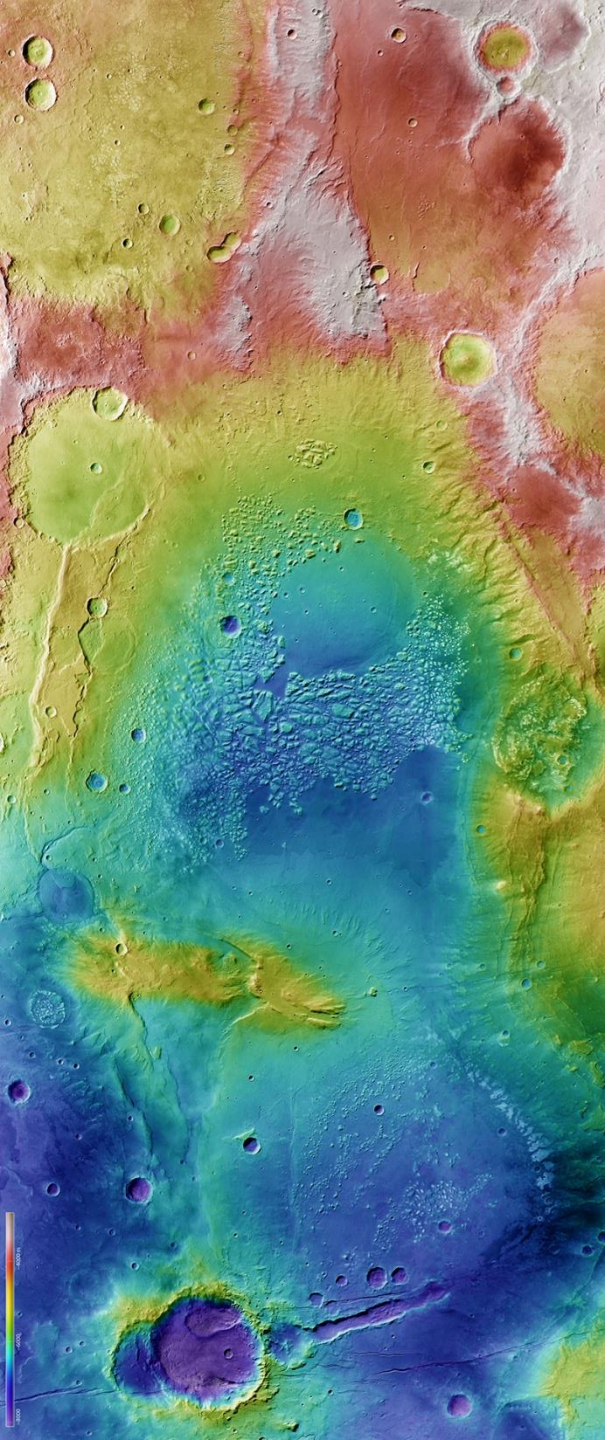
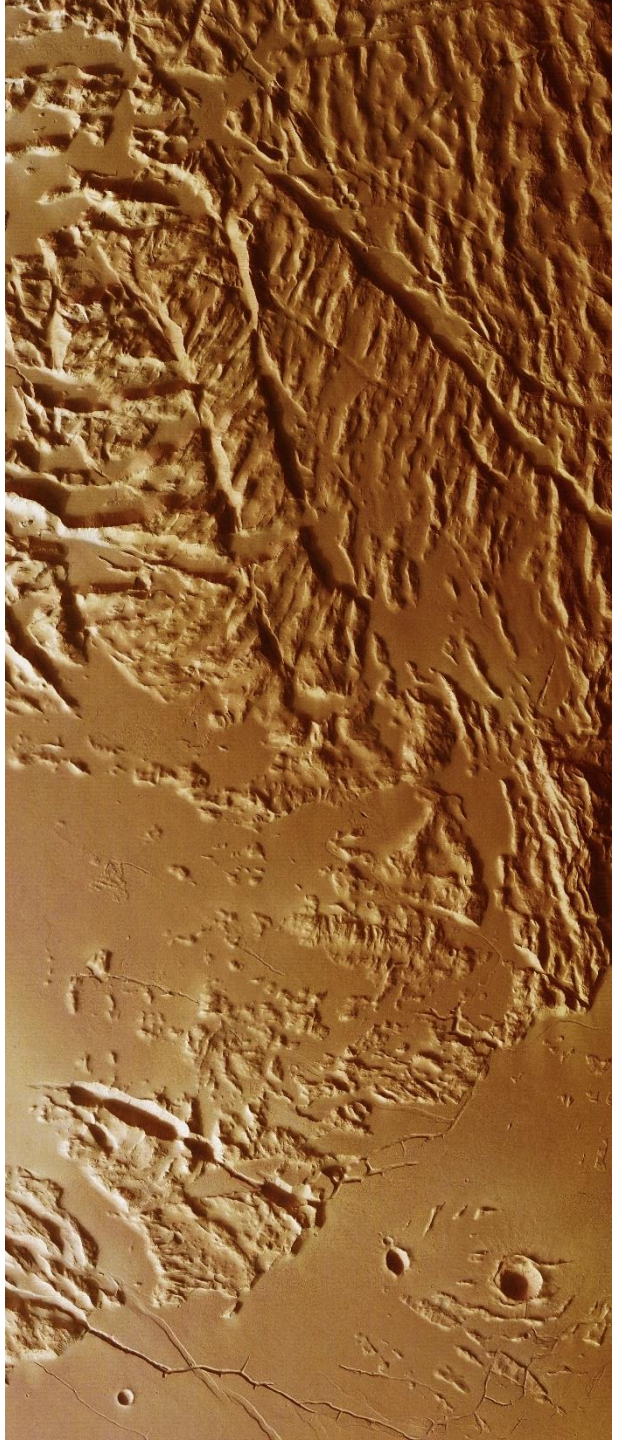
Mars Express Power Challenge

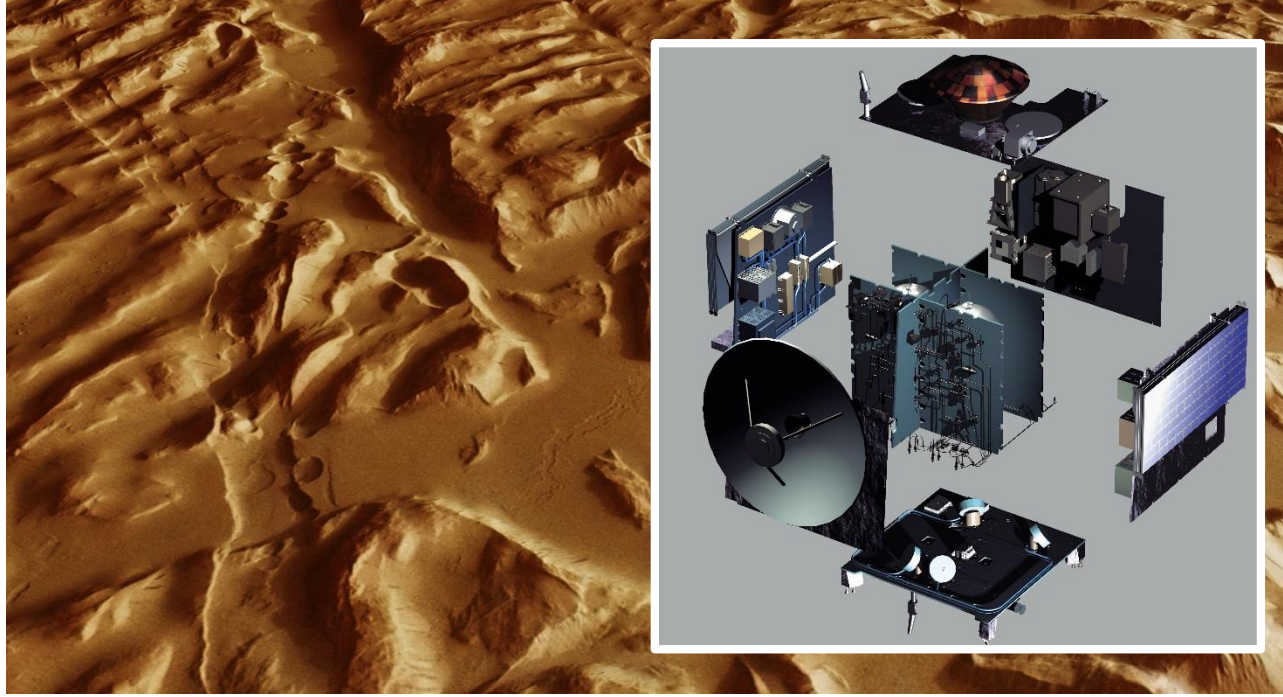
CC71Q - Introducción a la Minería de Datos

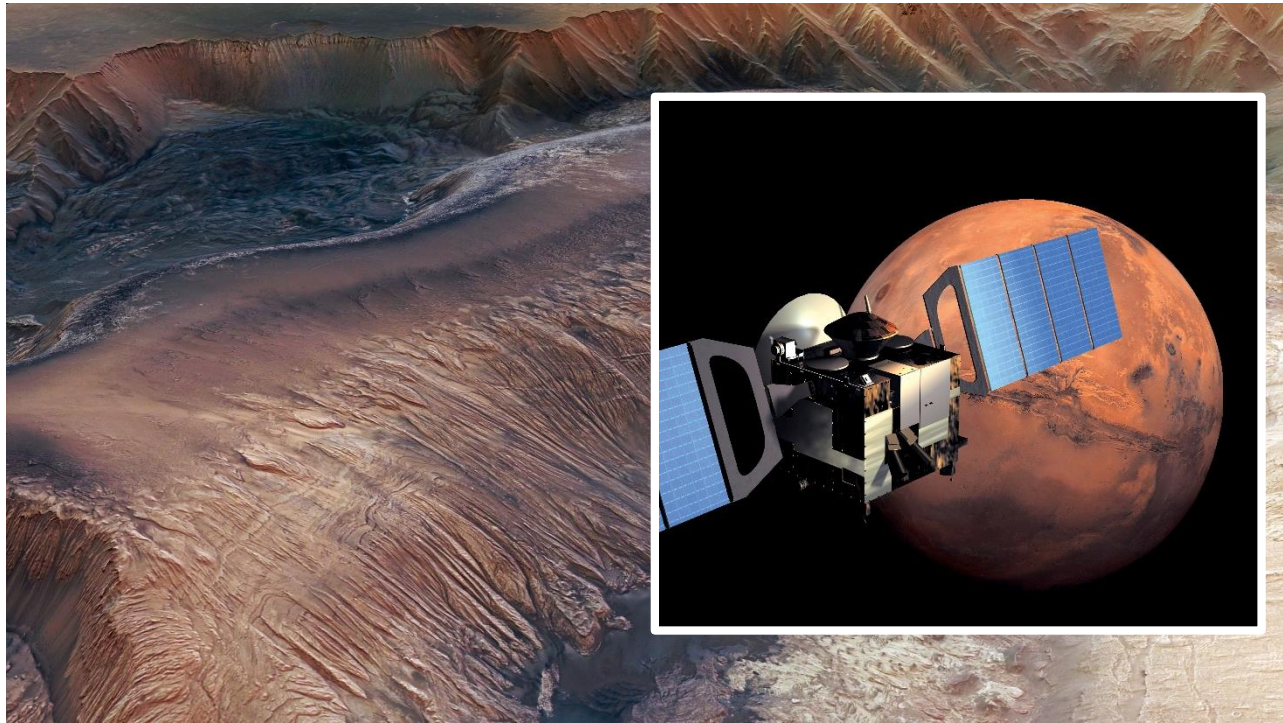
Gabriel De La Parra

30.Mayo.2016





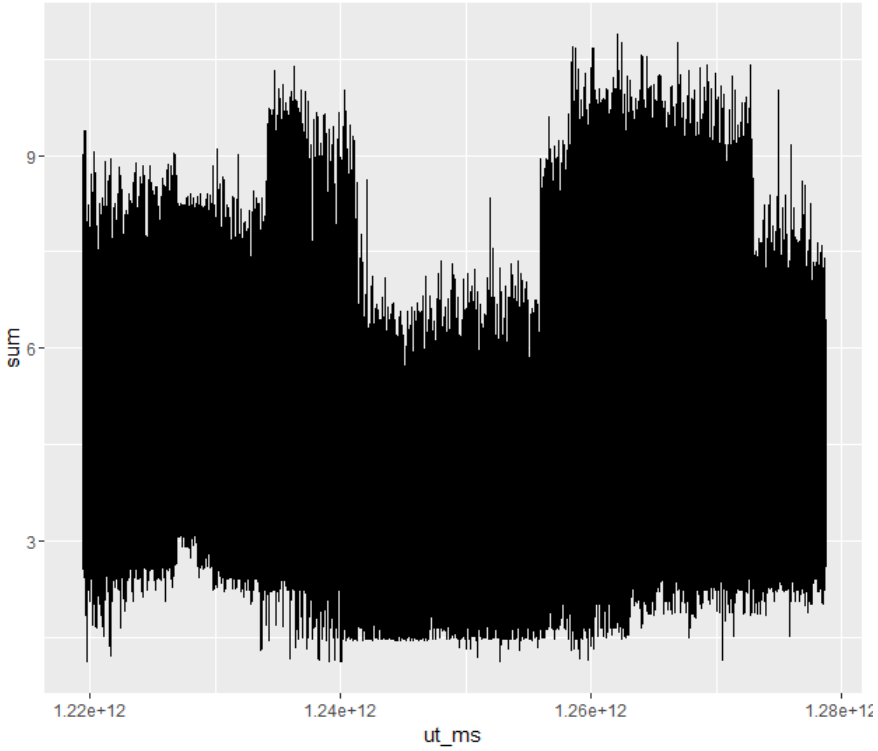




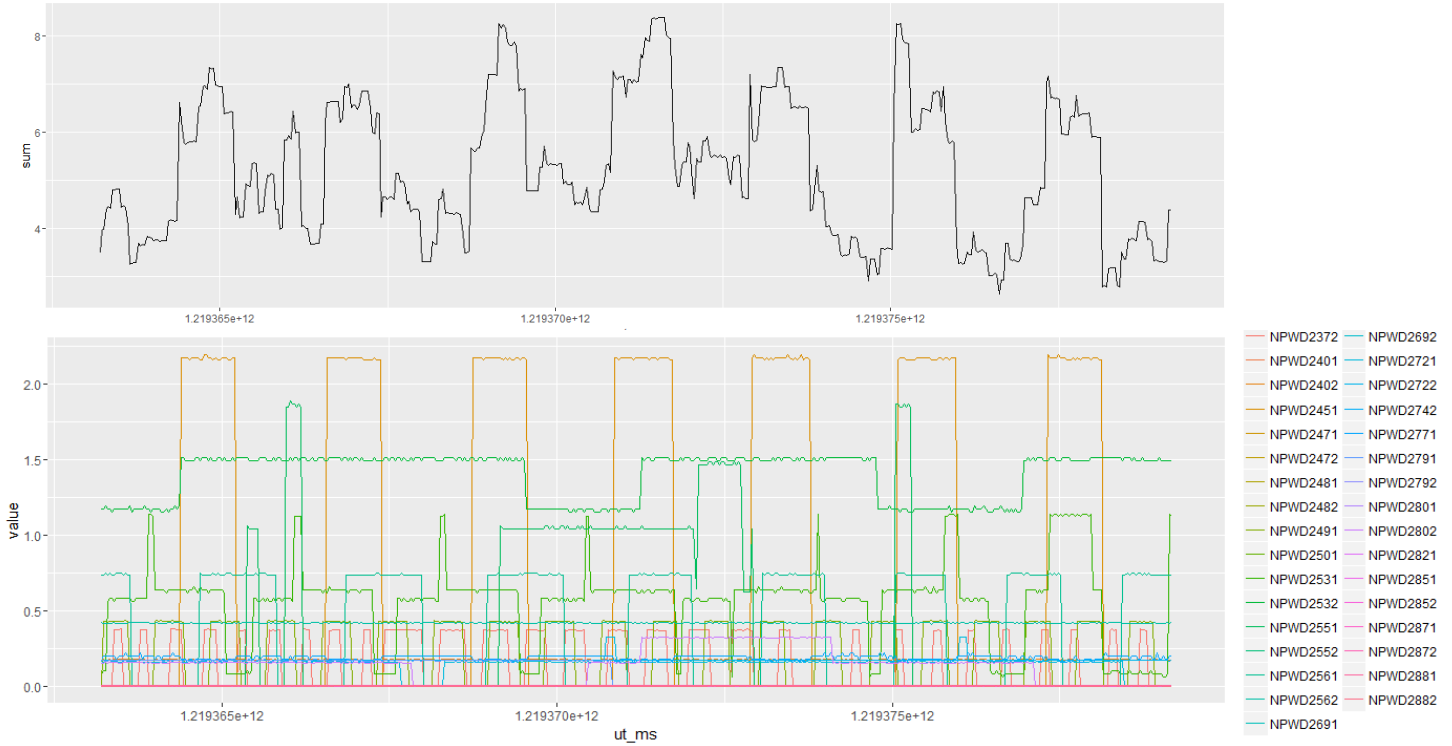
Mars Express Power Challenge

Problema: Predicción del consumo energético del satélite

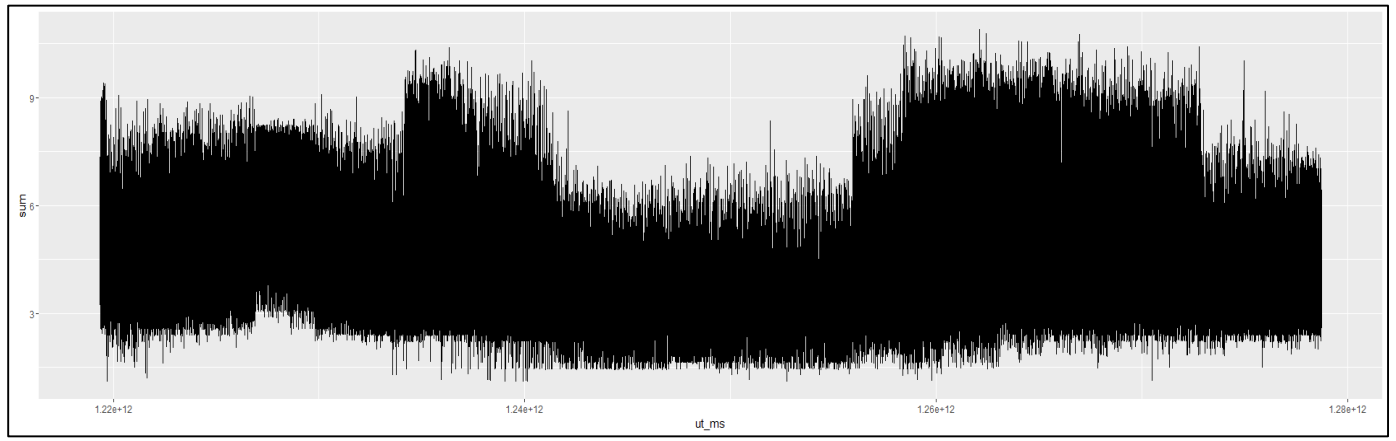
$E_{Ciencia} = E_{Solar} - E_{Nave} - E_{Climatización}$

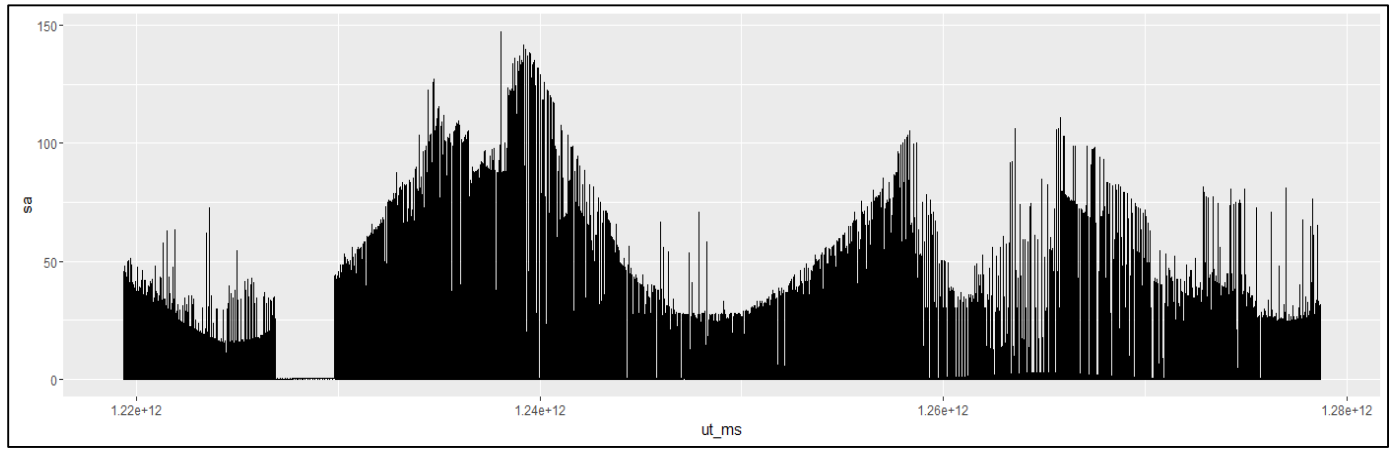


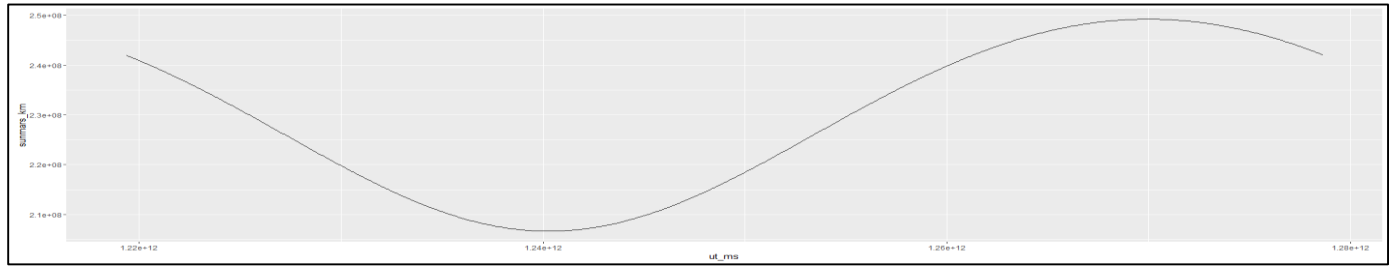
Consumo: 1 año, Suma circuitos



Consumo [1:500]







Pre-procesamiento

- Ajuste escala temporal
- Agrupamiento: Promedio por hora
- Match: Time Scale
- Join *
- Interpolación de valores faltantes
- Split: Train, Test

Procesamiento

- Regresión: Random Forest
- Evaluación: Root Mean Square

Pre-procesamiento
Ajuste escala temporal

ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451
1219363209542	0.001821	0.00147389	0.172173	0.0050155
1219363241537	0.0021852	0.00176867	0.17744	0.0060186
1219363273542	0.0021852	0.00147389	0.172173	0.0050155
1219363305537	0.001821	0.00147389	0.172173	0.0050155
1219363337542	0.001821	0.00176867	0.17744	0.0050155
1219363369537	0.0021852	0.00147389	0.172173	0.0050155
1219363401542	0.373872	0.00176867	0.17744	0.0050155
1219363433537	0.373872	0.00176867	0.17744	0.0050155
1219363465560	0.373872	0.00147389	0.17744	0.0050155

ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451
21/08/2008 20:00:41	0.001821	0.00147389	0.172173	0.0050155
21/08/2008 20:01:13	0.0021852	0.00176867	0.17744	0.0060186
21/08/2008 20:01:45	0.0021852	0.00147389	0.172173	0.0050155
21/08/2008 20:02:17	0.001821	0.00147389	0.172173	0.0050155
21/08/2008 20:02:49	0.001821	0.00176867	0.17744	0.0050155
21/08/2008 20:03:21	0.0021852	0.00147389	0.172173	0.0050155
21/08/2008 20:03:53	0.373872	0.00176867	0.17744	0.0050155
21/08/2008 20:04:25	0.373872	0.00176867	0.17744	0.0050155
21/08/2008 20:04:57	0.373872	0.00147389	0.17744	0.0050155

```
power1DT <- power1
power1DT$ut_ms <- as.POSIXct((((power1['ut_ms'])/1000)[,]), origin="1970-01-01")
```

Pre-procesamiento
Join & Match temporal

ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451	...	NPWD2882	sa	sx	sy	sz	sunmars_km	earthmars_km	sunmarsearthangle_deg	solarconstantmars
21/08/2008 20:00:00	NA	NA	NA	NA	...	NA	NA	NA	NA	NA	241938908	355756044	19.56508	522.264
21/08/2008 20:00:09	0.001821	0.00147389	0.172173	0.0050155	...	0.0032275	NA	NA	NA	NA	NA	NA	NA	NA
21/08/2008 20:00:13	NA	NA	NA	NA	...	NA	0.32	14.55	90.32	104.55	NA	NA	NA	NA
21/08/2008 20:00:35	NA	NA	NA	NA	...	NA	0.34	14.56	90.34	104.55	NA	NA	NA	NA
21/08/2008 20:00:41	0.0021852	0.00176867	0.17744	0.0060186	...	0.0032275	NA	NA	NA	NA	NA	NA	NA	NA
21/08/2008 20:01:13	0.0021852	0.00147389	0.172173	0.0050155	...	0.0032275	NA	NA	NA	NA	NA	NA	NA	NA
21/08/2008 20:01:35	NA	NA	NA	NA	...	NA	0.34	14.56	90.34	104.55	NA	NA	NA	NA
21/08/2008 20:01:45	0.001821	0.00147389	0.172173	0.0050155	...	0.0032275	NA	NA	NA	NA	NA	NA	NA	NA
21/08/2008 20:02:17	0.001821	0.00176867	0.17744	0.0050155	...	0.002582	NA	NA	NA	NA	NA	NA	NA	NA
21/08/2008 20:02:35	NA	NA	NA	NA	...	NA	0.34	14.56	90.34	104.55	NA	NA	NA	NA

```
power1DTHourMean<-merge(x=power1DTHourMean, y=saaf1DTHourMean, by="ut_ms", all.x=TRUE)

power1DTHourMeanMS <- power1DTHourMean$ut_ms
for (i in 1:nrow(saaf1DTHourMean)) {
  saaf1DTHourMean$ut_ms[i] <-
power1DTHourMeanMS[findInterval(saaf1DTHourMean$ut_ms[i],power1DTHourMeanMS)]
}
```

Pre-procesamiento
Agrupamiento por hora

ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451
21/08/2008 20:00:41	0.001821	0.00147389	0.172173	0.0050155
21/08/2008 20:01:13	0.0021852	0.00176867	0.17744	0.0060186
21/08/2008 20:01:45	0.0021852	0.00147389	0.172173	0.0050155
21/08/2008 20:02:17	0.001821	0.00147389	0.172173	0.0050155
21/08/2008 20:02:49	0.001821	0.00176867	0.17744	0.0050155
21/08/2008 20:03:21	0.0021852	0.00147389	0.172173	0.0050155
21/08/2008 20:03:53	0.373872	0.00176867	0.17744	0.0050155
21/08/2008 20:04:25	0.373872	0.00176867	0.17744	0.0050155
21/08/2008 20:04:57	0.373872	0.00147389	0.17744	0.0050155

ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451
21/08/2008 20:00:00	0.13358171	0.00155215	0.1754357	0.6572397
21/08/2008 21:00:00	0.21700089	0.001497578	0.1758881	0.8549184
21/08/2008 22:00:00	0.15592227	0.001502585	0.1753891	1.079016
21/08/2008 23:00:00	0.08144031	0.001531793	0.1753238	0.720677
22/08/2008 0:00:00	0.10377346	0.001505194	0.1742239	0.8285784
22/08/2008 1:00:00	0.118108	0.001555481	0.175653	1.0110647
22/08/2008 2:00:00	0.13978257	0.001546933	0.1758552	0.3453059
22/08/2008 3:00:00	0.12790922	0.001502842	0.1757	0.4254372
22/08/2008 4:00:00	0.23511138	0.001520846	0.1753891	0.5488608

```
power1DT$ut_ms <- cut(power1DT$ut_ms, breaks="hour")
power1DTHourMean <- power1DT %>% group_by(ut_ms) %>% summarise_each(funs(mean))
```

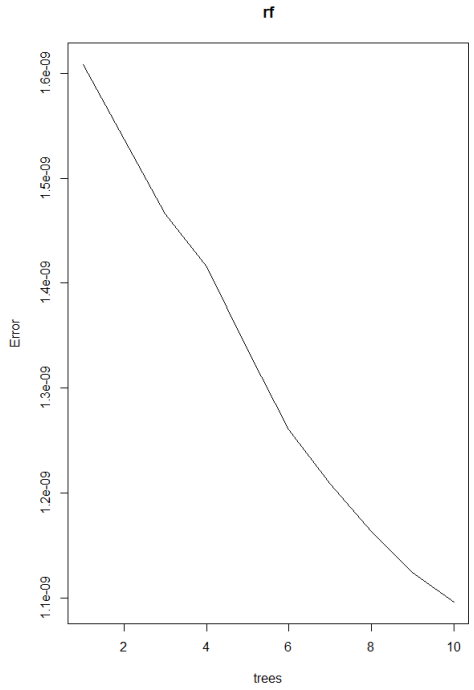
Pre-procesamiento
Aproximación de valores faltantes

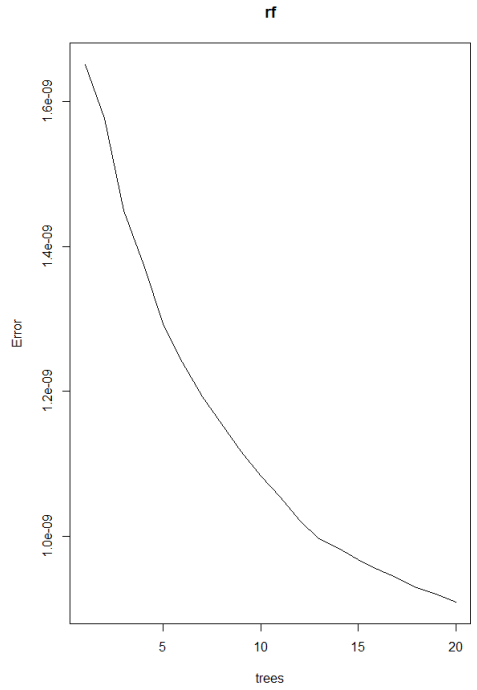
ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451	NPWD2882	sa	sx	sy	sz	sunmars_km	earthmars_km	sunmarsearthangle_deg	solarconstantmars
21/08/2008 20:00	0.13358171	0.00155215	0.1754357	0.6572397	0.003381735	0.34	14.55	90.34	104.55	241938908	355756044	19.56508	522.264
21/08/2008 21:00	0.21700089	0.001497578	0.1758881	0.8549184	0.003267844	6.62	18.01	90.5	90.396	*NA*	*NA*	*NA*	*NA*
...	...	...	...	...	...	...	...	...	...				
22/08/2008 0:00	0.10377346	0.001505194	0.1742239	0.8285784	0.00337031	0.34	14.52	90.34	104.52	*NA*	*NA*	*NA*	*NA*
22/08/2008 1:00	0.118108	0.001555481	0.175653	1.0110647	0.003354295	0.34	14.52	90.34	104.51	*NA*	*NA*	*NA*	*NA*
...	...	...	...	...	...	...	...	...	...	*NA*	*NA*	*NA*	*NA*
22/08/2008 20:00	0.09485976	0.001489682	0.1749946	1.0102144	0.003331241	0.34	14.32	90.34	104.32	241800160	356303701	19.39007	522.8635

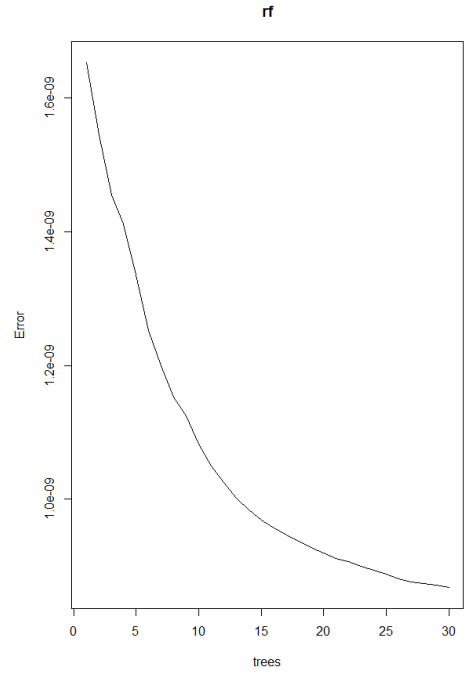
ut_ms	NPWD2372	NPWD2401	NPWD2402	NPWD2451	NPWD2882	sa	sx	sy	sz	sunmars_km	earthmars_km	sunmarsearthangle_deg	solarconstantmars
21/08/2008 20:00	0.13358171	0.00155215	0.1754357	0.6572397	0.6572397	0.34	14.55	90.34	104.55	241938908	355756044	19.56508	522.264
21/08/2008 21:00	0.21700089	0.001497578	0.1758881	0.8549184	0.8549184	6.62	18.01	90.5	90.396	241933149	355779026	19.55779	522.2889
21/08/2008 22:00	0.15592227	0.001502585	0.1753891	1.079016	1.079016	30.9	37.51	64.62	86.911	241927389	355801995	19.55051	522.3137
21/08/2008 23:00	0.08144031	0.001531793	0.1753238	0.720677	0.720677	1.53	15.22	89.97	105.2	241921626	355824949	19.54322	522.3386
22/08/2008 0:00	0.10377346	0.001505194	0.1742239	0.8285784	0.8285784	0.34	14.52	90.34	104.52	241915861	355847889	19.53593	522.3635
22/08/2008 1:00	0.118108	0.001555481	0.175653	1.0110647	1.0110647	0.34	14.52	90.34	104.51	241910095	355870814	19.52865	522.3884

```
power1DTHourMean$sunmars_km <- na.spline(power1DTHourMean[,grep("sunmars_km",
colnames(power1DTHourMean))],na.rm = FALSE)
```

Mars Express Power Challenge
RandomForest (ntree=10, 20, 30)



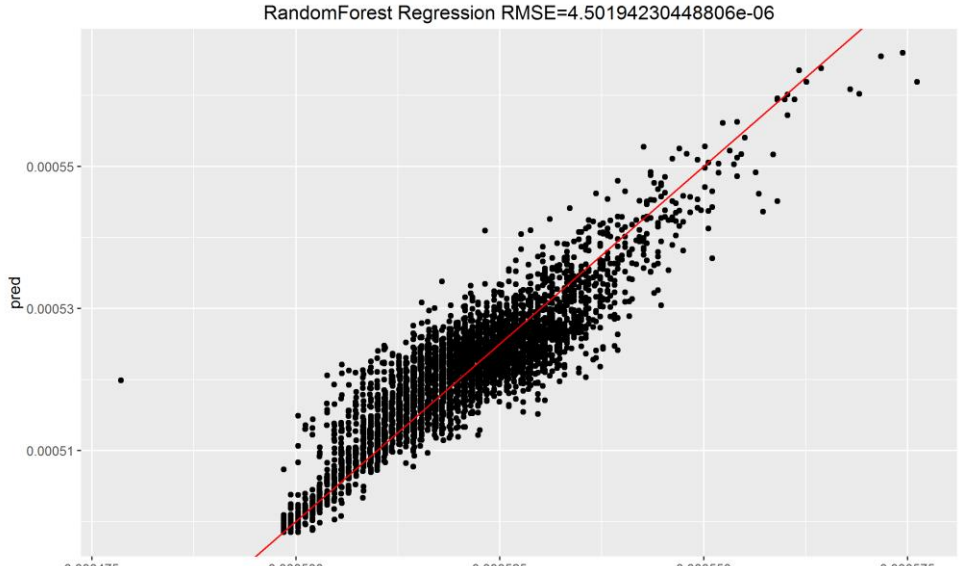


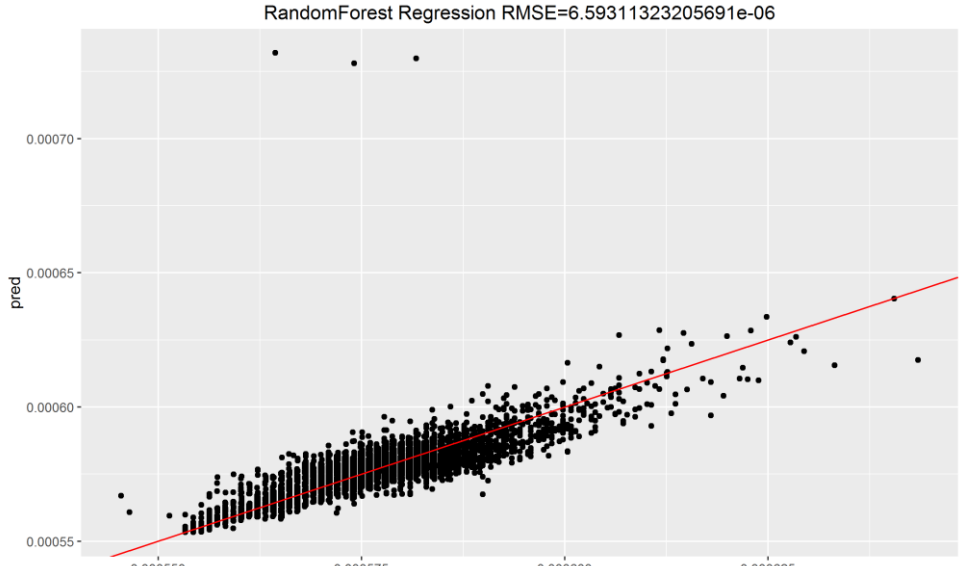


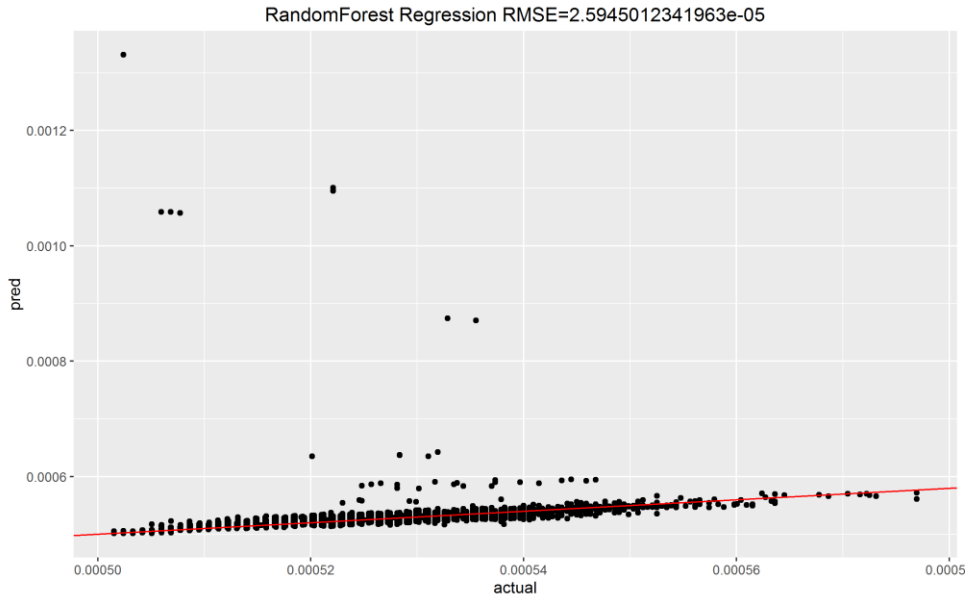

```
train <- power1DTHourMean[1:12000,-1]
 test <- power1DTHourMean[12001:16000,-1]
rf <- randomForest(as.formula(paste(colName," ~ sa + sx + sy + sz + sunmars_km + earthmars_km +
sunmarsearthangle_deg + solarconstantmars + eclipseduration_min + occultationduration_min")) ,data=train, ntree=30)
```

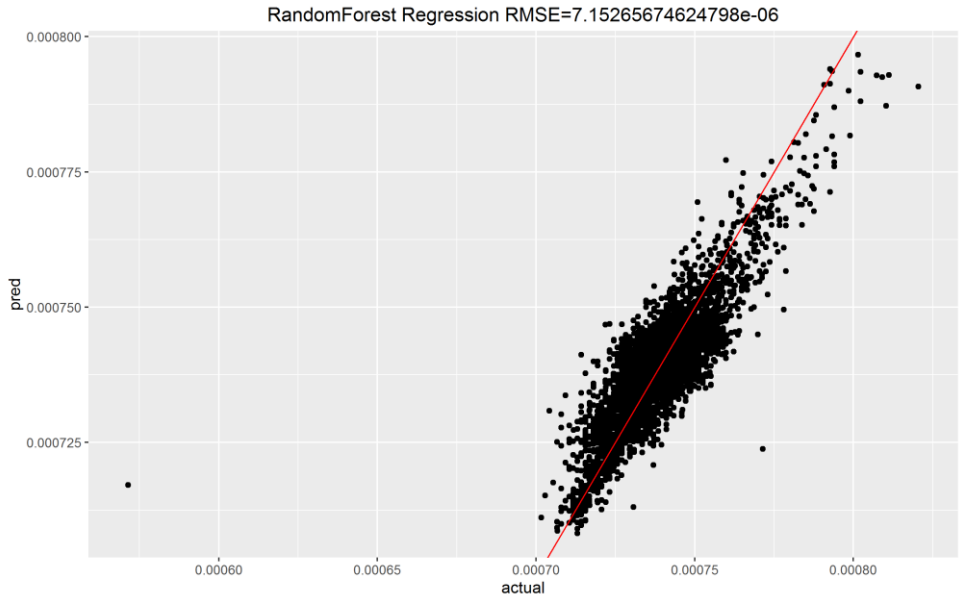
Mars Express Power Challenge

Resultados positivos (ntree=10)



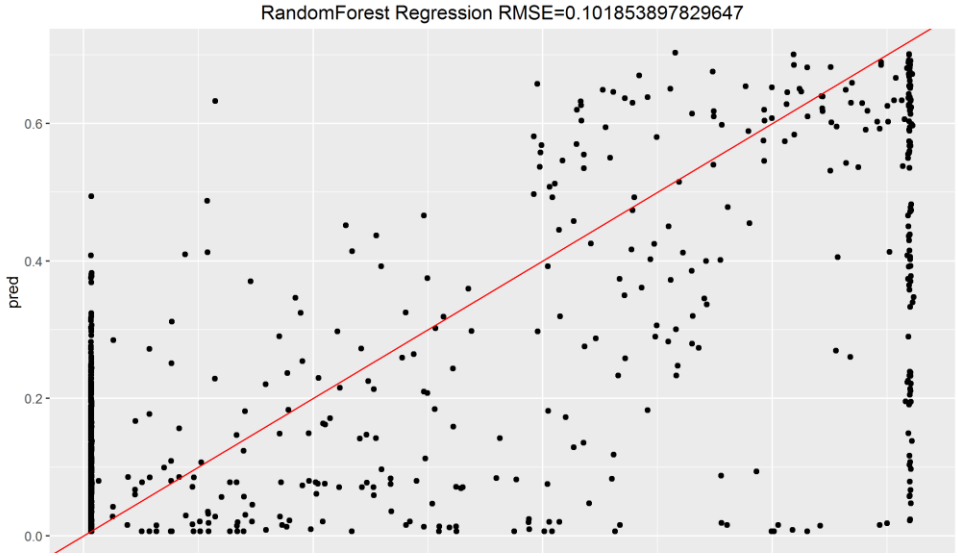


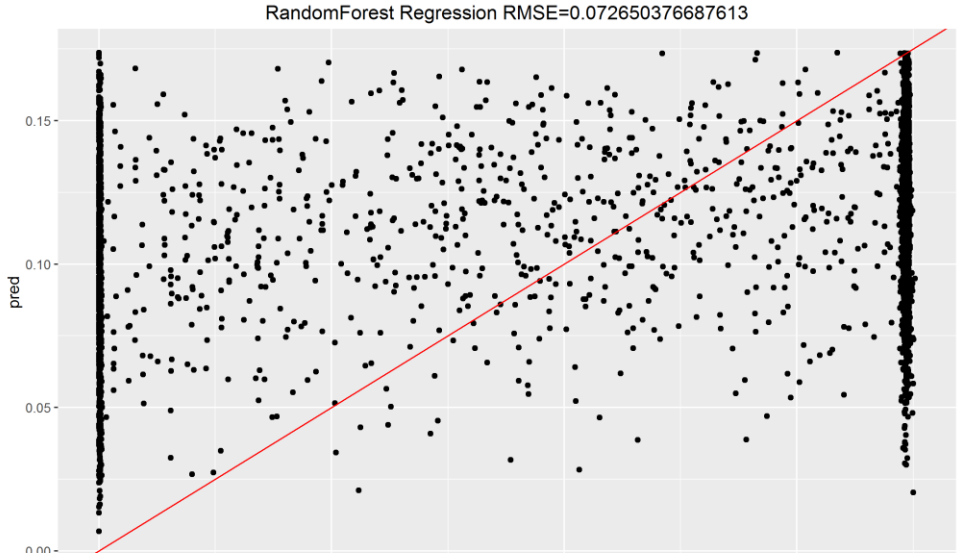


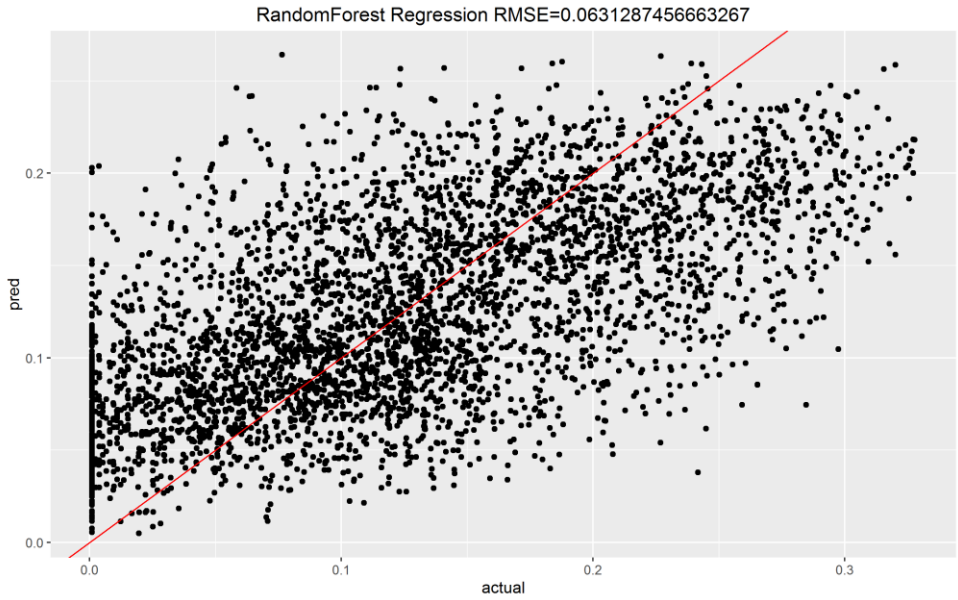


Mars Express Power Challenge

Resultados no tan positivos





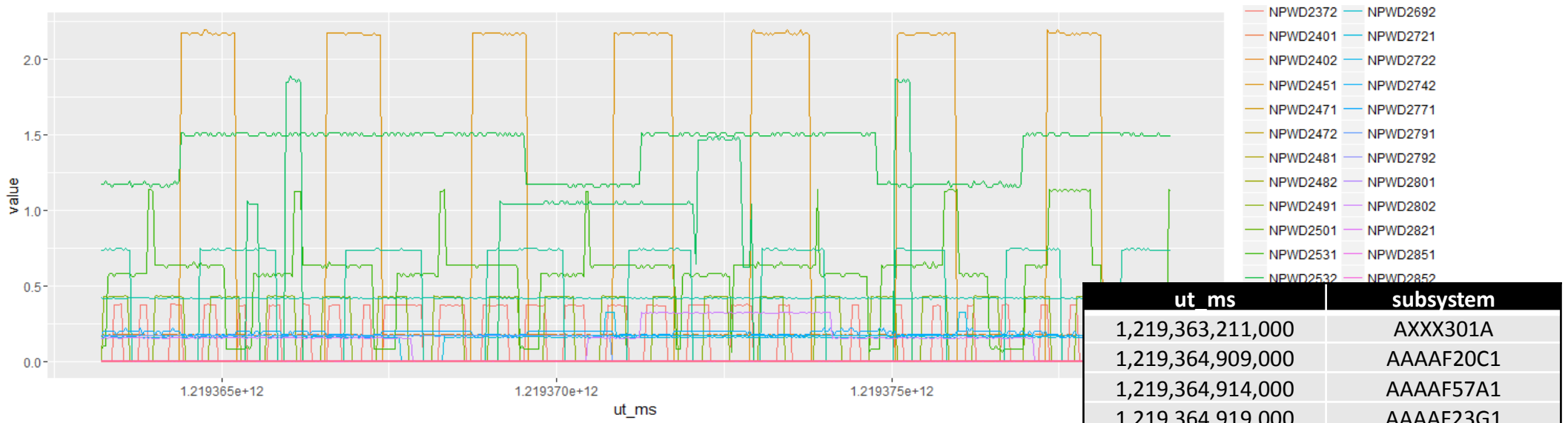


Mars Express Power Challenge
Resultados y Conclusiones

- Error total: 0.06579785 (6%)
- (Pre)Procesamiento = +Tiempo
- Recursos (hardware)
- Generación ~ Consumo
- Predicción macro: Ok
- Predicción micro: Falta procesamiento: DMOP, EVTF, FTL

Mars Express Power Challenge

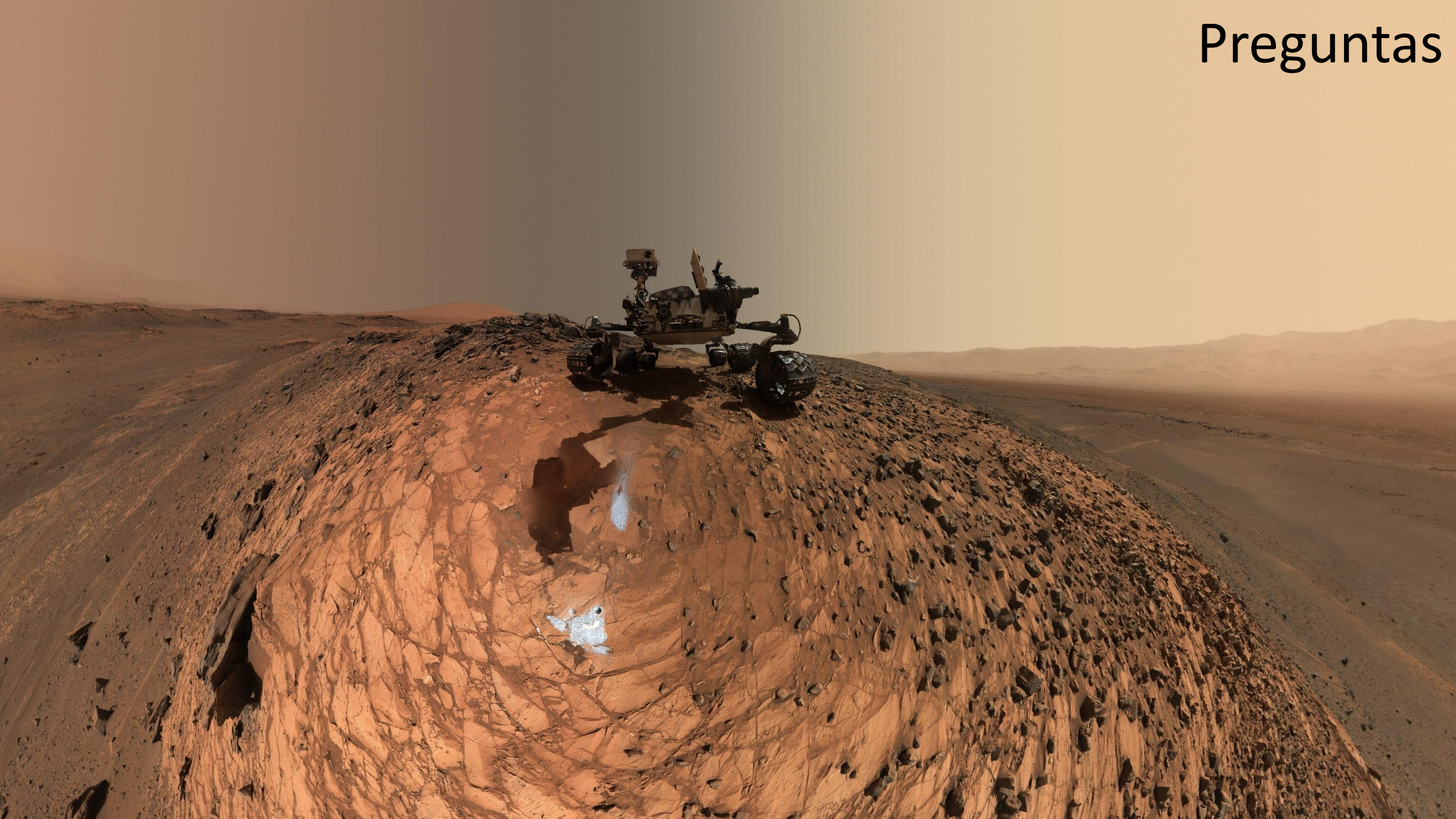
Próximos avances



ut_ms	subsystem
1,219,363,211,000	AXXX301A
1,219,364,909,000	AAAAF20C1
1,219,364,914,000	AAAAF57A1
1,219,364,919,000	AAAAF23G1
1,219,364,924,000	AAAAF60A1
1,219,365,615,000	AXXX305A
1,219,366,035,000	AXXX380A
1,219,366,635,000	ASEQ4200
1,219,367,381,000	ATTTF301E
1,219,369,061,000	ATTTF310A
1,219,369,181,000	APSF01A2
1,219,369,301,000	APSF02A1
1,219,369,306,000	APSF89A1

- Comandos (DMOP)
- Neural? Fuzzy?

Preguntas



```r
#Escala Tiempo:
power1DT <- power1
power1DT$ut_ms <- as.POSIXct((((power1['ut_ms'])/1000)[,]), origin="1970-01-01")

#Agrupamiento: Promedio por hora:
power1DT$ut_ms <- cut(power1DT$ut_ms, breaks="hour")
power1DTHourMean <- power1DT %>% group_by(ut_ms) %>% summarise_each(funs(mean))

#Match: TimeScale
power1DTHourMeanMS <- power1DTHourMean$ut_ms
for (i in 1:nrow(saaf1DTHourMean)) {
  saaf1DTHourMean$ut_ms[i] <- power1DTHourMeanMS[findInterval(saaf1DTHourMean$ut_ms[i],power1DTHourMeanMS)]
}

#Merge:
power1DTHourMean<-merge(x=power1DTHourMean, y=saaf1DTHourMean, by="ut_ms", all.x=TRUE)
power1DTHourMean<-merge(x=power1DTHourMean, y=ltdata1DTHourMean, by="ut_ms", all.x=TRUE)

#Interpolacion:
power1DTHourMean$sunmars_km <- na.spline(power1DTHourMean[,grep("sunmars_km",
colnames(power1DTHourMean))],na.rm = FALSE)
power1DTHourMean$earthmars_km <- na.spline(power1DTHourMean[,grep("earthmars_km",
colnames(power1DTHourMean))],na.rm = FALSE)
```

```r
#Regresión
rmseSum <- 0
for(i in 1:33){
  predictField <- i #Campo a predecir
  predictCols <- colnames(power1DT[,-1]) #Columnas en juego:
  #Set de entrenamiento y pruebas
  train <- power1DTHourMean[1:12000,-1]
  test <- power1DTHourMean[12001:16000,-1]
  colName <- predictCols[predictField]
  #Entrenamiento:
  rf <- randomForest(as.formula(paste(colName," ~ sa + sx + sy + sz + sunmars_km + earthmars_km + sunmarsearthangle_deg + solarconstantmars + eclipseduration_min + occultationduration_min")) ,data=train, ntree=10)
  #Prueba:
  predicted <- predict(rf, test)
  predCol <- test[,c(colName)]
  #Medición error:
  r2 <- RMSE(predCol, predicted)
    #Graficar Predicción vs. Referencia:
  p <- ggplot(aes(x=actual, y=pred), data=data.frame(actual=predCol, pred=predict(rf, test)))
  p <- p + geom_point() +  geom_abline(color="red") + ggtitle(paste("RandomForest Regression RMSE=", r2, sep=""))
  rmseSum <- rmseSum + r2 #Acumulación Error
  ggsave(paste("Predict",i,".png"), p) #Guardar Imagen
}
errorTotal<-rmseSum/33
```