



Using Natural Language Processing to Convert Mud-Log Chip Descriptions to Useful Data Tables

Marcelo Guarido, David Emery, and Kris Innanen

Banff, December 2nd, 2022

www.crewes.org





Today's Agenda

Part 1:

Chip Descriptions to Table

Part 2:

Synthetic DTC

Part 3:

Conclusions

Part 1:

CHIP DESCRIPTIONS TO
TABLE

```
mirror object to mirror  
mirror_mod.mirror_object  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True
```

```
selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob)  
mirror_ob.select = 0  
= bpy.context.selected_object  
data.objects[one.name].select  
print("please select exactly
```

-- OPERATOR CLASSES -----

```
types.Operator):  
X mirror to the selected  
object.mirror_mirror_x"  
mirror X"
```

```
context):  
active_object is not
```

Chip Descriptions in Mud-Logging

- Part of the mud-logging report
- Descriptive mineral analysis per depth interval
- Used by interpreters
- Supporting document
- Usually in PDF files
- Hard to use as data





ConocoPhillips
Cuttings Descriptions Report

Well Name : Poseidon-2			Print Date 8/07/2010	
Wellsite Geologist(s) : M Boyd M Ortiz S Phillips J Bardelosa M Ortiz M Warrington				
Interval (m)	%	Lithology / Show Descriptions	Ca (%)	Mg (%)
Main				
2429.0 - 2430.0	100	ARGILLACEOUS CALCILUTITE: (Sample from Bit Junk Slot), dark greenish grey, firm, sticky, plastic, 40-50% clay material, strongly calcareous, grading to Marl?.	52	0
2430.0 - 2440.0	70	CALCARENITE: yellowish grey, very pale brown to very pale orange, trace off white to very light grey, moderately hard, blocky to sub-blocky, sucrosic in part, occasional fine calcite grains, 2% pyrite nodules, rare very fine disseminated pyrite in part.	78	0
	30	CEMENT:		
2440.0 - 2450.0	95	CALCARENITE: as above, very pale orange to very pale brown, trace to 2% pyrite nodules & forams, 3% chert in part, very fine quartz grains in part.	79	1
	5	CALCILUTITE: white, off white, firm to generally moderately hard, blocky to subblocky, cryptocrystalline.		
2450.0 - 2460.0	100	CALCARENITE: very pale brown to very pale orange, rare off white to white, moderately hard, sub-blocky to trace blocky, cryptocrystalline when white, occasionally sucrosic, trace pyrite nodules, trace forams and fossil fragments.	84	1
2460.0 - 2470.0	100	CALCARENITE: as above, very fine quartz grain inclusions in part.		
2470.0 - 2480.0	95	CALCARENITE: as above, very pale brown to very pale orange, moderately hard, sub-blocky to blocky, occasionally sucrosic, trace pyrite nodules.		
	5	CALCILUTITE: medium light grey, medium grey, firm to moderately hard, sub-blocky to blocky, slight to moderately argillaceous,		
2480.0 - 2490.0	95	CALCARENITE:	84	1
	5	CALCILUTITE: as above.		
2490.0 - 2500.0	97	CALCARENITE: as above, 3% chert.	87	1
	3	CHERT: very pale to pale orange, occasionally translucent, hard, angular, conchoidal fracture.		
2500.0 - 2510.0	95	CALCARENITE: as above, very pale brown to very pale orange, rare off white to white, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 5% chert.	87	1
	5	CHERT: as above.		
2510.0 - 2520.0	97	CALCARENITE: as above.	85	1
	5	CHERT: as above.		
2520.0 - 2530.0	92	CALCARENITE: as above, very pale brown to very pale orange, rare off white and pale orange, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 3-5% chert fragments.		
	5	CALCILUTITE: medium light grey, medium grey, medium bluish grey, firm, sub-blocky to occasionally blocky, slightly argillaceous,		
	3	CHERT: as above.		
2530.0 - 2540.0	90	CALCARENITE: as above.		
	5	CALCILUTITE: as above.		
	5	CHERT: as above.		
2540.0 - 2550.0	95	CALCARENITE: as above.	82	1
	5	CHERT: as above.		
2550.0 - 2560.0	98	CALCARENITE: very pale brown, yellowish grey, rare white, moderately hard, predominantly fine fragments, occasionally medium, sub-blocky to blocky, nil to trace chert, Claystone and orange translucent fragments.		
	2	CHERT: as above.		

© Copyright 2002 - Petroleum Data Systems International Pty Ltd
 All Rights Reserved

Page : 1 of 32

PDF to Table



ConocoPhillips
Cuttings Descriptions Report

Well Name :		Poseidon-2		Print Date		8/07/2010	
Wellsite Geologist(s) :		M Boyd M Ortiz S Phillips		J Bardelosa M Ortiz		M Warrington	
Interval (m)	%	Lithology / Show Descriptions				Ca (%)	Mg (%)
Main							
2429.0 - 2430.0	100	ARGILLACEOUS CALCILUTITE: (Sample from Bit Junk Slot), dark greenish grey, firm, sticky, plastic, 40-50% clay material, strongly calcareous, grading to Marl?.				52	0
2430.0 - 2440.0	70	CALCARENITE: yellowish grey, very pale brown to very pale orange, trace off white to very light grey, moderately hard, blocky to sub-blocky, sucrosic in part, occasional fine calcite grains, 2% pyrite nodules, rare very fine disseminated pyrite in part.				78	0
	30	CEMENT:					
2440.0 - 2450.0	95	CALCARENITE: as above, very pale orange to very pale brown, trace to 2% pyrite nodules & forams, 3% chert in part, very fine quartz grains in part.				79	1
	5	CALCILUTITE: white, off white, firm to generally moderately hard, blocky to subblocky, cryptocrystalline.					
2450.0 - 2460.0	100	CALCARENITE: very pale brown to very pale orange, rare off white to white, moderately hard, sub-blocky to trace blocky, cryptocrystalline when white, occasionally sucrosic, trace pyrite nodules, trace forams and fossil fragments.				84	1
2460.0 - 2470.0	100	CALCARENITE: as above, very fine quartz grain inclusions in part.					
2470.0 - 2480.0	95	CALCARENITE: as above, very pale brown to very pale orange, moderately hard, sub-blocky to blocky, occasionally sucrosic, trace pyrite nodules.					
	5	CALCILUTITE: medium light grey, medium grey, firm to moderately hard, sub-blocky to blocky, slight to moderately argillaceous,					
2480.0 - 2490.0	95	CALCARENITE:				84	1
	5	CALCILUTITE: as above.					
2490.0 - 2500.0	97	CALCARENITE: as above, 3% chert.				87	1
	3	CHERT: very pale to pale orange, occasionally translucent, hard, angular, conchoidal fracture.					
2500.0 - 2510.0	95	CALCARENITE: as above, very pale brown to very pale orange, rare off white to white, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 5% chert.				87	1
	5	CHERT: as above.					
2510.0 - 2520.0	97	CALCARENITE: as above.				85	1
	5	CHERT: as above.					
2520.0 - 2530.0	92	CALCARENITE: as above, very pale brown to very pale orange, rare off white and pale orange, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 3-5% chert fragments.					
	5	CALCILUTITE: medium light grey, medium grey, medium bluish grey, firm, sub-blocky to occasionally blocky, slightly argillaceous,					
2530.0 - 2540.0		CHERT: as above.					
	3						
2530.0 - 2540.0	90	CALCARENITE: as above.					
	5	CALCILUTITE: as above.					
2540.0 - 2550.0		CHERT: as above.					
	5						
2540.0 - 2550.0	95	CALCARENITE: as above.				82	1
	5	CHERT: as above.					
2550.0 - 2560.0	98	CALCARENITE: very pale brown, yellowish grey, rare white, moderately hard, predominantly fine fragments, occasionally medium, sub-blocky to blocky, nil to trace chert, Claystone and orange translucent fragments.					
	2	CHERT: as above.					

AutoSaveOffchip_description_Resampled_poseidon-2 • Saved

Search (Alt+Q)

Marcelo Guarido

FileHomeInsertDrawPage LayoutFormulasDataReviewViewHelp

UndoPasteCopyFormat PainterClipboardFontAlignmentNumberStylesCellsEditingAnalysis

Calibri11A[^]A_v

B*I*U

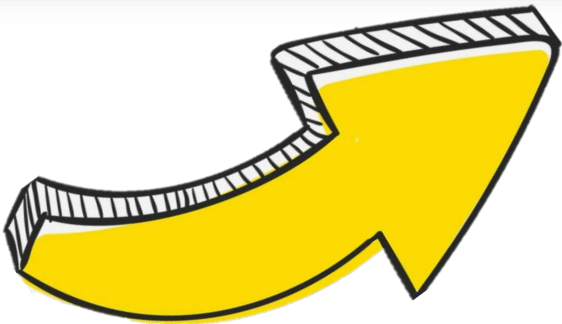
Wrap Text**Merge & Center**

General\$%

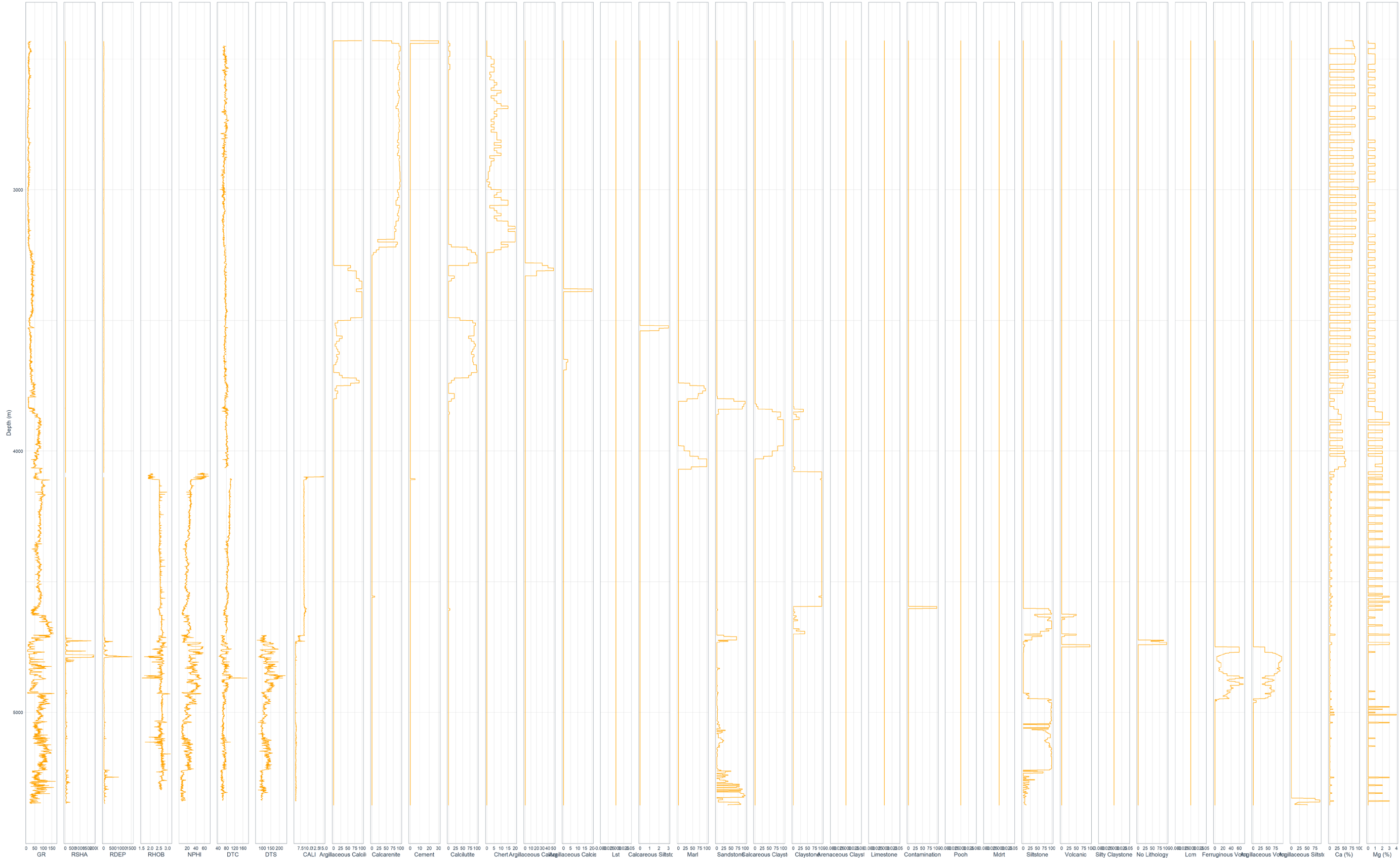
Conditional FormattingFormat as TableCell StylesInsertDeleteFormat

AutoSumFillClearSort & FilterFind & SelectAnalyze Data

E60



PDF to Table





☐ Copyright 2002 - Petroleum Data Systems International Pty Ltd
 All Rights Reserved

PDF to Table

SANDSTONE: light olive grey to pale yellowish brown, olive grey, translucent, predominantly moderately hard aggregates, predominantly fine to commonly medium, moderately well sorted, sub-rounded to occasionally sub-angular, moderately strong siliceous cement, minor silty matrix in part grading to trace very fine arenaceous **SILTSTONE**, predominantly fair to good visible porosity, poor porosity in very fine to silty aggregates, no fluorescence. **SILTSTONE:** olive black, commonly grading to olive grey, firm, blocky, weakly calcareous, moderately argillaceous, commonly arenaceous.

REGEX → Regular Expression

[A-Z]+:?:?([A-Z]+[A-Z \d]+:?)

PDF to Table

Tabulizer

- Select the table on the first page
- Read other pages
- Extract long table

Processing

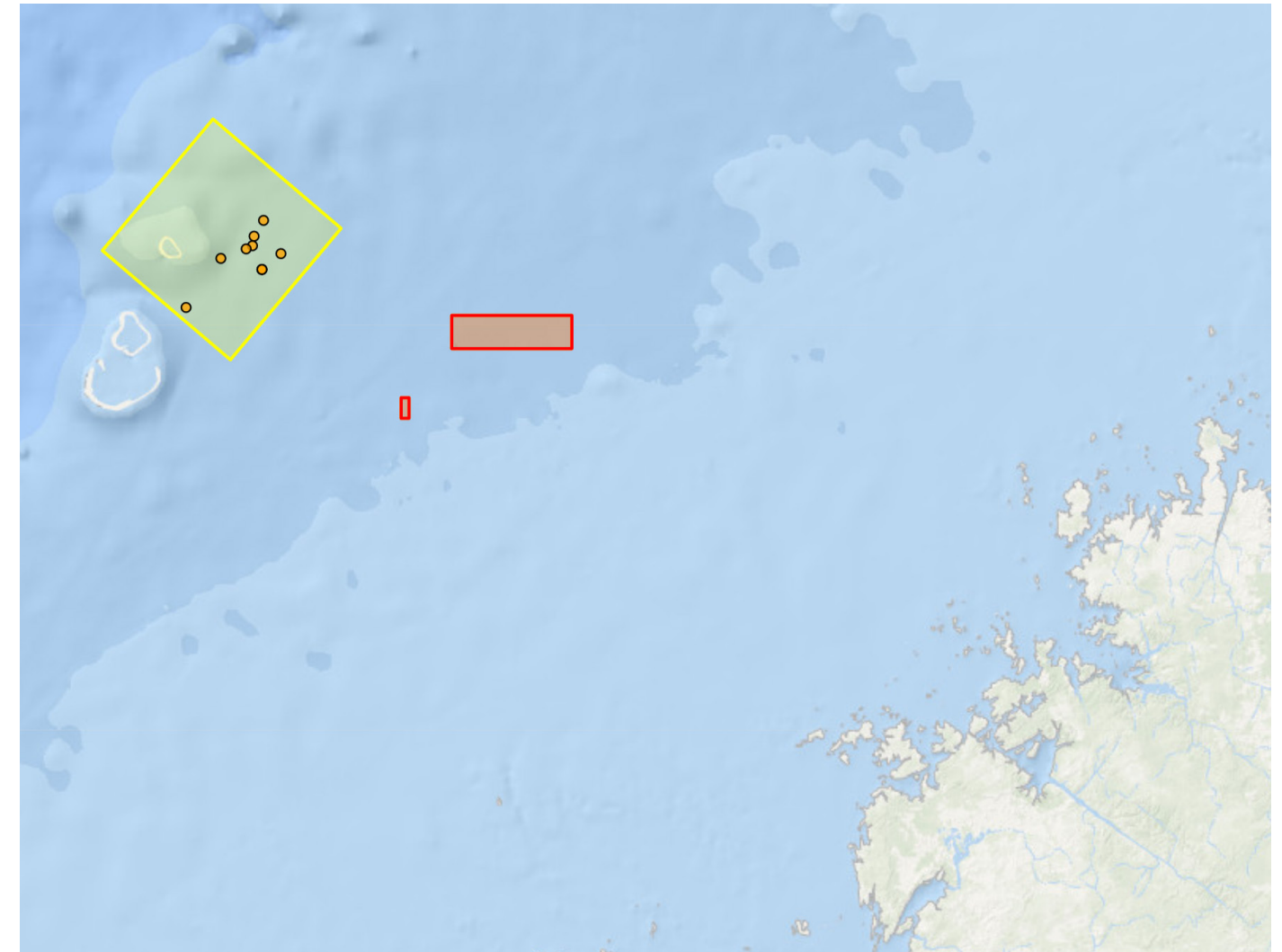
- Regularize table and merge rows
- Locate minerals with NLP and separate into columns
- Feed the new columns with percentages

Export

- Resample to the desired sample rate
- Export to CSV

Synthetic DTC

- Poseidon 3D (Australia)
- Mud-logging from well *Poseidon-2*
- Chip descriptions + GR + RDEP (LWD)
- De-trend logs
- Standardize the data (Z-Score)
- Predict DTC
 - Linear regression
 - Step-wise regression
 - XGBoost



Synthetic DTC: LWD

40/60 data split

```
Call:
lm(formula = DTC_Detrended ~ GR + RDEP + `Depth (m)`, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-27.379  -3.160  -0.161   3.037  33.115

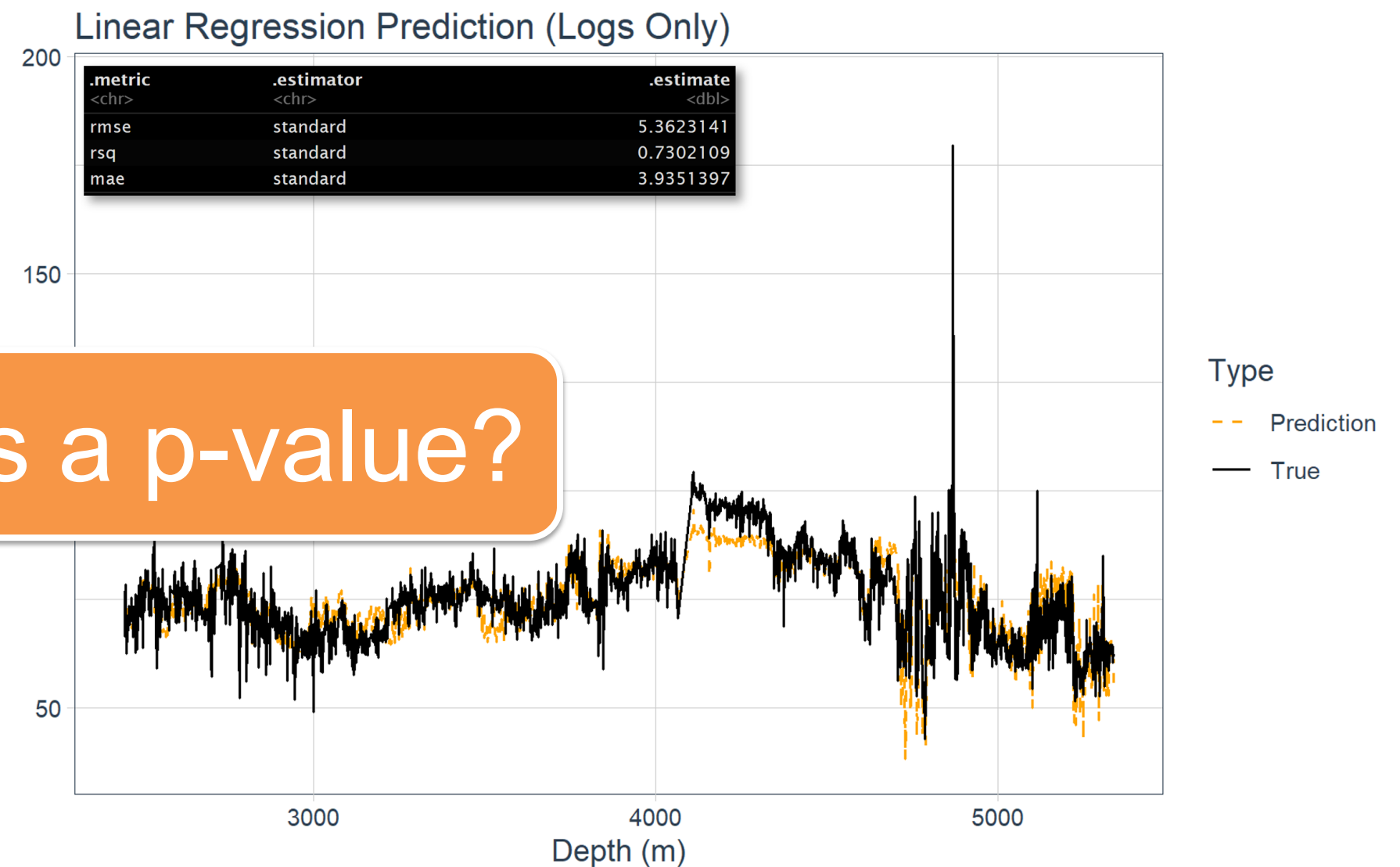
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.11421    0.11001  -1.038   0.299
GR           1.91385    0.12934  14.797 <2e-16 **
RDEP        -7.62717    0.12629 -60.395 <2e-16 **
`Depth (m)` -0.07992    0.10949  -0.730   0.466
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.244 on 2269 degrees of freedom
Multiple R-squared:  0.7318,    Adjusted R-squared:  0.7315
F-statistic: 2064 on 3 and 2269 DF,  p-value: < 2.2e-16
```

What is a p-value?

Not statistically significant

Significance level: 0.05



$R^2 = 0.73$

Hypothesis Testing

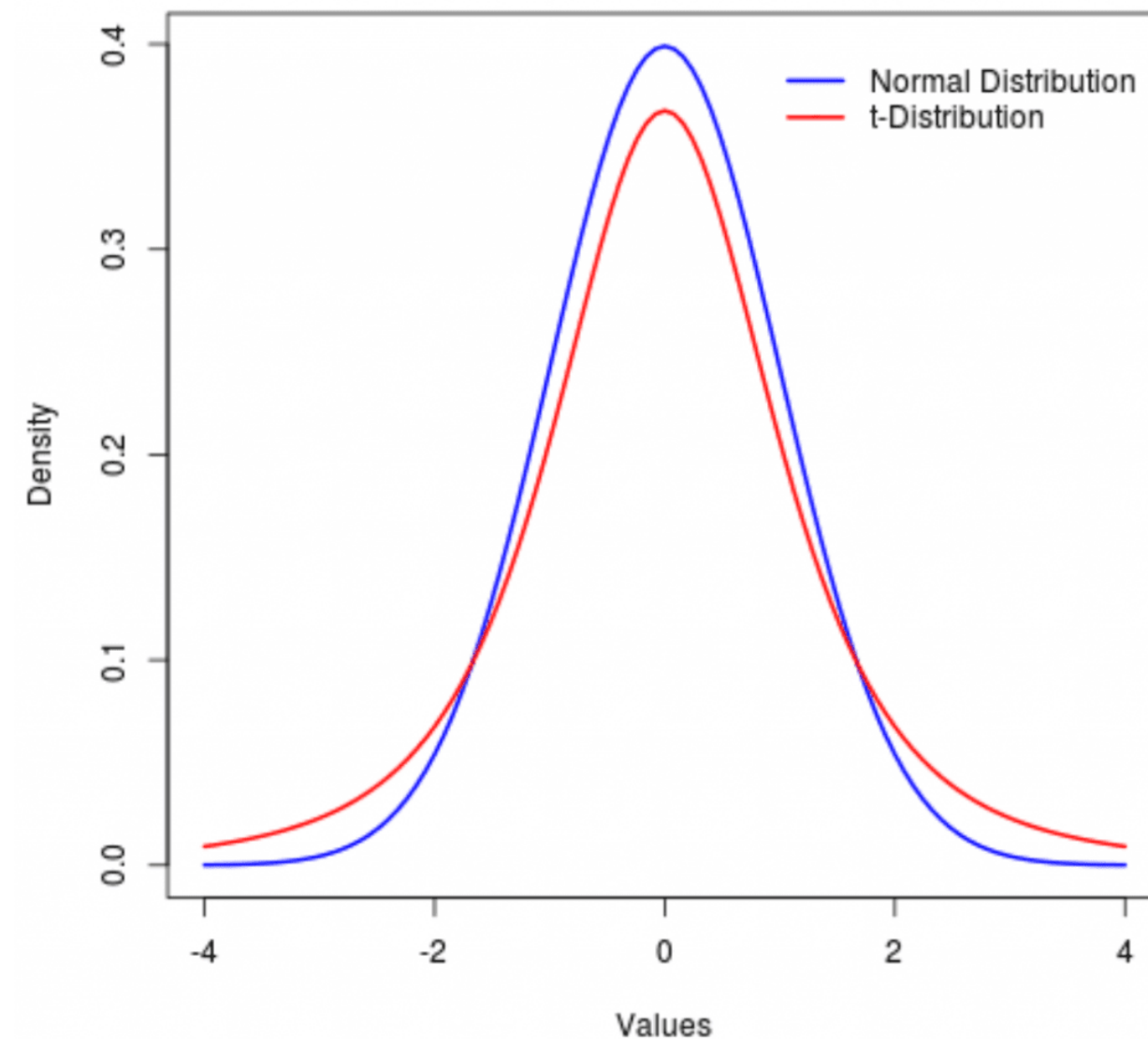
Null hypothesis $\rightarrow H_0: w = 0$
Alternate hypothesis $\rightarrow H_a: w \neq 0$

P-value (or *probability value*) is the probability of the Null Hypothesis being true.

The p-value is determined from a statistical value of the population distribution.

But what is the population distribution?

Hypothesis Testing



Significance level: 0.05

- **t distribution**
- **Used when:**
 - Small amount of data
 - Population's standard deviation is unknown
- **t-statistic**

$$t = \frac{w - w_0}{se(w)}$$

Where:

- **w** is the estimated weight
- **w₀** is the hypothesis weight (equals zero)
- **se(w)** is the standard error of the weight

Synthetic DTC: LWD + Chip Descriptions

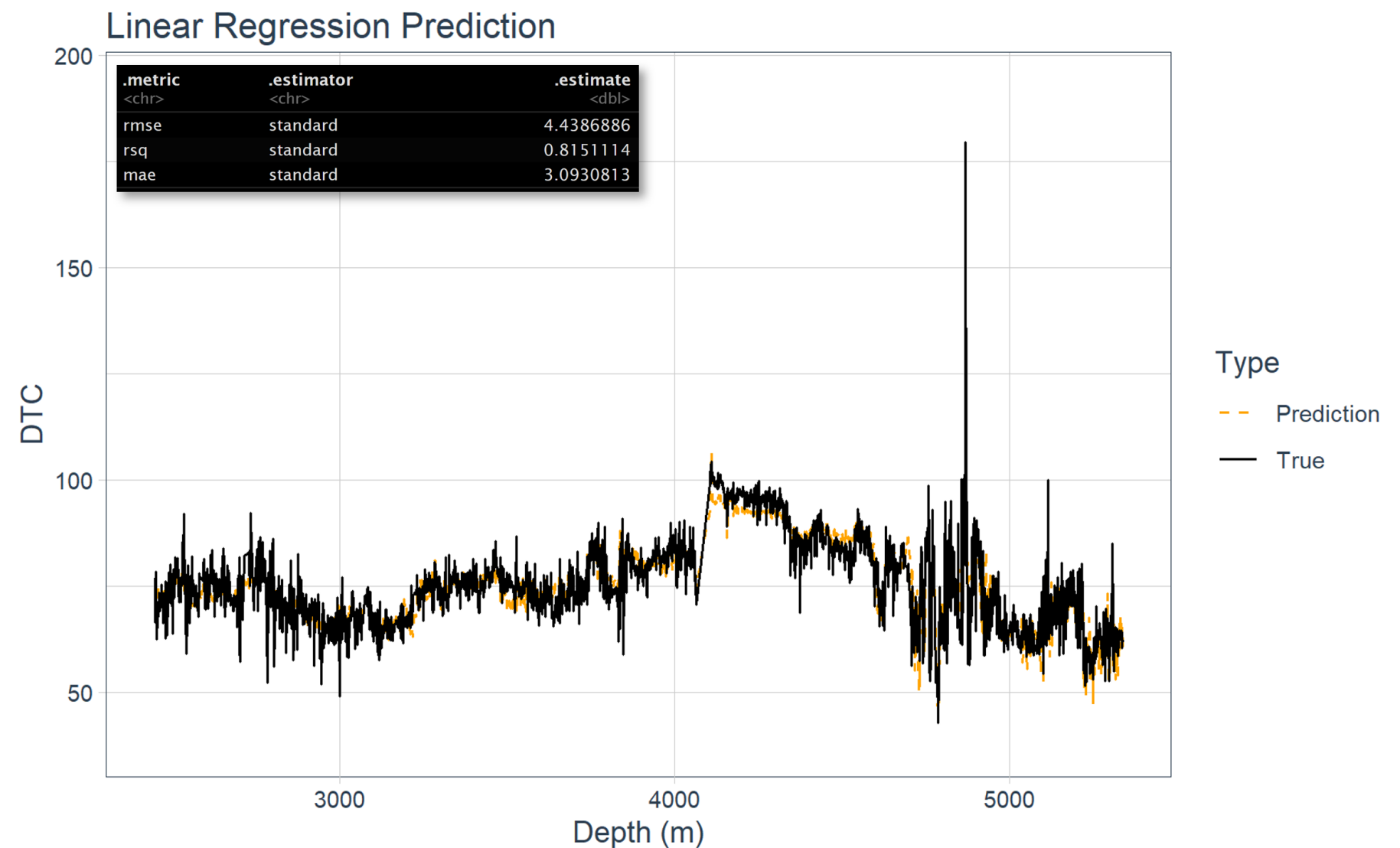
40/60 data split

```
Call:
lm(formula = DTC_Detrended ~ . - Depth - Trend_DTC - DTC, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-24.2906  -2.3174  -0.0601   2.4379  22.1395

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.07621    0.09041  -0.843  0.399360
Depth (m)     -5.10753    0.43234 -11.814 < 2e-16 ***
Argillaceous Calcilutite  0.28764    0.42570   0.676  0.499306
Calcarenite    -2.53034    0.75655  -3.345  0.000838 ***
Cement         0.34663    0.08092   4.284  1.92e-05 ***
Calcilutite    0.06983    0.44667   0.156  0.875787
Calcarenite Calcilutite -0.47073    0.14862  -3.167  0.001560 **
Chert         -0.44993    0.16017  -2.809  0.005010 **
Calcarenite Chert -0.40398    0.15794  -2.558  0.010602 *
Calcilutite Argillaceous Calcarenite 0.05612    0.14630   0.384  0.701306
Argillaceous Calcarenite -0.27558    0.12137  -2.271  0.023270 *
Argillaceous Calcisiltite 0.12594    0.08487   1.484  0.137975
Calcilutite 1  0.12880    0.11180   1.170  0.245495
Calcarenite Siltstone -0.13957    0.09888  -1.412  0.156227
Argillaceous Calcilutite 1 0.34530    0.09703   3.559  0.000381 ***
Marl          0.68583    0.26453   2.593  0.009587 **
Sandstone     0.24984    0.27965   0.893  0.371752
Marl Calcilutite 0.12855    0.13569   0.947  0.343524
Calcarenite Claystone 0.08277    0.35293   0.235  0.814508
Claystone     2.97673    0.60937   4.885  1.11e-06 ***
Claystone 1   0.20035    0.17072   1.174  0.240709
Contamination 0.29400    0.11964   2.457  0.014075 *
Siltstone     0.24599    0.51090   0.481  0.630218
Volcanic      0.38974    0.12842   3.035  0.002434 **
No Lithology  0.30034    0.13603   2.208  0.027349 *
Ferruginous Volcanic 1.10406    0.20897   5.283  1.39e-07 ***
Argillaceous Volcanic 0.30196    0.30407   0.993  0.320785
Siltstone 1   0.08152    0.11029   0.739  0.459887
Siltstone Siltstone 1 0.14090    0.06590   2.138  0.032627 *
Argillaceous Siltstone 0.31283    0.13287   2.354  0.018638 *
Ca (%)       -1.68043    0.67561  -2.487  0.012945 *
GR           2.27713    0.13907  16.374 < 2e-16 ***
RDEP        -4.72598    0.15345 -30.799 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.281 on 2240 degrees of freedom
Multiple R-squared:  0.8235,    Adjusted R-squared:  0.821
F-statistic: 326.7 on 32 and 2240 DF,  p-value: < 2.2e-16
```



$R^2 = 0.82$

Synthetic DTC: Step-wise regression

40/60 data split

```
Call:
lm(formula = DTC_Detrended ~ RDEP + Claystone + GR + Siltstone +
  Chert + Calcilutite + Marl + Cement + `Ferruginous Volcanic` +
  `Depth (m)` + `Ca (%)` + Calcarenite + `Argillaceous Calcilutite 1` +
  Volcanic + `Calcarenite Calcilutite` + Contamination + `Calcarenite Chert` +
  `No Lithology` + `Argillaceous Ca
  Siltstone Siltstone 1` + `Calci
  data = train)
```

Residuals:

Min	1Q	Median	3Q
-23.7128	-2.3492	-0.0624	2.4954

Coefficients:

	Estimate
(Intercept)	-0.06767
RDEP	-4.63892
Claystone	2.64004
GR	2.26257
Siltstone	-0.61817
Chert	-0.35745
Calcilutite	0.29630
Marl	0.73528
Cement	0.35096
`Ferruginous Volcanic`	1.05934
`Depth (m)`	-4.89432
`Ca (%)`	-2.69692
Calcarenite	-1.87118
`Argillaceous Calcilutite 1`	0.28086
Volcanic	0.31937
`Calcarenite Calcilutite`	-0.36713
Contamination	0.23988
`Calcarenite Chert`	-0.30700
`No Lithology`	0.23056
`Argillaceous Calcarenite`	-0.23653
`Argillaceous Siltstone`	0.25202
`Siltstone Siltstone 1`	0.13833
`Calcilutite 1`	0.21857
`Argillaceous Calcisiltite`	-0.12187

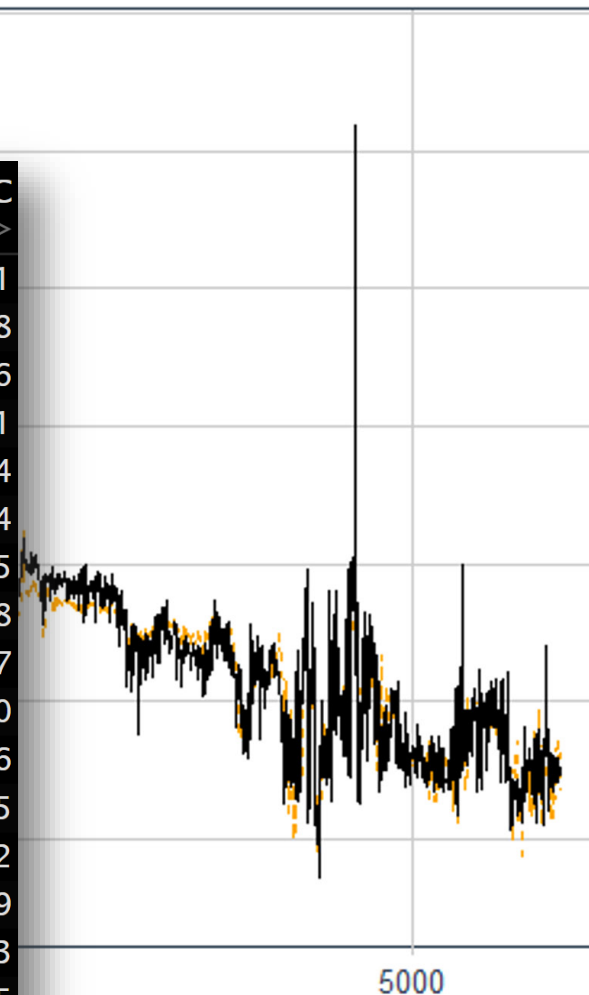
Signif. codes: 0 '***' 0.001 '**' 0.

Residual standard error: 4.281 on 224
Multiple R-squared: 0.8229, Adjusted
F-statistic: 454.2 on 23 and 2249 DF,

Step-wise Model Prediction

.metric <chr>	.estimator <chr>	.estimate <dbl>
rmse	standard	4.4545413
rsq	standard	0.8137509

Step <S3: AsIs>	Df <dbl>	Deviance <dbl>	Resid. Df <dbl>	Resid. Dev <dbl>	AIC <dbl>
	NA	NA	2272	232626.37	10522.201
+ RDEP	-1	164201.15914	2271	68425.21	7742.748
+ Claystone	-1	11905.59089	2270	56519.62	7310.256
+ GR	-1	3638.88705	2269	52880.73	7160.991
+ Siltstone	-1	4409.50446	2268	48471.22	6965.084
+ Chert	-1	1948.92161	2267	46522.30	6873.804
+ Calcilutite	-1	433.89126	2266	46088.41	6854.505
+ Marl	-1	411.73326	2265	45676.68	6836.108
+ Cement	-1	404.94381	2264	45271.73	6817.867
+ `Ferruginous Volcanic`	-1	407.18323	2263	44864.55	6799.330
+ `Depth (m)`	-1	370.41554	2262	44494.14	6782.486
+ `Ca (%)`	-1	1190.78662	2261	43303.35	6722.825
+ Calcarenite	-1	814.37080	2260	42488.98	6681.672
+ `Argillaceous Calcilutite 1`	-1	222.65322	2259	42266.33	6671.729
+ Volcanic	-1	179.73710	2258	42086.59	6664.043
+ `Calcarenite Calcilutite`	-1	173.82514	2257	41912.76	6656.635
+ Contamination	-1	107.77090	2256	41804.99	6652.783
+ `Calcarenite Chert`	-1	110.49629	2255	41694.50	6648.767
+ `No Lithology`	-1	99.43980	2254	41595.06	6645.340
+ `Argillaceous Calcarenite`	-1	96.34674	2253	41498.71	6642.069
+ `Argillaceous Siltstone`	-1	82.81256	2252	41415.90	6639.529
+ `Siltstone Siltstone 1`	-1	85.12202	2251	41330.77	6636.852
+ `Calcilutite 1`	-1	82.97880	2250	41247.80	6634.284
+ `Argillaceous Calcisiltite`	-1	39.81617	2249	41207.98	6634.089



Type

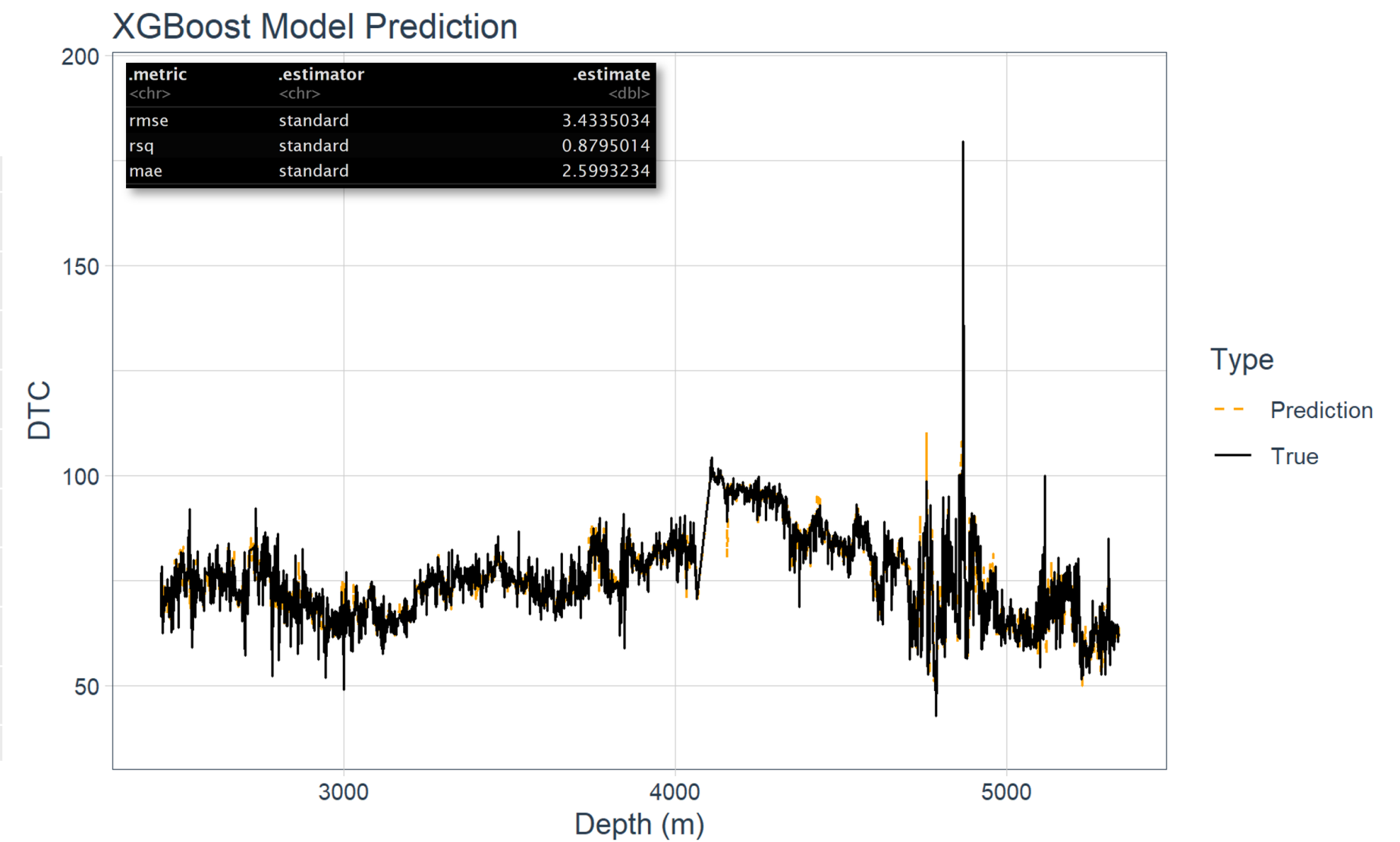
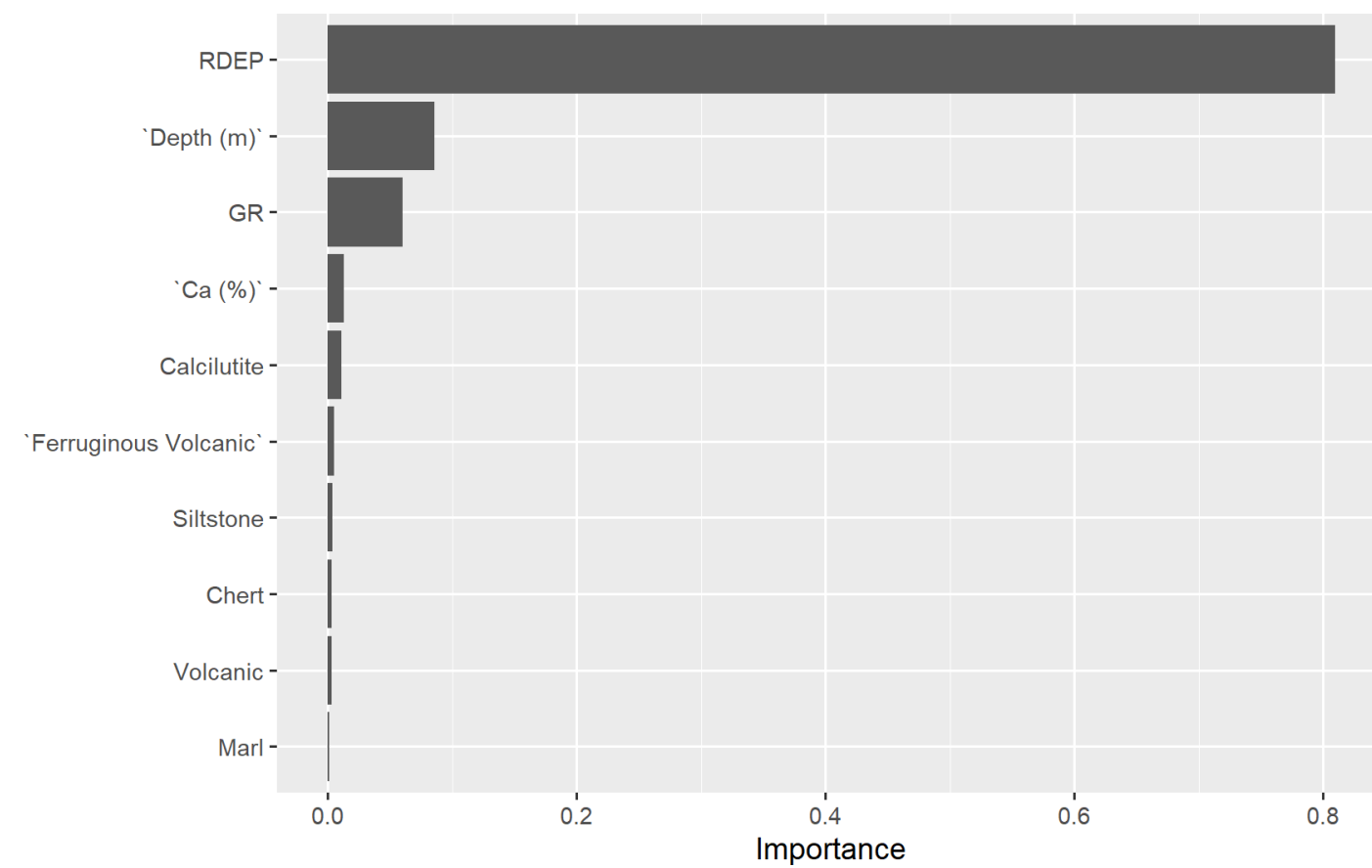
--- Prediction
— True

AIC = Akaike Information Criterion

Evaluate the relative quality of models.

Synthetic DTC: XGBoost

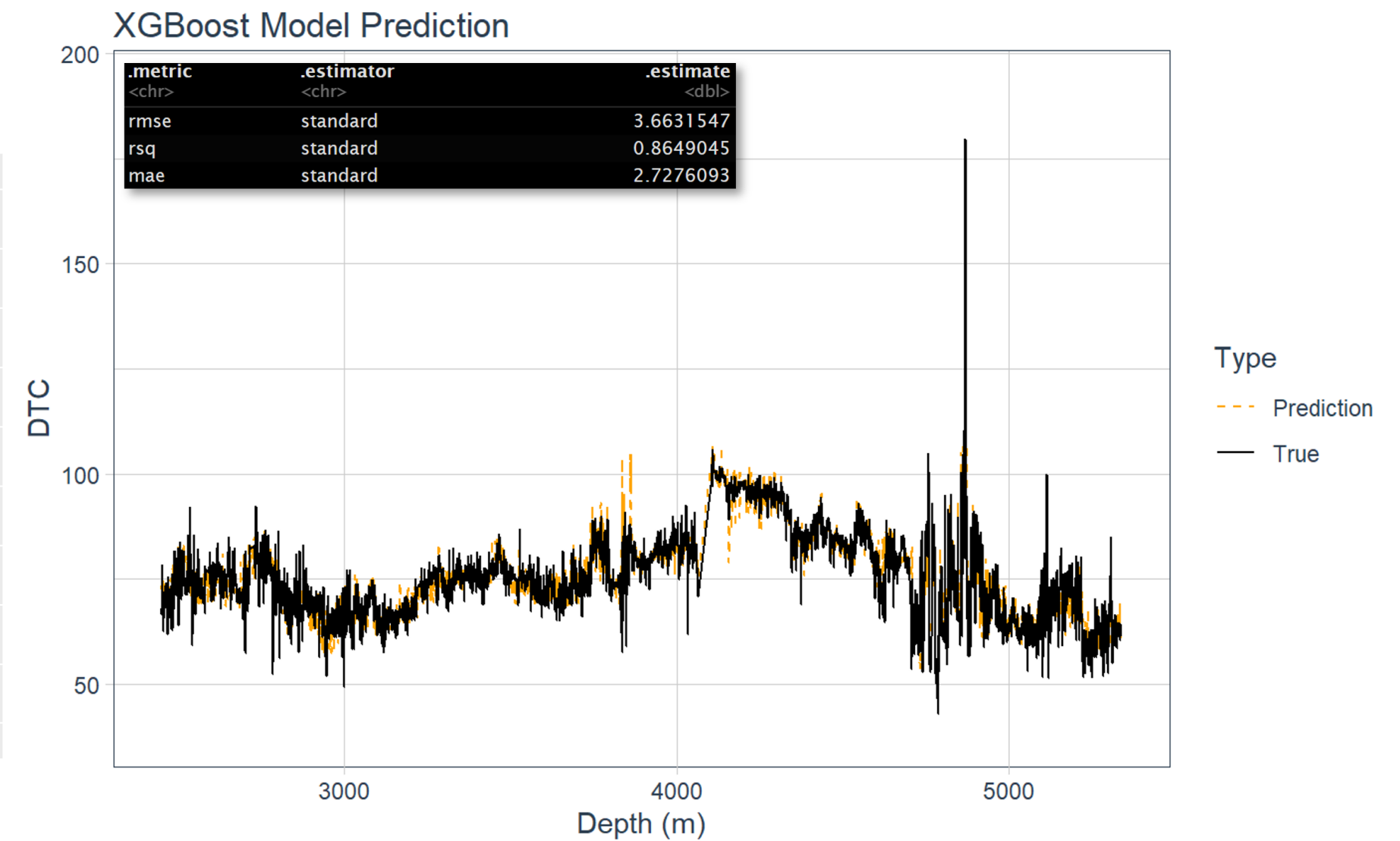
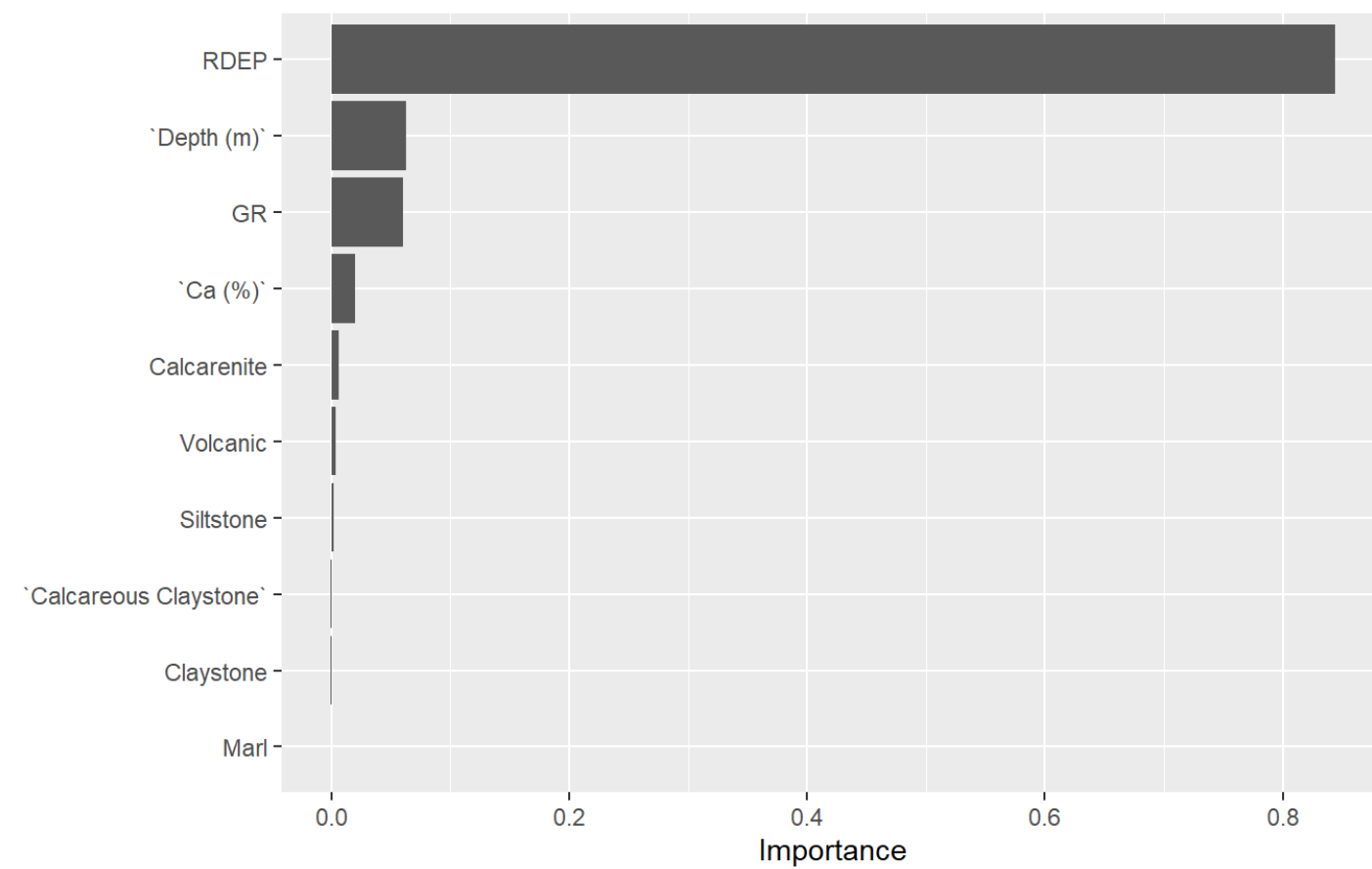
40/60 data split



$$R^2 = 0.88$$

Synthetic DTC: XGBoost

5/95 data split



$$R^2 = 0.86$$

Part 3:

CONCLUSIONS



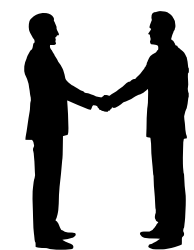
Conclusions

PDF to Table

- Mud-logging chip descriptions
- Complex tables in PDF
- Successfully converted to CSV tables
- Redo processes to other templates
- Build the app

Synthetic DTC

- Chip descriptions improve DTC predictions
- Step-wise regression selected statistically significant features
- XGBoost relies on LWD/Wireline logs to predict DTC
- All models were robust



Acknowledgments

- CREWES Sponsors
- NSERC
- CFREF and GRI
- All the co-authors
- CREWES students and staff

Thank You!



www.crewes.org