   2008:560
## 3rd Qu.   :-1620.8   EMP-GAModelWindow  : 240                   2009:560
## Max.      : -865.4   EMP-ReducedGAModelWindow: 240                   2010:560
##              (Other)      :1920
##           regions
## Kanto      :960
```

```
## Kansai :720
## Tohoku :720
## EastJapan:960
##
##
##
```

ANOVA - Specific analysis somente com Cluster.

Seleciono somente as áreas com dados do SC e os modelos apropriados.

```
subTabela3 = subTabela[subTabela$region=='Kanto'|subTabela$region=='EastJapan',]
subTabela3 = subTabela3[subTabela3$model=='ReducedGAModelSC'|subTabela3$model=='GAModelSC'|
                        subTabela3$model=='EMP-ReducedGAModelSC'|subTabela3$model=='EMP-GAModelSC'|
                        subTabela3$model=='ReducedGAModel'|subTabela3$model=='GAModel'|
                        subTabela3$model=='EMP-ReducedGAModel'|subTabela3$model=='EMP-GAModel',]
summary(subTabela3)
```

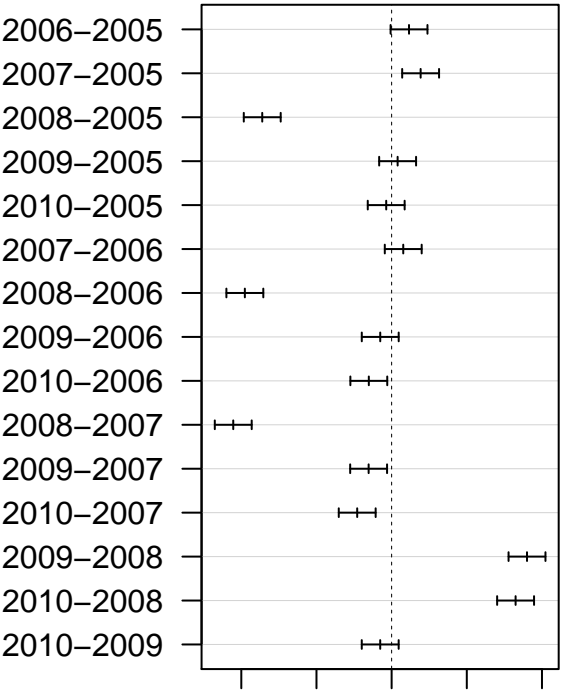
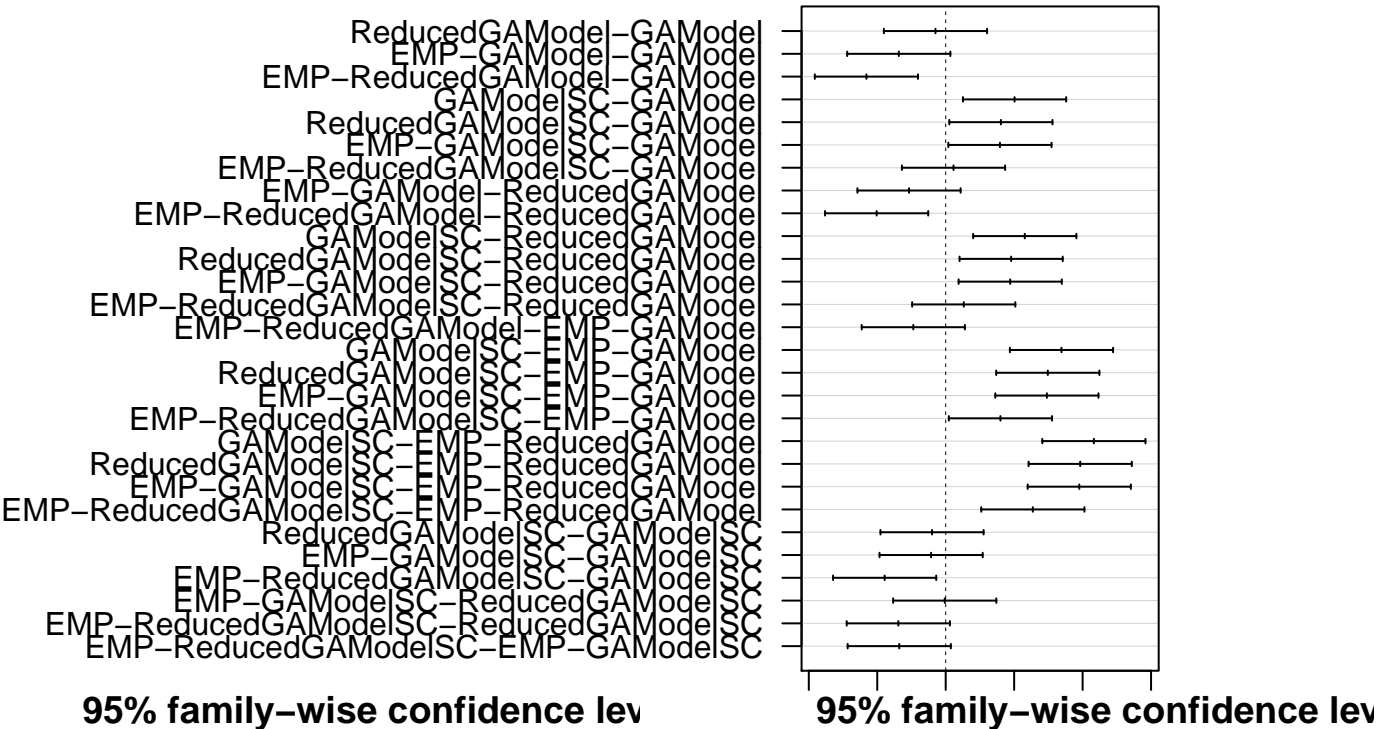
```
## loglikeValues          model      depths      years
## Min.      :-4617   GAModel          :120   100:960   2005:160
## 1st Qu.   :-2429   ReducedGAModel    :120    25 : 0   2006:160
## Median   :-2278   EMP-GAModel        :120    60 : 0   2007:160
## Mean      :-2425   EMP-ReducedGAModel:120                2008:160
## 3rd Qu.   :-2148   GAModelSC          :120                2009:160
## Max.      :-1811   ReducedGAModelSC   :120                2010:160
##              (Other)                :240
##      regions
## Kanto      :480
## Kansai     : 0
## Tohoku     : 0
## EastJapan:480
##
##
##
```

```
resultANOVA = aov(loglikeValues~model+years+regions , data = subTabela3)
summary(resultANOVA)
```

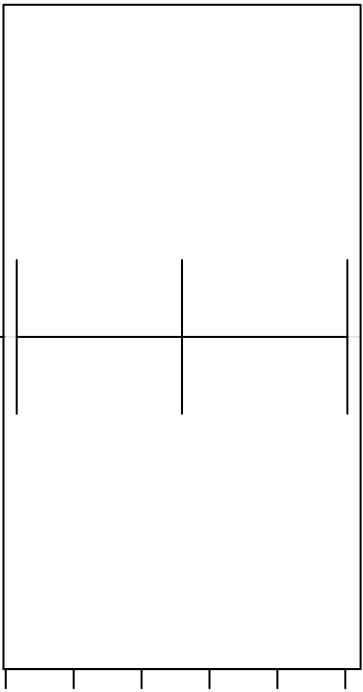
```
##              Df      Sum Sq  Mean Sq F value Pr(>F)
## model          7  19542417   2791774   18.9 <2e-16 ***
## years          5  119065905  23813181  161.2 <2e-16 ***
## regions        1   19918319  19918319   134.8 <2e-16 ***
## Residuals     946  139746434    147724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tuk = TukeyHSD(resultANOVA)
# par(mfrow=c(2,2))
op <- par(mar = c(1,21,4,2) + 0.1)
plot(tuk,las=1)
```

95% family-wise confidence lev



EastJapan-Kanto



```
# print(tuk)
```

Retiro o o EMP-ReducedGAModele e o emp-gamodel pelo desempenho ruim e refaço as análises sem eles.

```
subTabela3 = subTabela[subTabela$region=='Kanto'|subTabela$region=='EastJapan',]
subTabela3 = subTabela3[subTabela3$model=='ReducedGAModelSC'|subTabela3$model=='GAModelSC'|
                        subTabela3$model=='EMP-ReducedGAModelSC'|subTabela3$model=='EMP-GAModelSC'|
                        subTabela3$model=='GAModel'|subTabela3$model=='EMP-GAModel',]
summary(subTabela3)
```

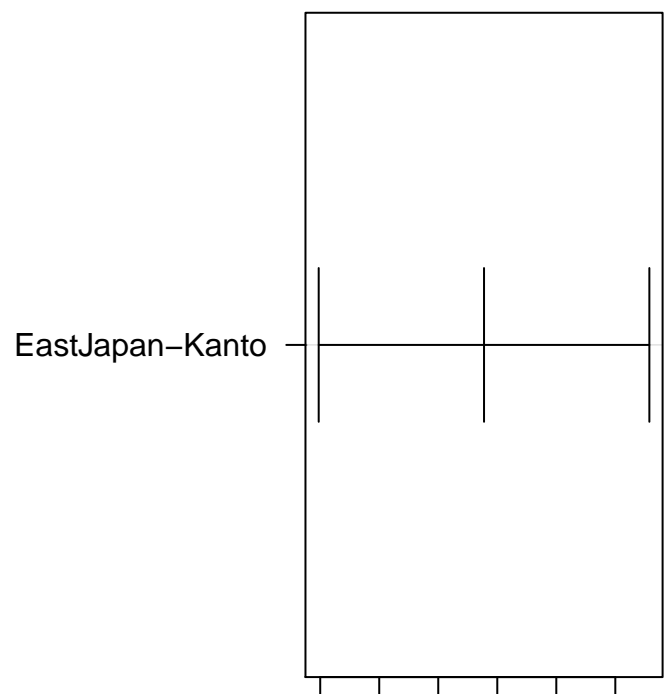
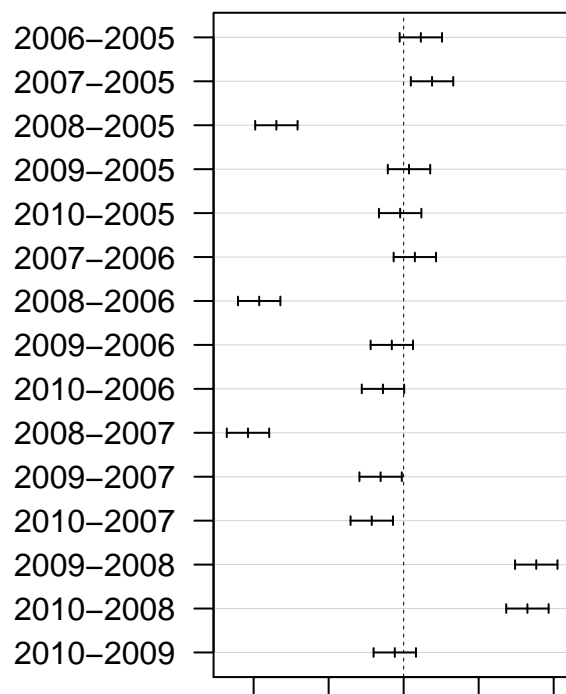
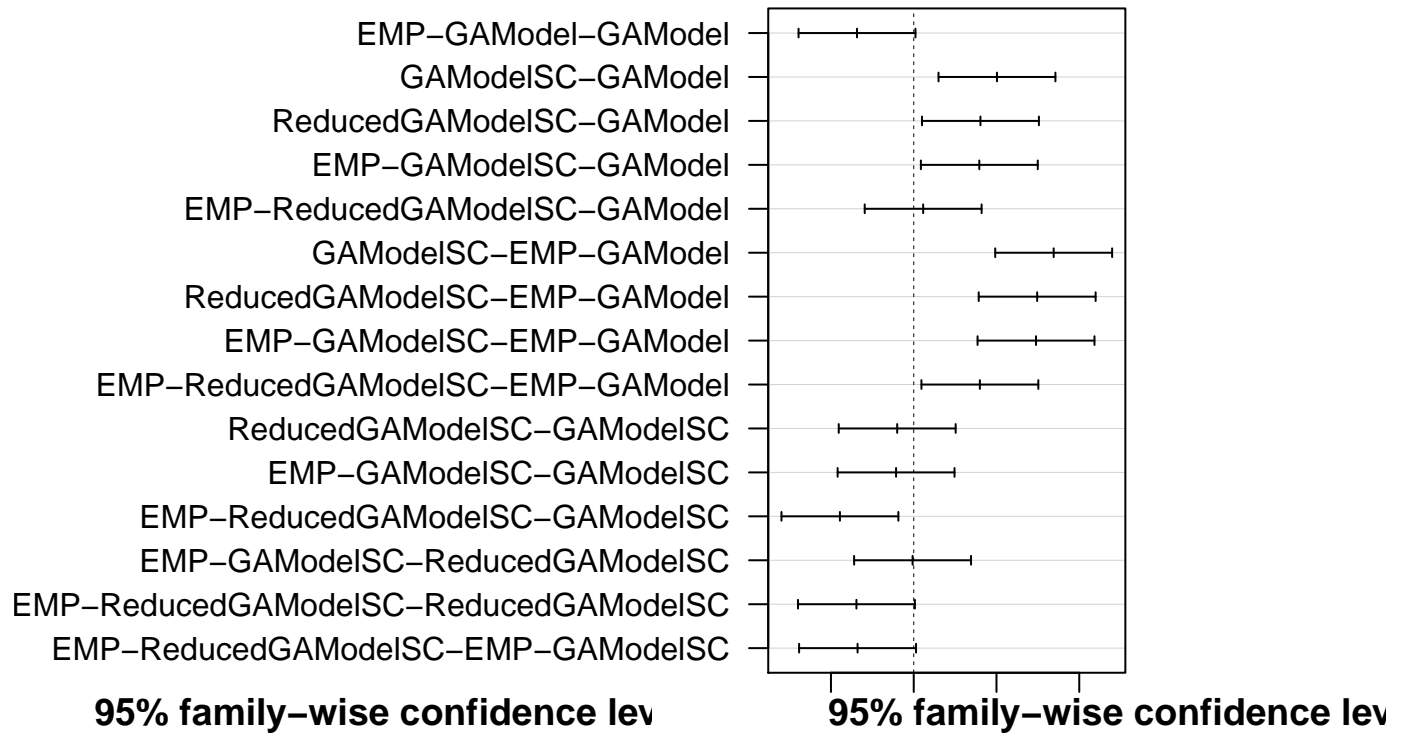
```
## loglikeValues          model      depths      years
## Min.      :-4424    GAModel          :120    100:720    2005:120
## 1st Qu.   :-2345    EMP-GAModel       :120     25 : 0    2006:120
## Median    :-2229    GAModelSC         :120     60 : 0    2007:120
## Mean      :-2375    ReducedGAModelSC   :120                2008:120
## 3rd Qu.   :-2113    EMP-GAModelSC      :120                2009:120
## Max.      :-1811    EMP-ReducedGAModelSC:120                2010:120
##              (Other)          : 0
##      regions
## Kanto      :360
## Kansai     : 0
## Tohoku     : 0
## EastJapan:360
##
##
##
```

```
resultANOVA = aov(loglikeValues~model+years+regions , data = subTabela3)
summary(resultANOVA)
```

```
##           Df      Sum Sq Mean Sq F value    Pr(>F)
## model      5  9999878  1999976    13.64 9.85e-13 ***
## years      5  86778328 17355666   118.41 < 2e-16 ***
## regions    1  14568011 14568011    99.39 < 2e-16 ***
## Residuals 708 103774355   146574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tuk = TukeyHSD(resultANOVA)
# par(mfrow=c(2,2))
op <- par(mar = c(1,21,4,2) + 0.1)
plot(tuk,las=1)
```

95% family-wise confidence lev



```
# print(tuk)
```

Como sempre é mais interessante utilizar o SC, refaço as análises só com eles para garantir que são estatisticamente iguais.

```
subTabela3 = subTabela[subTabela$region=='Kanto'|subTabela$region=='EastJapan',]
subTabela3 = subTabela3[subTabela3$model=='ReducedGAModelSC'|subTabela3$model=='GAModelSC'|
                        subTabela3$model=='EMP-ReducedGAModelSC'|subTabela3$model=='EMP-GAModelSC'],]
summary(subTabela3)
```

```
## loglikeValues          model      depths      years
## Min.      :-4204   GAModelSC      :120   100:480   2005:80
## 1st Qu.   :-2282   ReducedGAModelSC :120   25 : 0     2006:80
## Median    :-2173   EMP-GAModelSC    :120   60 : 0     2007:80
## Mean      :-2307   EMP-ReducedGAModelSC:120           2008:80
## 3rd Qu.   :-2099   GAModel          : 0           2009:80
## Max.      :-1811   ReducedGAModel    : 0           2010:80
##              (Other)      : 0
##      regions
## Kanto      :240
## Kansai     : 0
## Tohoku     : 0
## EastJapan:240
##
##
##
```

```
resultANOVA = aov(loglikeValues~model+years+regions , data = subTabela3)
summary(resultANOVA)
```

```
##           Df    Sum Sq Mean Sq F value    Pr(>F)
## model      3  2184319   728106   4.851  0.00247 **
## years      5 60607233 12121447  80.755 < 2e-16 ***
## regions    1  5457185   5457185  36.356 3.33e-09 ***
## Residuals 470 70548068   150102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Teste pareado para esses modelos: GAModelSC, GAModelWindow, ReducedGAModelSC, Emp-GAModelSC.

```
ttestPaired= function(region){
  subTabela6 = subTabela[subTabela$regions==region,]
  aggfinaldata<-aggregate(loglikeValues~years:model, data=subTabela6,FUN=mean)
  # Perform paired t-test
  cat('in', region, 'the t.test between the models GAModelSC and ReducedGAModelSC is: ')
  difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[7:12])
  print(t.test(difTimes))
  cat('in', region, 'the t.test between the models GAModelSC and Emp-GAModelSC is: ')
  difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[13:18])
  print(t.test(difTimes))
  cat('in', region, 'the t.test between the models ReducedGAModelSC and Emp-GAModelSC is: ')
}
```

```

difTimes<-with(aggfinaldata,loglikeValues[7:12]-loglikeValues[13:18])
print(t.test(difTimes))
}

# ttestPaired('Kansai')
# ttestPaired('Tohoku')
ttestPaired('EastJapan')

```

```

## in EastJapan the t.test between the models GAModelSC and ReducedGAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 0.060707, df = 5, p-value = 0.9539
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -77.42738 81.17290
## sample estimates:
## mean of x
## 1.872764
##
## in EastJapan the t.test between the models GAModelSC and Emp-GAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 25.208, df = 5, p-value = 1.834e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 174.6612 214.3277
## sample estimates:
## mean of x
## 194.4944
##
## in EastJapan the t.test between the models ReducedGAModelSC and Emp-GAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 5.1102, df = 5, p-value = 0.003738
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 95.7269 289.5165
## sample estimates:
## mean of x
## 192.6217

```

```

ttestPaired('Kanto')

```

```

## in Kanto the t.test between the models GAModelSC and ReducedGAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 6.4009, df = 5, p-value = 0.00138
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:

```

```

## 34.68124 81.23101
## sample estimates:
## mean of x
## 57.95612
##
## in Kanto the t.test between the models GAModelSC and Emp-GAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 13.918, df = 5, p-value = 3.441e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 64.90515 94.31048
## sample estimates:
## mean of x
## 79.60781
##
## in Kanto the t.test between the models ReducedGAModelSC and Emp-GAModelSC is:
## One Sample t-test
##
## data: difTimes
## t = 1.9367, df = 5, p-value = 0.1105
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -7.086516 50.389896
## sample estimates:
## mean of x
## 21.65169

```