Data Science for Smart Manufacturing

IIT Madras || University of Texas || GIAN

Surface Characteristic Prediction Of Grinding Process

Anil Kharde

IIT BBS

Need of Data Science

- In order to know the outcome of the process before undergoing it we need to take help of past data of process.
- By analyzing the data we can put forward model equation which governs the process.
- Parameters affecting the response can be controlled to get desired output.
- This can be achieved by Data Science…

What is Data Science...?

- Scientific process of converting raw data into knowledge to support decision making.
- Science which deals with collecting, processing and analyzing the data to draw some insights which helps in decision making.
- Applications:

Manufacturing	Defense
Healthcare	Banking
Supply chain	Sports
E- commerce	Airlines
Tourism	Retail

- Goal of data science is to make business more competitive and improve it.
- Data analytics approaches provide an automated and cost-effective way to
- assure quality in manufactured products
- Manage nation’s engineering and information assets

Classification

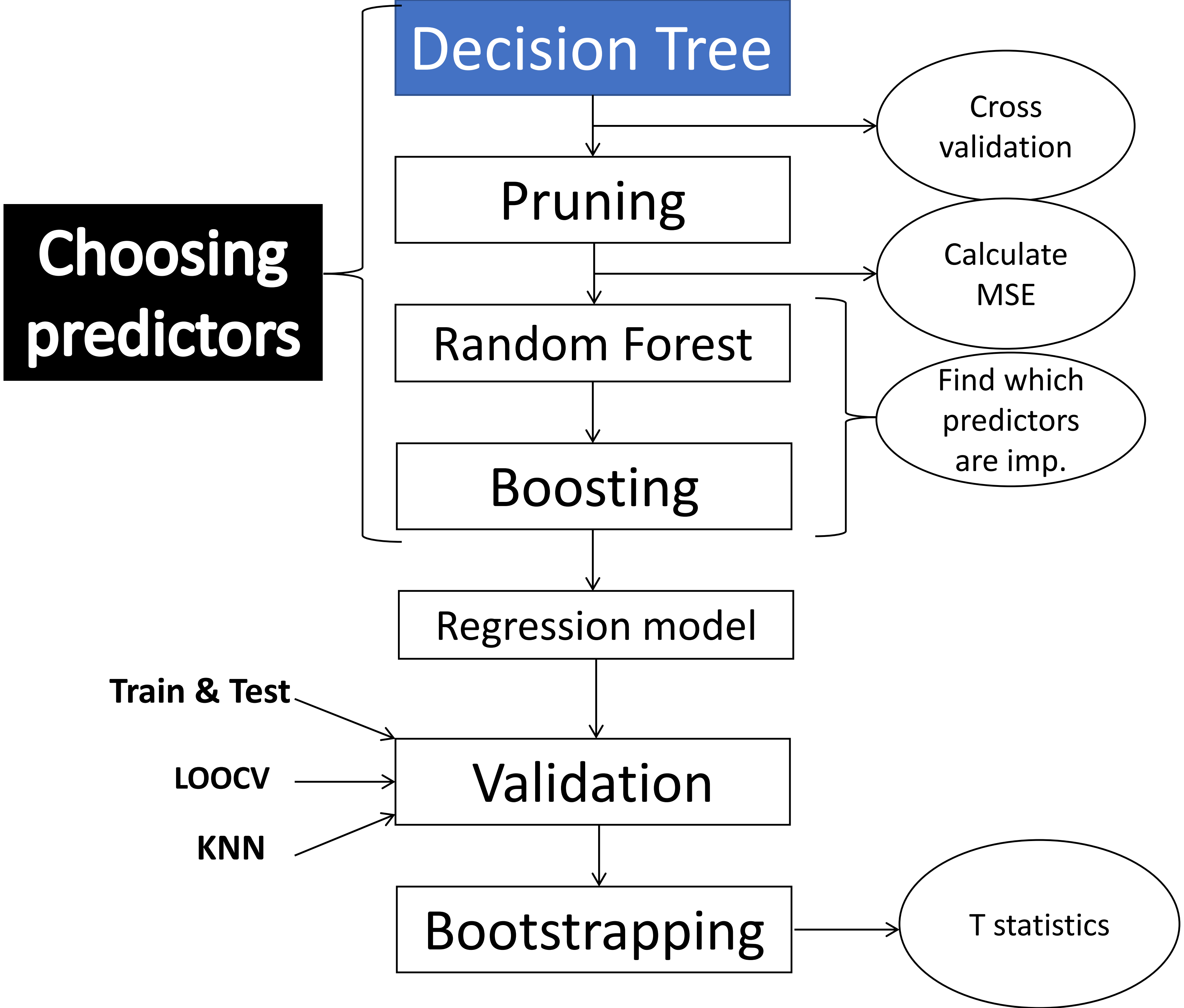
- Descriptive Analytics- It describes the data and gives information about the past.
- Predictive Analytics- Next step to descriptive analytics. It predicts the future outcome from the past data.
- Prescriptive analytics- Next step of predictive analytics. It involves the decision making and deciding the course of action.

Surface Roughness Prediction in Grinding process

- We have a data for various speed and feed
- Accelerometers, voltage and current sensors are mounted to check vibration and power parameters.
- Total 38 workpiece are grinded using various combination of controllable parameters in four stages.
- Wheel speed= 25, 35, 45 rpm
- Work speed= 100,200,300 rpm
- Feed rate=0.01, 0.05, 0.2, 0.4, 1.2 mm/rev
- Surface profilometer measures the surface roughness in µm.
- From power and accelerometer sensors various parameters are extracted

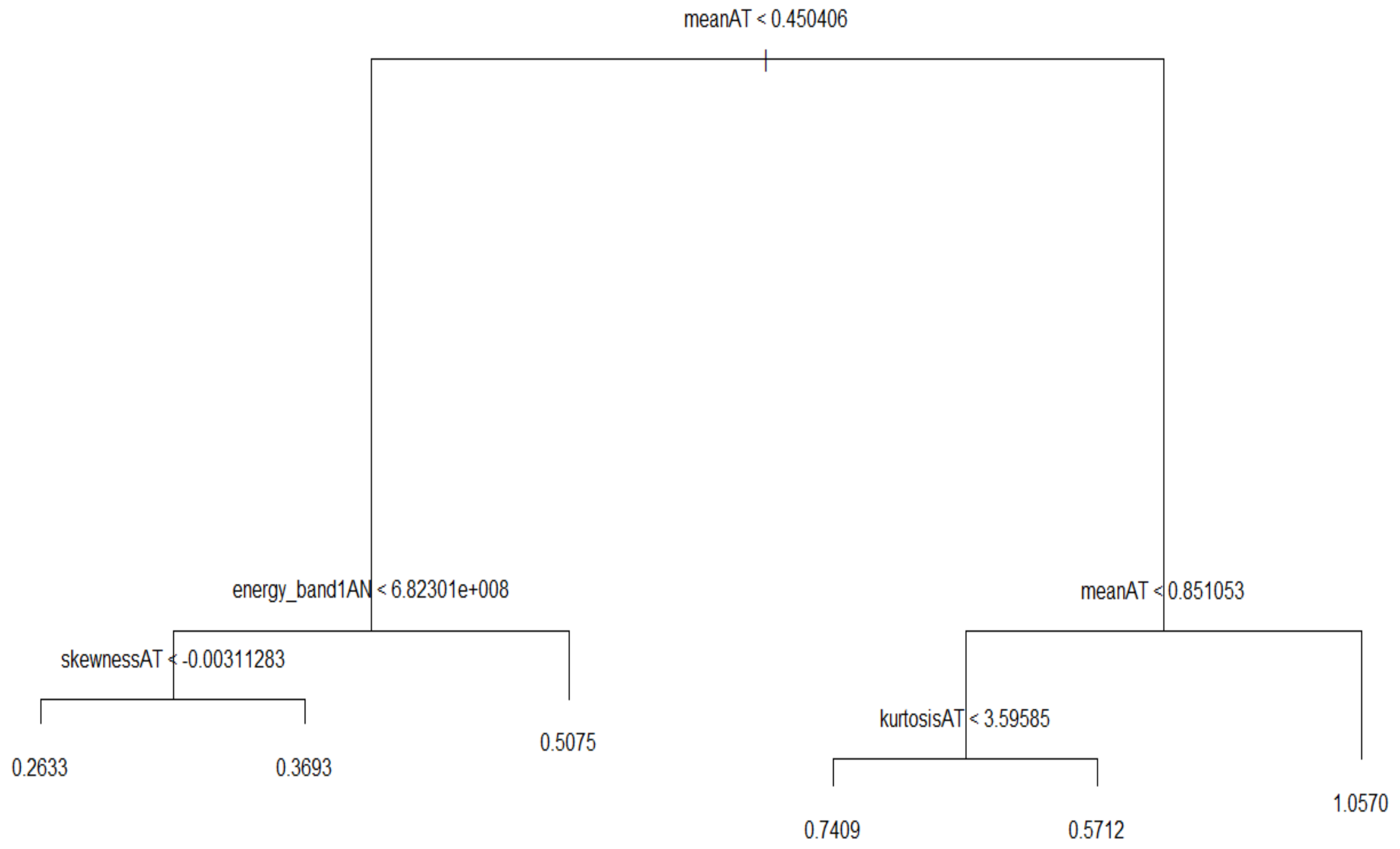
Objective

- To design a Regression model by analyzing the collected data set which can predict the surface roughness (Ra Value) of a randomly chosen work piece.
- Dependent variable: Roughness(Ra)
- Independent variables are 24
- Data
- Data visualization


Decision Tree
Cross validation
Pruning
Calculate MSE
Choosing predictors
Random Forest
Find which predictors are imp.
Boosting
Regression model
Train & Test
LOOCV
KNN
Validation
Bootstrapping
T statistics

Choosing the Right ones

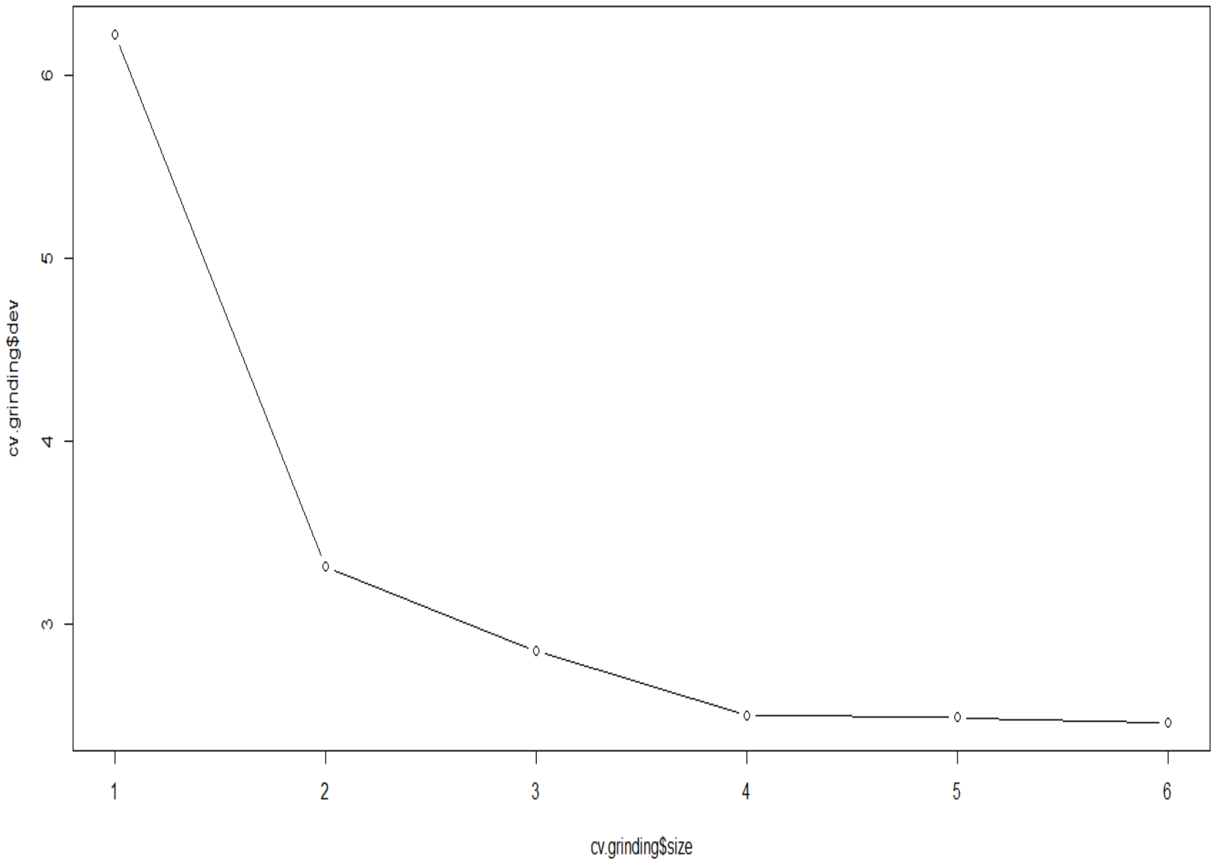
- An analysis was done to judge the better prediction terms among all the given parameters using Decision Trees.



Summary of the plot

Regression tree:
tree(formula = Ra ~ ., data = Final1, subset = train)
Variables actually used in tree construction:
[1] "meanAT"   "energy_band1AN" "skewnessAT"
"kurtosisAT"
Number of terminal nodes: 6
Residual mean deviance: 0.0115 = 1.081 / 94
Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.2429000	-0.0633100	0.0004025	0.0000000	0.0531300	0.4226000



Pruning as an option

- Pruning is done to ensure the decision tree has a better performance. It makes some changes in the tree structure only for the better.
- So through Cross Validation we decide the size which will be best suited for pruning the tree.

Summary of Cross Validation

```
cv.grinding
$size
[1] 6 5 4 3 2 1
$dev
[1] 2.462333 2.489448 2.503179 2.853785 3.314332 6.226398
$k
[1] -Inf 0.1483040 0.1657847 0.4142310 0.7798620 3.5092065
$method
[1] "deviance"
attr(,"class")
[1] "prune" "tree.sequence“
```

So it is to be noted that 6 is the size which we initially got. Hence no changes are to be made. We go for MSE next.

Calculations of MSE

```
yhat=predict(tree.grinding,newdata=Final1[-train,])
grinding.test=Final1[-train,"Ra"]
mean((yhat-grinding.test)^2)
```

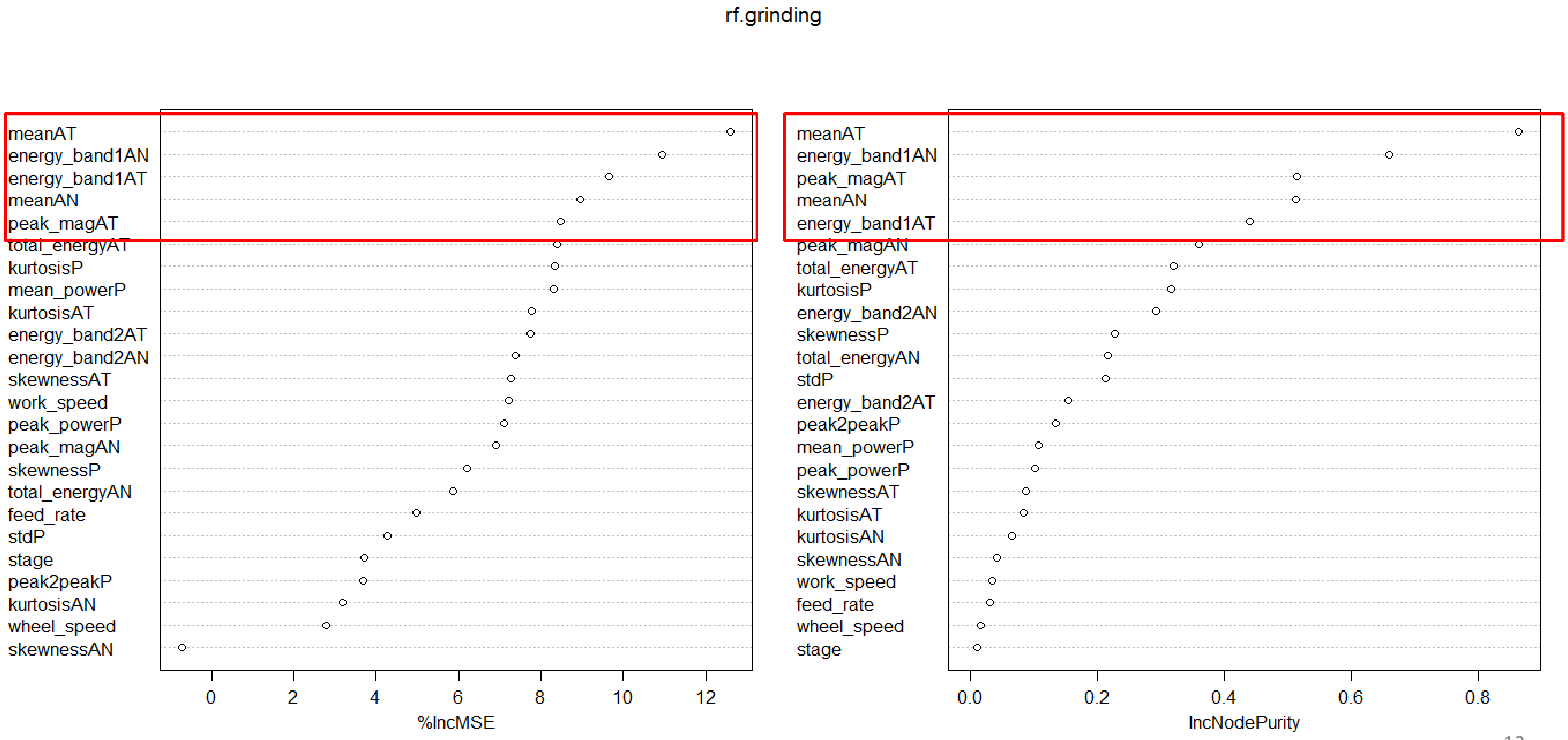
MSE: 0.01275293

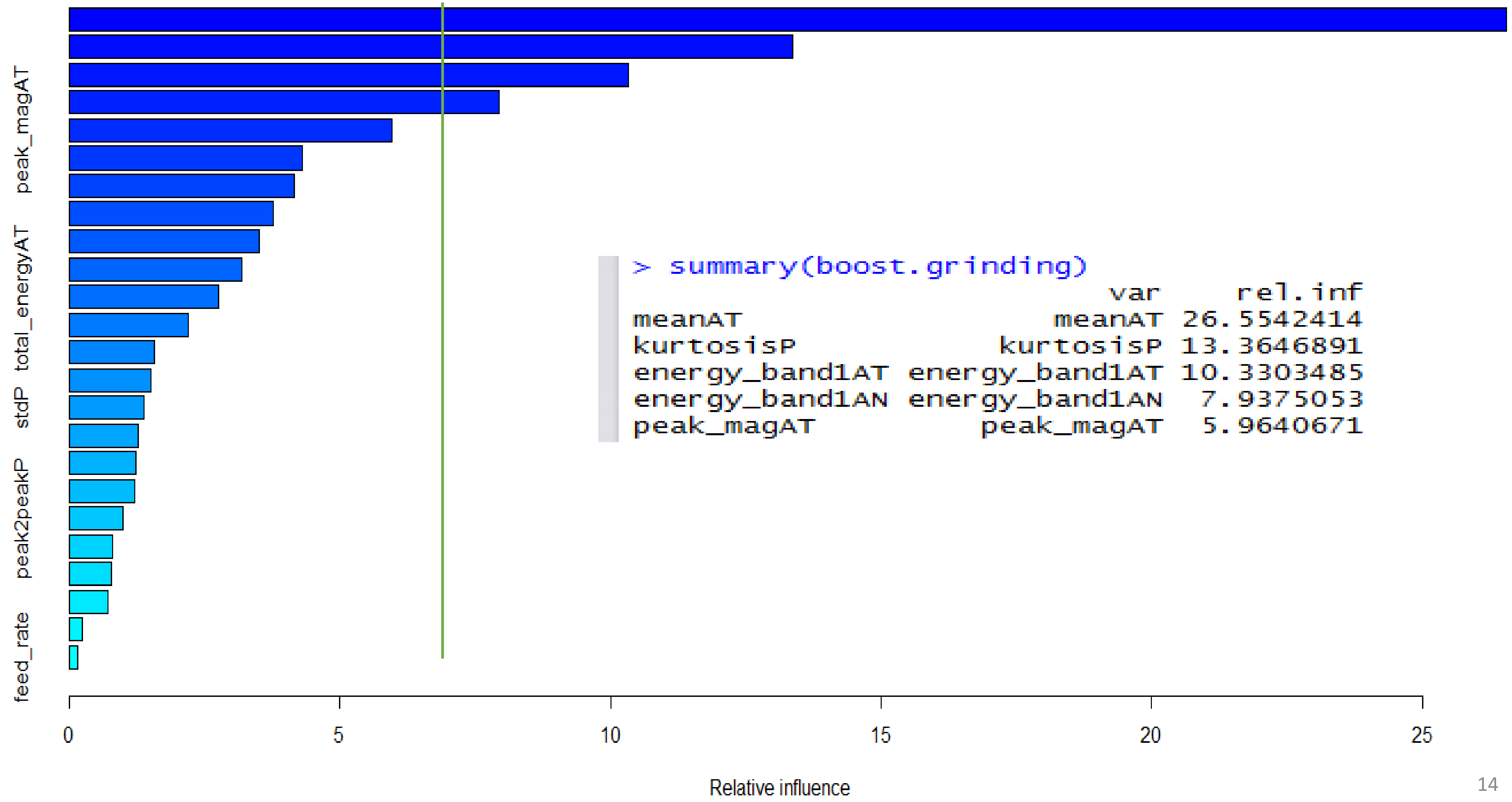
Random Forest…

- This method is far more popular and advanced.
- randomForest(formula = Ra ~ ., data = Final1, mtry = 5, importance = TRUE, subset = train)

- Type of random forest: regression
- Number of trees: 500
- No. of variables tried at each split: 5
- Mean of squared residuals: 0.01500107
- % Var explained: 75.4

Importance of each variable





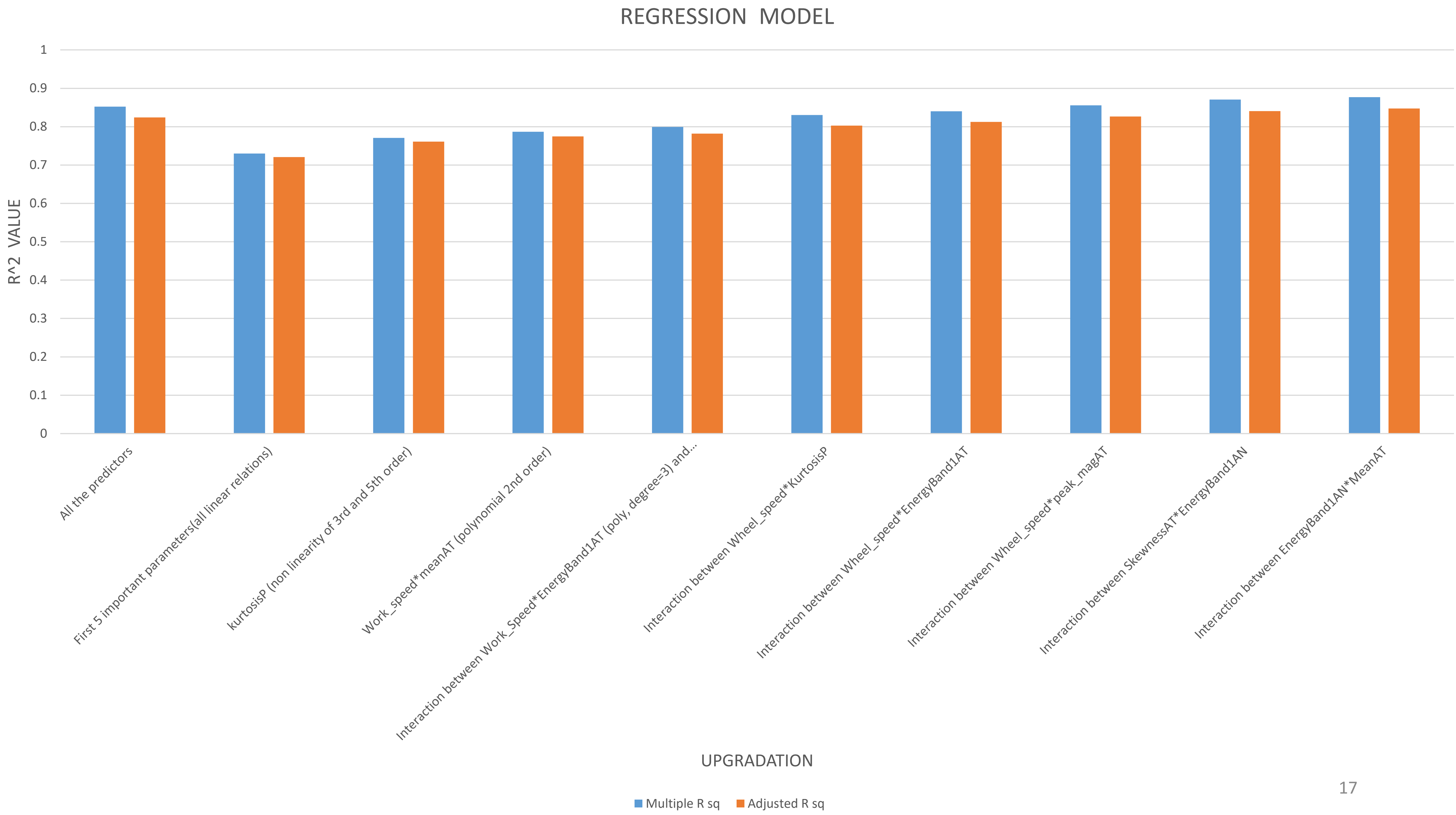

```
> summary(boost.grinding)
                         var    rel.inf
meanAT                meanAT 26.5542414
kurtosisP          kurtosisP 13.3646891
energy_band1AT energy_band1AT 10.3303485
energy_band1AN energy_band1AN  7.9375053
peak_magAT        peak_magAT  5.9640671
```

Building the Model…

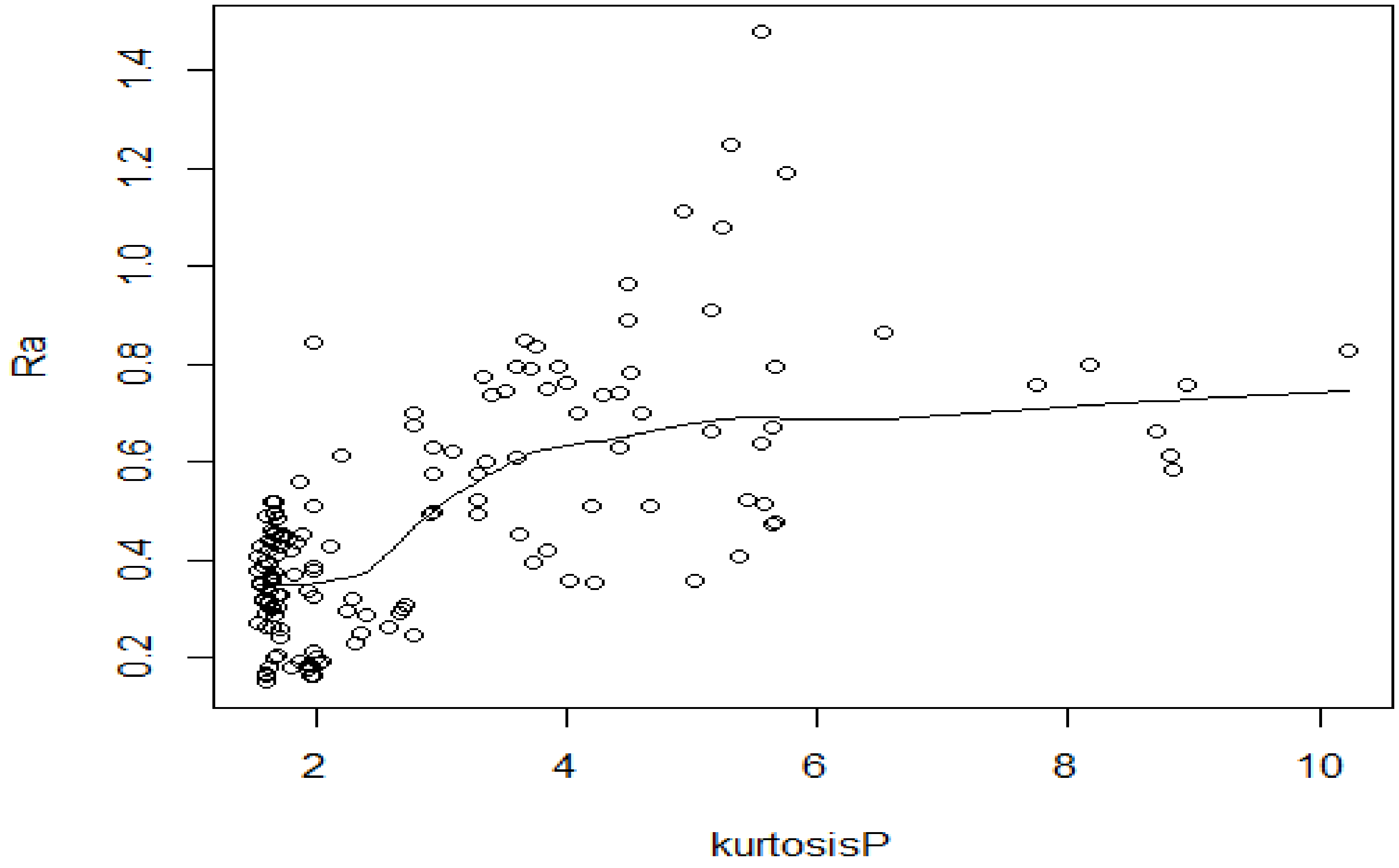
Iteration no.	Upgradation	Residual SE	Multiple R sq	Adjusted R sq
1	All the predictors	0.101	0.8523	0.8239
2	First 5 important parameters (all linear relations)	0.1273	0.7301	0.7207
3	kurtosisP (non linearity of 3rd and 5th order)	0.1177	0.7708	0.7612
4	Work_speed*meanAT (polynomial 2nd order)	0.1144	0.7866	0.7745
5	Interaction between Work_Speed*EnergyBand1AT (poly, degree=3) and Work_speed*EnergyBand1AN	0.1125	0.7994	0.7818
6	Interaction between Wheel_speed*KurtosisP	0.107	0.8306	0.8028
7	Interaction between Wheel_speed*EnergyBand1A T	0.1043	0.8403	0.8126

Iteration no.	Upgradation	Residual SE	Multiple R sq	Adjusted R sq
8	Interaction between Wheel_speed*peak_magAT	0.1003	0.8556	0.8265
9	Interaction between SkewnessAT*EnergyBand1AN	0.09609	0.8707	0.8408
10	Interaction between EnergyBand1AN*MeanAT	0.09413	0.877	0.8473

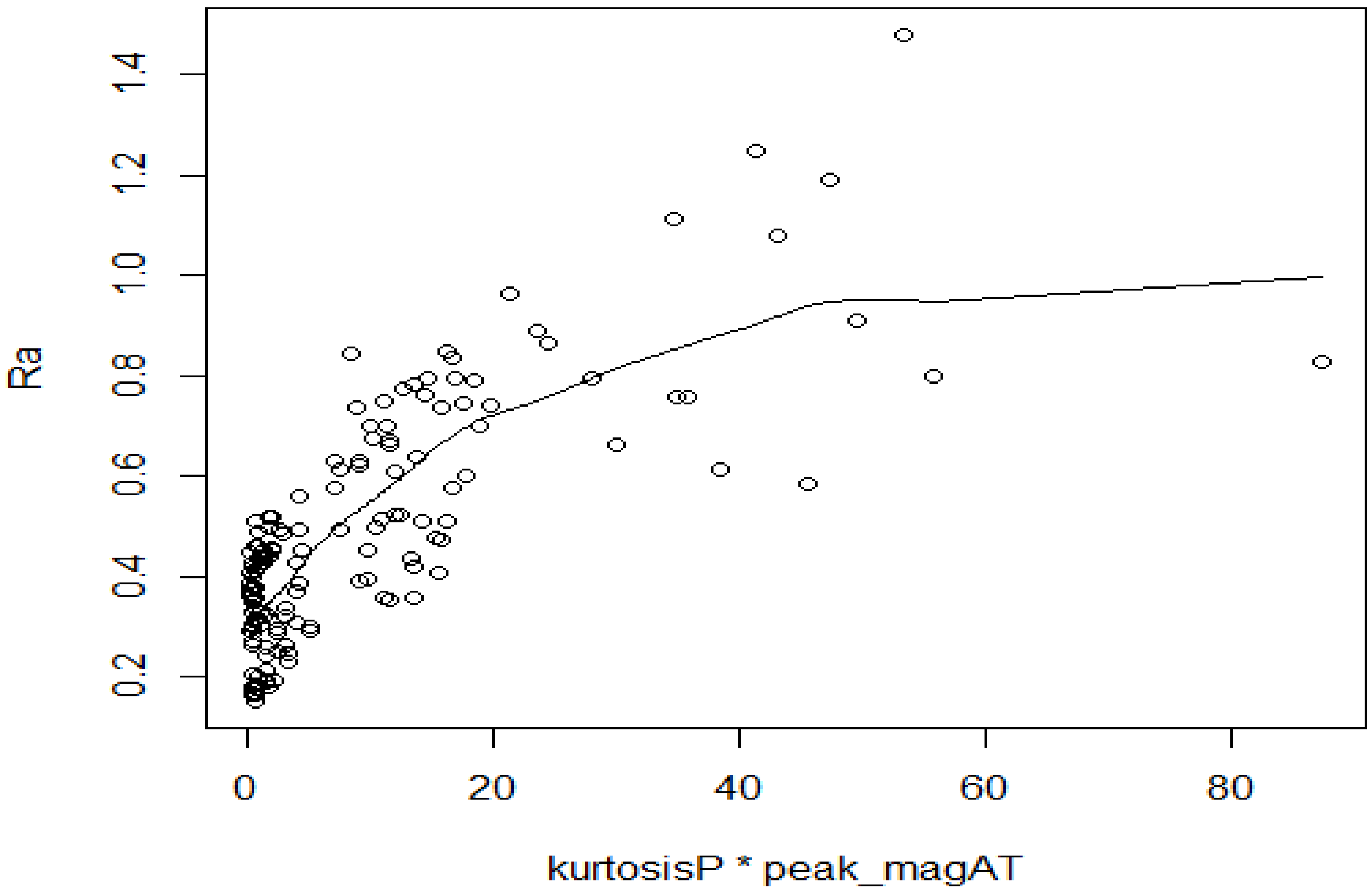
```
lm.fit=lm(Ra~meanAT+I(kurtosisP^3)+I(kurtosisP^5)+energy_band1AT+energy_band1AN+peak_magAT+poly(work_speed*meanAT,2)+poly(work_speed*energy_band1AT,3)+work_speed*energy_band1AN+poly(feed_rate*peak_magAT,5)+poly(wheel_speed*meanAT,3)+wheel_speed:kurtosisP+wheel_speed:energy_band1AT+wheel_speed*energy_band1AN+poly(wheel_speed*peak_magAT,2)+poly(skewnessAT*energy_band1AN,3)+energy_band1AN*meanAT, data=Final1)
```


REGRESSION MODEL
R^2 VALUE
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
All the predictors
First 5 important parameters(all linear relations)
kurtosisP (non linearity of 3rd and 5th order)
Work_speed*meanAT (polynomial 2nd order)
Interaction between Work_Speed*EnergyBand1AT (poly, degree=3) and...
Interaction between Wheel_speed*KurtosisP
Interaction between Wheel_speed*EnergyBand1AT
Interaction between Wheel_speed*peak_magAT
Interaction between SkewnessAT*EnergyBand1AN
Interaction between EnergyBand1AN*MeanAT
UPGRADATION
Multiple R sq
Adjusted R sq

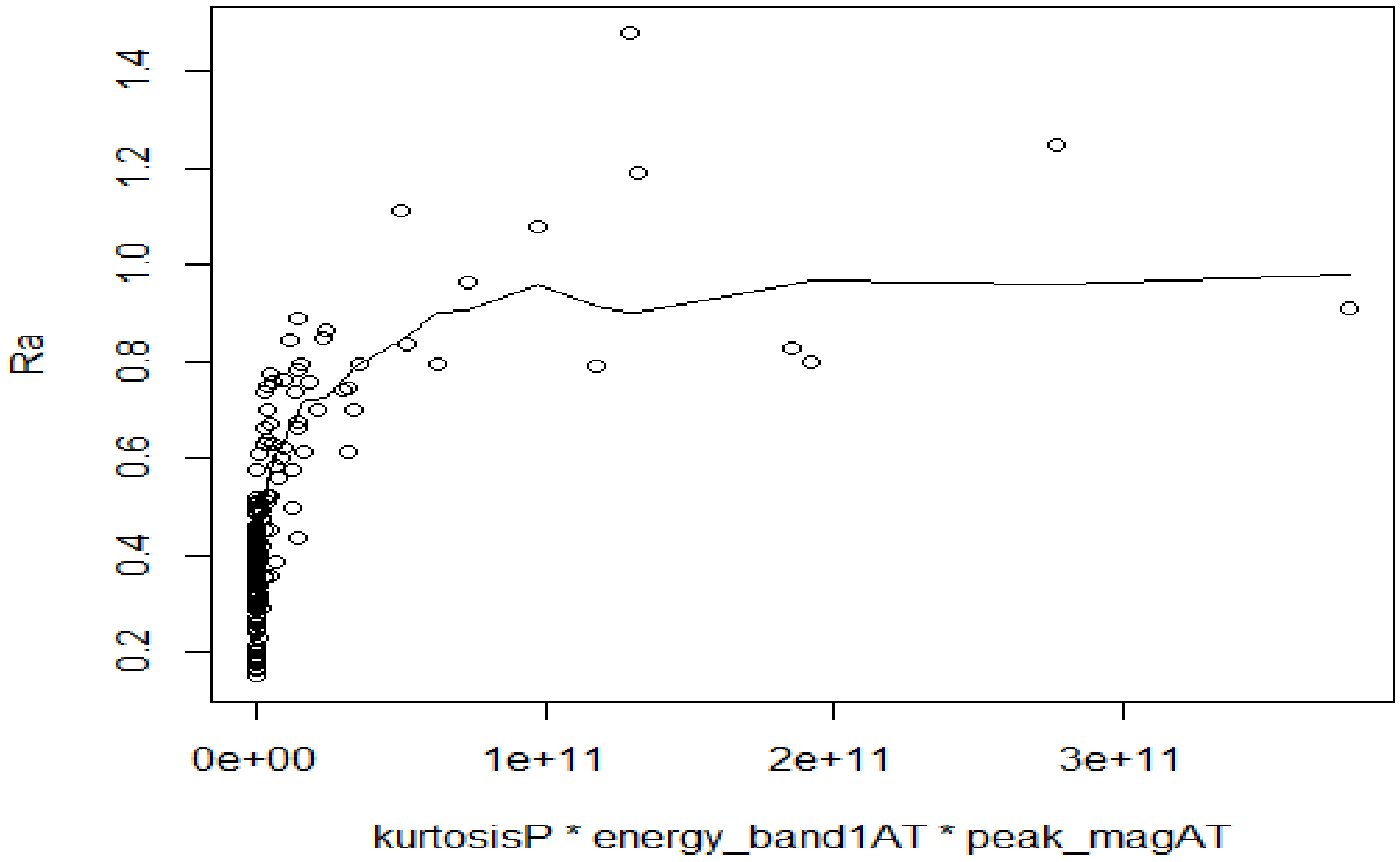
How to decide which order polynomial will help?



Why use interaction of predictors?



Why reject some predictors?

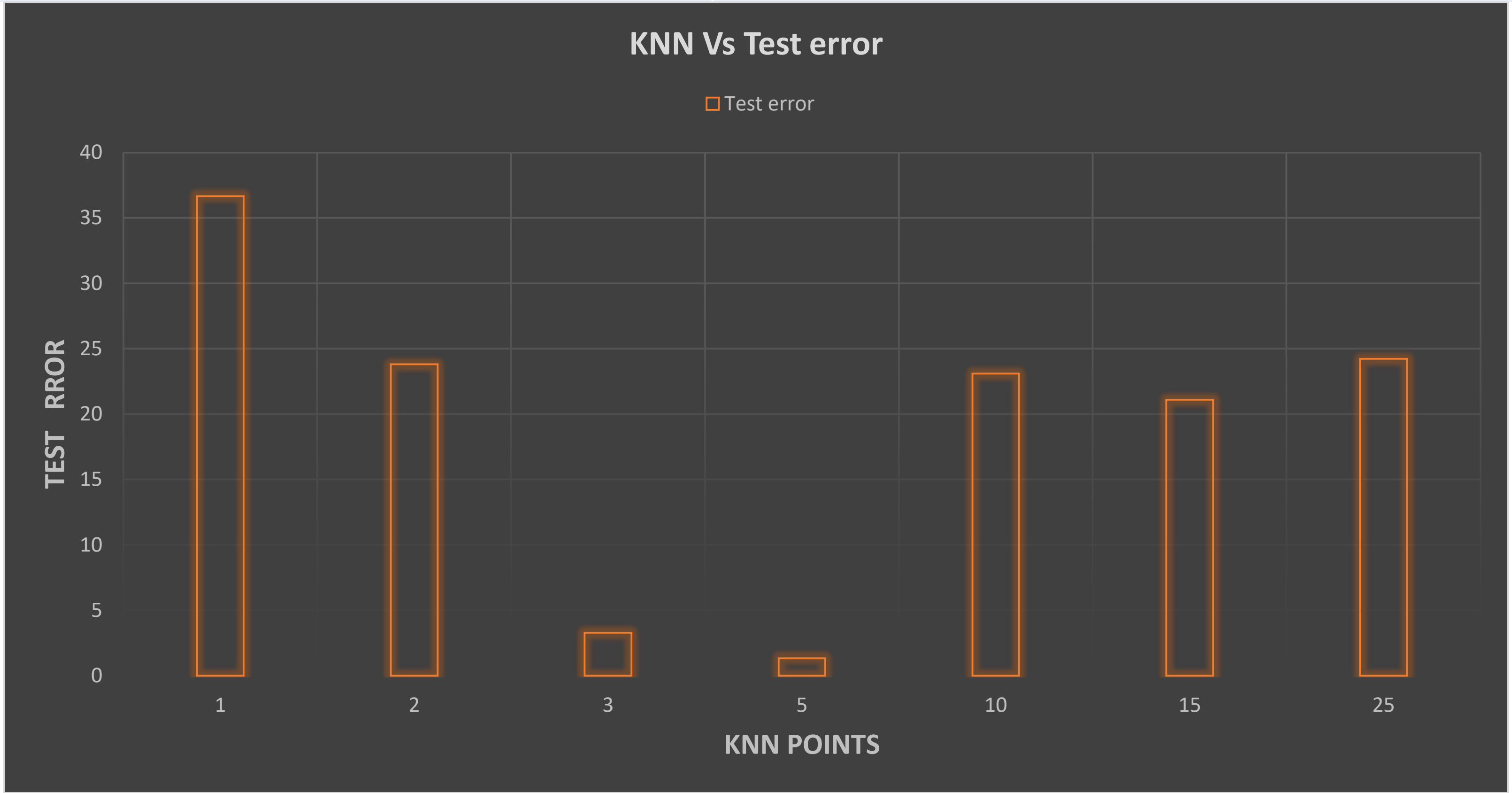


R implementation…

```
##validation set approach
library(boot)
##By taking sample
lm.fit=lm(Ra~meanAT+I(kurtosisP^3)+I(kurtosisP^5)+energy_band1AT+energy_band1AN+peak_
magAT+poly(work_speed*meanAT,2)+poly(work_speed*energy_band1AT,3)+I(work_speed*ener
gy_band1AN)+poly(feed_rate*peak_magAT,5)+poly(wheel_speed*meanAT,3)+wheel_speed:kurt
osisP+wheel_speed:energy_band1AT+I(wheel_speed*energy_band1AN)+poly(wheel_speed*pea
k_magAT,2)+poly(skewnessAT*energy_band1AN,3)+I(energy_band1AN*meanAT),data=Final1,su
bset=train)
mean((Ra-predict(lm.fit,Final1))[-train]^2)
##LOOCV
glm.fit=glm(Ra~meanAT+I(kurtosisP^3)+I(kurtosisP^5)+energy_band1AT+energy_band1AN+peak
_magAT+poly(work_speed*meanAT,2)+poly(work_speed*energy_band1AT,3)+I(work_speed*ene
rgy_band1AN)+poly(feed_rate*peak_magAT,5)+poly(wheel_speed*meanAT,3)+wheel_speed:kur
tosisP+wheel_speed:energy_band1AT+I(wheel_speed*energy_band1AN)+poly(wheel_speed*pe
ak_magAT,2)+poly(skewnessAT*energy_band1AN,3)+I(energy_band1AN*meanAT),data=Final1)
cv.error=cv.glm(Final1,glm.fit)
cv.error$delta
##K fold cross validation
cv.error.KNN=cv.glm(Final1,glm.fit,K=5)
cv.error.KNN$delta
```

Validation

Validation Method	Test Error
Train & Test Validation set Approach	0.01362
LOOCV	27.54



Bootstrapping

```
##Bootstrap
boot.fn=function(Final1,index)return(coef(lm(Ra~meanAT+I(kurtosisP^3)+I(kurtosisP^5)+ene
rgy_band1AT+energy_band1AN+peak_magAT+poly(work_speed*meanAT,2)+poly(work_spee
d*energy_band1AT,3)+I(work_speed*energy_band1AN)+poly(feed_rate*peak_magAT,5)+pol
y(wheel_speed*meanAT,3)+wheel_speed:kurtosisP+wheel_speed:energy_band1AT+I(wheel_
speed*energy_band1AN)+poly(wheel_speed*peak_magAT,2)+poly(skewnessAT*energy_ban
d1AN,3)+I(energy_band1AN*meanAT),data=Final1,subset=index)))
boot.fn(Final1,1:150)

boot.fn(Final1,sample(150,150,replace=T))
boot(Final1,boot.fn,1000)

summary(lm(Ra~meanAT+I(kurtosisP^3)+I(kurtosisP^5)+energy_band1AT+energy_band1AN+
peak_magAT+poly(work_speed*meanAT,2)+poly(work_speed*energy_band1AT,3)+I(work_sp
eed*energy_band1AN)+poly(feed_rate*peak_magAT,5)+poly(wheel_speed*meanAT,3)+whe
el_speed:kurtosisP+wheel_speed:energy_band1AT+I(wheel_speed*energy_band1AN)+poly(
wheel_speed*peak_magAT,2)+poly(skewnessAT*energy_band1AN,3)+I(energy_band1AN*m
eanAT),data=Final1))$coef
```

T-Statistics

Prediction terms	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.336925e-02	1.630400e-01	0.2046691	8.381774e-01
meanAT	1.360546e+00	5.681839e-01	2.3945520	1.818847e-02
I(kurtosisP^3)	2.187780e-03	9.565892e-04	2.2870635	2.394439e-02
I(kurtosisP^5)	-1.428913e-05	8.444362e-06	-1.6921507	9.321231e-02
energy_band1AT	3.519608e-10	2.077994e-10	1.6937531	9.290638e-02
energy_band1AN	-3.039573e-10	8.667912e-11	-3.5066961	6.390766e-04
peak_magAT	-1.420527e-02	7.256045e-02	-0.1957715	8.451201e-01
poly(work_speed * meanAT, 2)1	-6.586234e-01	7.173191e-01	-0.9181735	3.603698e-01
poly(work_speed * meanAT, 2)2	6.641240e-01	3.384097e-01	1.9624853	5.202031e-02
poly(work_speed * energy_band1AT, 3)1	-2.209709e+00	1.510211e+00	-1.4631792	1.460327e-01
poly(work_speed * energy_band1AT, 3)2	1.689080e+00	5.316178e-01	3.1772452	1.890785e-03
poly(work_speed * energy_band1AT, 3)3	1.257785e-01	2.259181e-01	0.5567438	5.787393e-01
I(work_speed * energy_band1AN)	6.058552e-13	1.730340e-13	3.5013661	6.507767e-04
poly(feed_rate * peak_magAT, 5)1	8.597668e-02	4.624705e-01	0.1859074	8.528313e-01
poly(feed_rate * peak_magAT, 5)2	6.628270e-01	2.691303e-01	2.4628482	1.520293e-02
poly(feed_rate * peak_magAT, 5)3	3.999004e-01	1.971681e-01	2.0282202	4.475181e-02
poly(feed_rate * peak_magAT, 5)4	4.870610e-02	1.875995e-01	0.2596281	7.955955e-01
poly(feed_rate * peak_magAT, 5)5	9.250714e-02	1.718965e-01	0.5381560	5.914661e-01
poly(wheel_speed * meanAT, 3)1	-9.657315e-01	2.630368e+00	-0.3671469	7.141561e-01
poly(wheel_speed * meanAT, 3)2	7.370640e-01	5.833318e-01	1.2635416	2.088438e-01
poly(wheel_speed * meanAT, 3)3	-5.298450e-01	1.911535e-01	-2.7718306	6.463545e-03
I(wheel_speed * energy_band1AN)	7.708471e-12	2.290698e-12	3.3651183	1.027990e-03
poly(wheel_speed * peak_magAT, 2)1	-3.210185e-01	1.921859e+00	-0.1670354	8.676233e-01
poly(wheel_speed * peak_magAT, 2)2	-1.207514e+00	4.204559e-01	-2.8719172	4.825164e-03
poly(skewnessAT * energy_band1AN, 3)1	-1.050139e+00	2.496081e-01	-4.2071501	5.014776e-05
poly(skewnessAT * energy_band1AN, 3)2	-2.055776e+00	5.027815e-01	-4.0888051	7.877960e-05
poly(skewnessAT * energy_band1AN, 3)3	-1.071476e+00	3.301953e-01	-3.2449779	1.522174e-03
I(energy_band1AN * meanAT)	-9.703458e-11	3.928192e-11	-2.4702100	1.490876e-02
wheel_speed:kurtosisP	-2.376067e-03	8.001043e-04	-2.9696971	3.600786e-03
energy_band1AT:wheel_speed	-8.983363e-13	7.259787e-12	-0.1237414	9.017270e-01

Conclusion

- The final model presented above has 29 prediction terms.
  - R^2value = 0.877 Adjusted R^2 = 0.8473
  - Residual Standard Error = 0.09413
  - Mean Square Error = 0.01362

Recap…

- Surface Roughness Prediction in Grinding process

Response- Surface roughness

Factors(24)- Process parameter(4), Acceleration and power signal(20) parameters

- Model R^2 was 87% , even then process parameter where not significant.
- Need to rebuild the model

- Phenomenon of “causation”
- Quote-

“Torture the data, and it will confess to anything.”- **Ronald Coase**, British economist and author

- An attempt is made to rebuild the model

Case 1: considering only process parameters

Case 2: considering process, acceleration and power signal parameter.

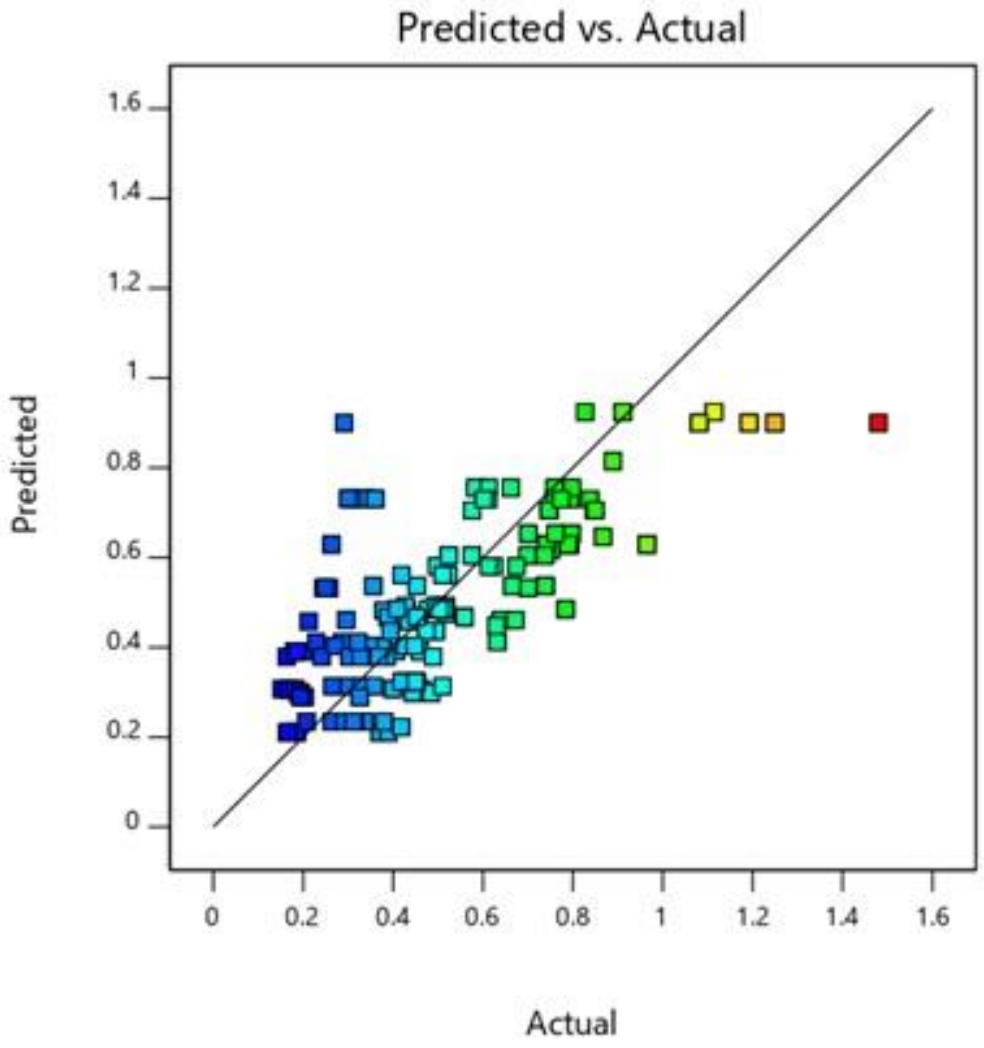
Case 1: a] Linear model

Roughness	=
+0.934089	
-0.008455	wheel speed
-0.000127	work speed
+0.242923	feed rate
-0.076451	stage

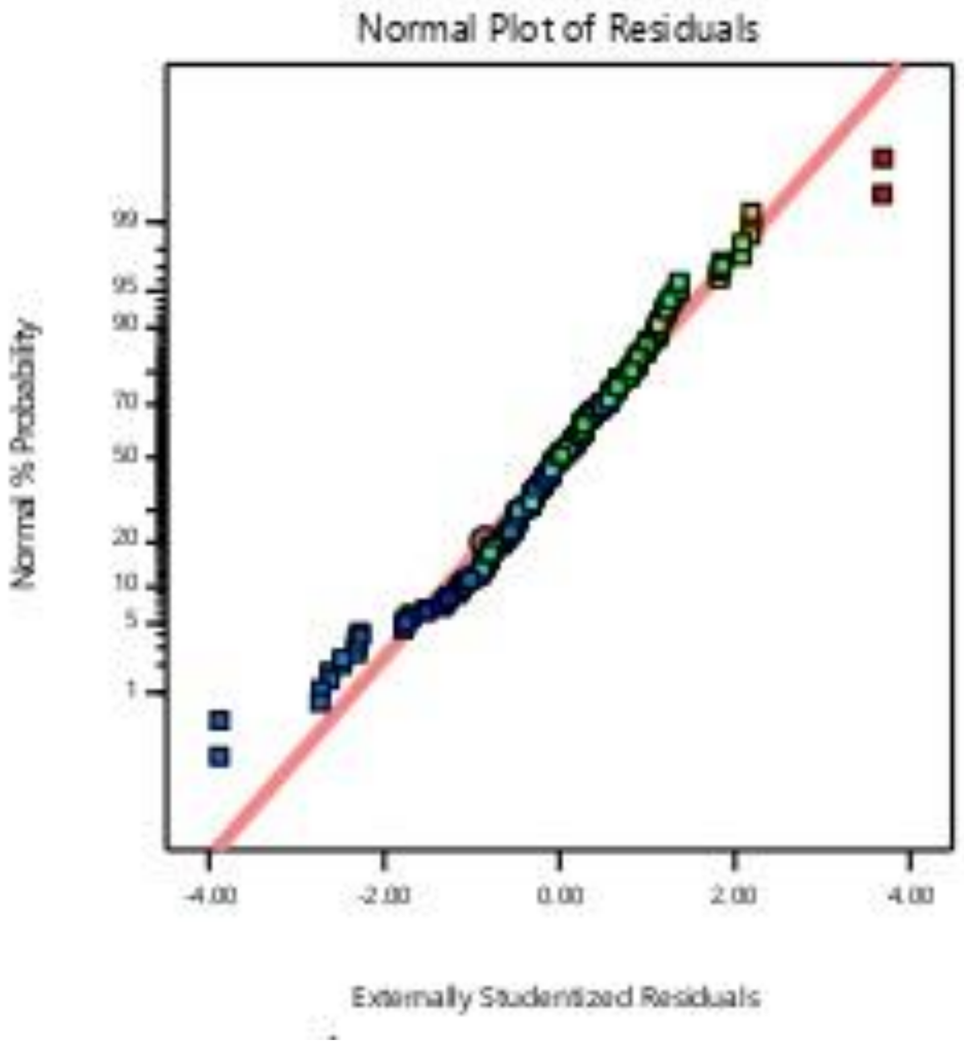
Std. Dev.	0.1628		R^2	0.5691
Mean	0.4984		Adjusted R^2	0.5610
C.V. %	32.66			

Source	Term df	Error df	F-value	p-value	
Subplot	4	315.39	105.13	< 0.0001	significant
a-wheel speed	1	318.97	74.21	< 0.0001	
B-work speed	1	291.31	1.37	0.2420	
C-feed rate	1	317.27	37.13	< 0.0001	
D-stage	1	317.15	32.08	< 0.0001	

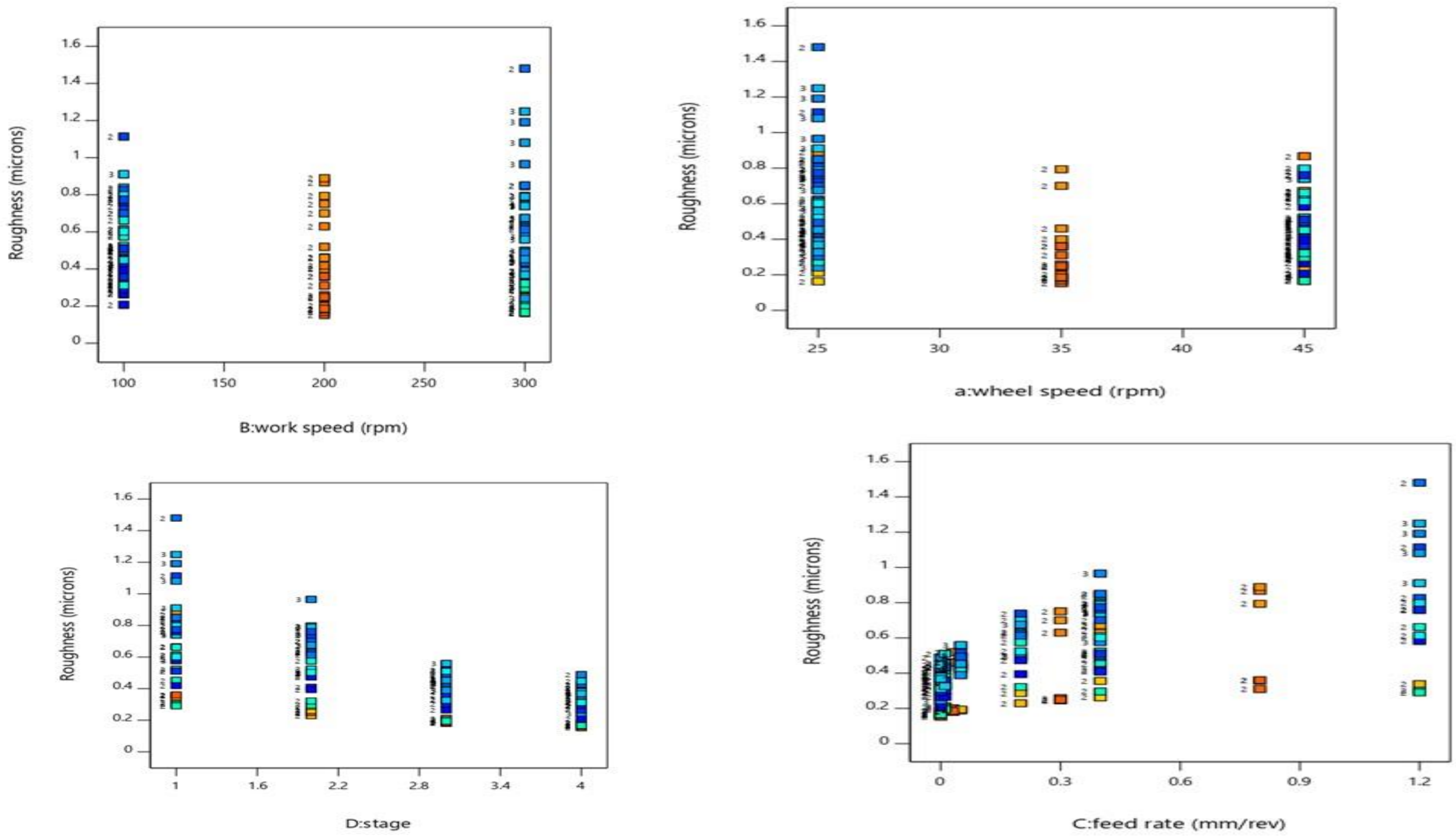
A, C, D significant



Predicted vs. Actual
Predicted
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Actual



Normal Plot of Residuals
Normal % Probability
99
95
90
70
50
20
10
5
1
-4.00
-2.00
0.00
2.00
4.00
Externally Studentized Residuals


Roughness (microns)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
100
150
200
250
300
B:work speed (rpm)
25
30
35
40
45
a:wheel speed (rpm)
1
1.6
2.2
2.8
3.4
4
D:stage
0
0.3
0.6
0.9
1.2
C:feed rate (mm/rev)

- Linear model is significant as all data points lie near the line plotted on predicted vs actual data points.
- Outliers in the normal probability plot are the significant factors
- Optimized factors to minimize roughness:

Feed rate: 0 to 0.3

Wheel speed: 35 rpm

Work speed: 200 rpm

Stage: 4

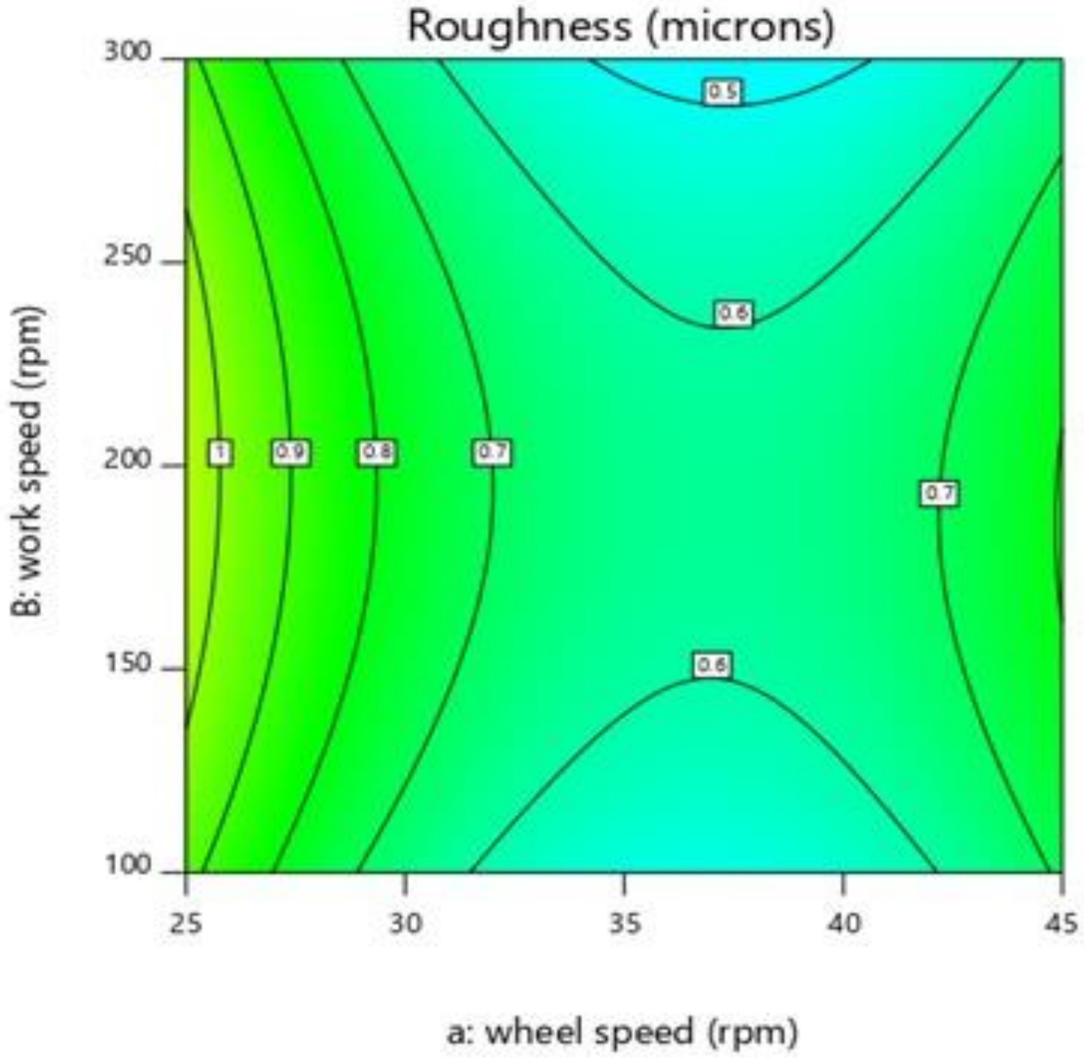
2] Quadratic:

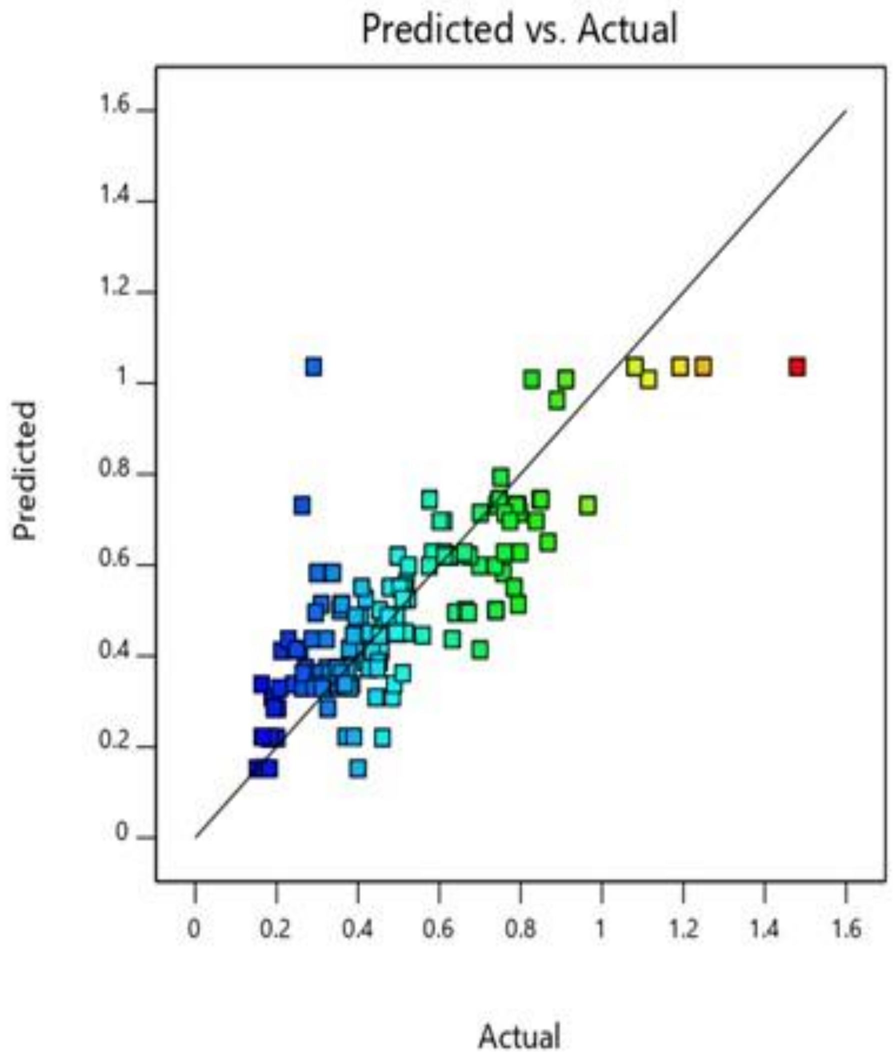
A, C, AB, AC, A^2, B^2 are significant

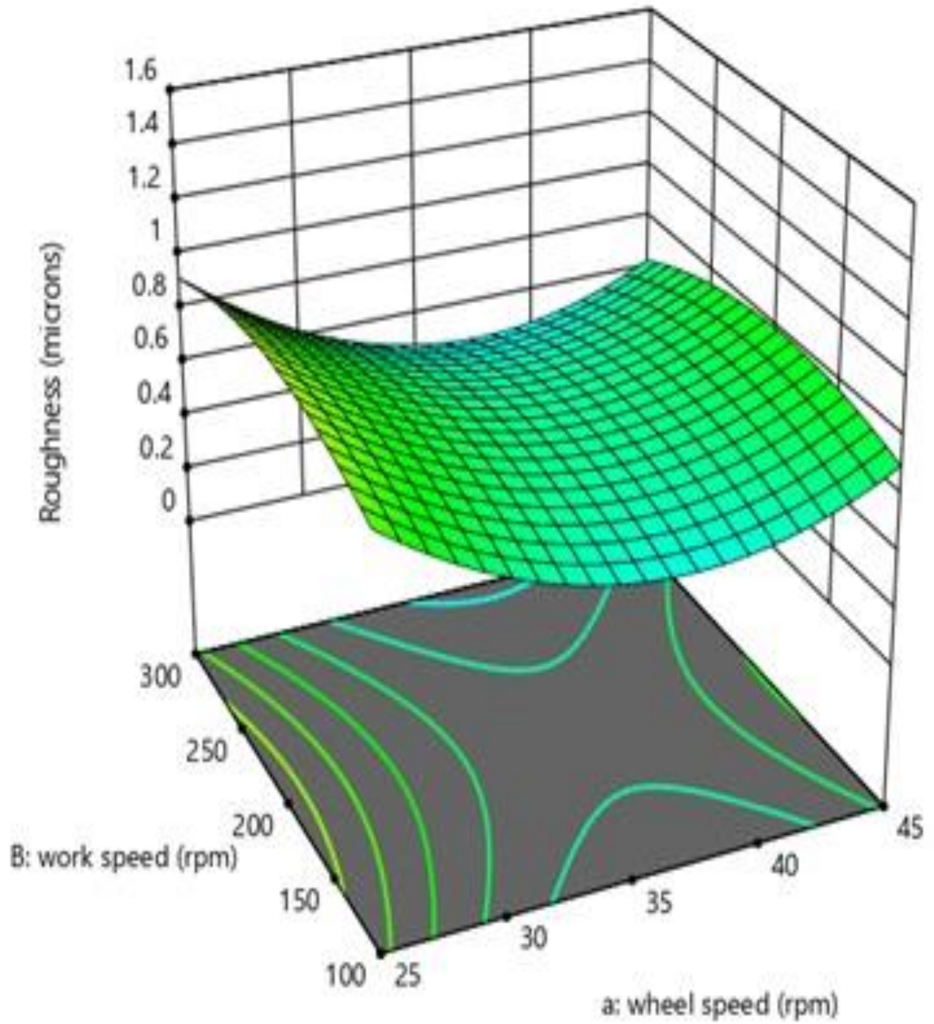
Std. Dev.	0.1392		R^2	0.6941
Mean	0.4984		Adjusted R^2	0.6782
C.V. %	27.94			

Roughness	=
+3.77855	
-0.206723	wheel speed
+0.006080	work speed
-0.089405	feed rate
-0.240050	stage
-0.000018	wheel speed * work speed
-0.013116	wheel speed * feed rate
+0.000379	wheel speed * stage
-0.000124	work speed * feed rate
-0.000154	work speed * stage
+0.432729	feed rate * stage
+0.002926	wheel speed²
-0.000013	work speed²
+0.241883	feed rate²
+0.030276	stage²

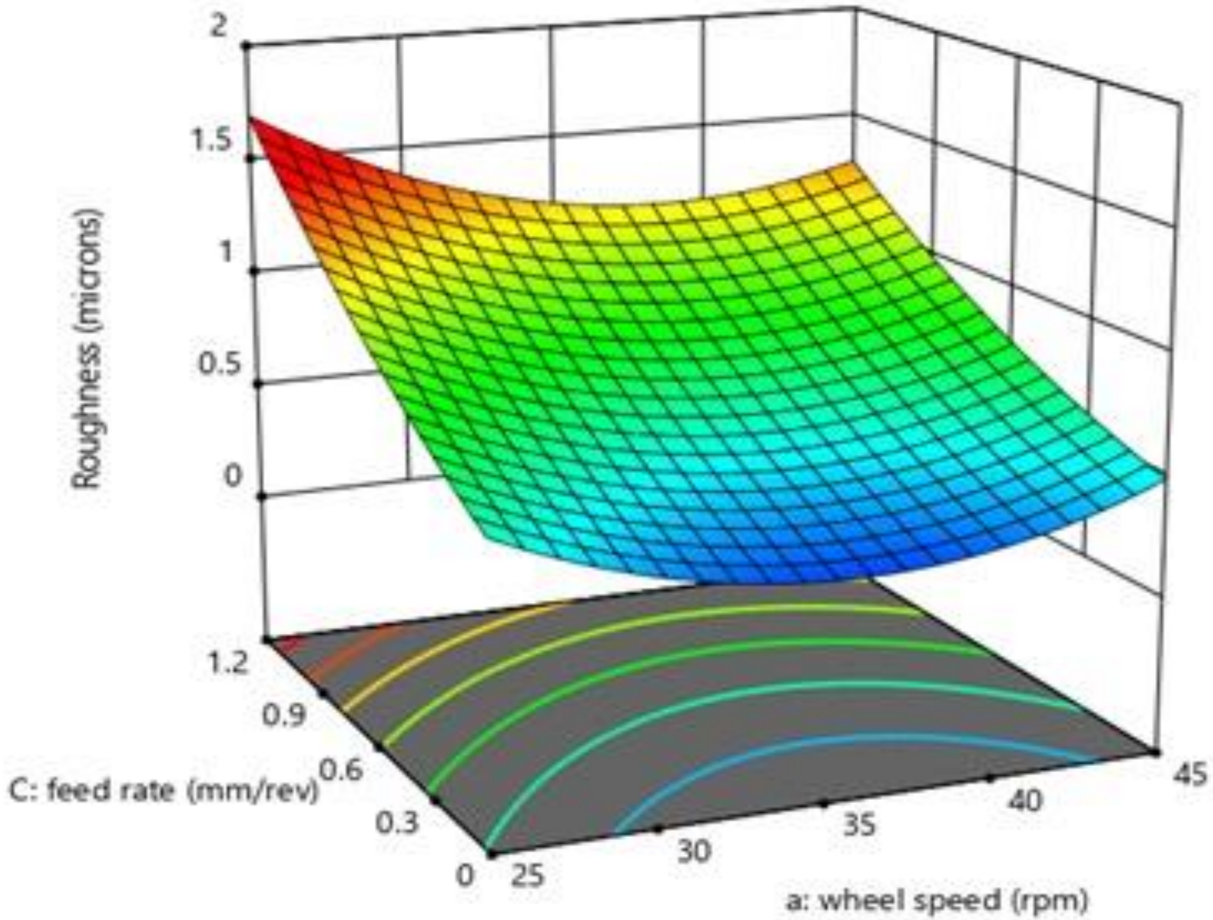
Source	Term df	Error df	F-value	p-value	
Subplot	14	309.00	50.08	< 0.0001	significant
a-wheel speed	1	309.00	76.75	< 0.0001	
B-work speed	1	309.00	1.92	0.1671	
C-feed rate	1	309.00	9.96	0.0018	
D-stage	1	309.00	2.35	0.1264	
aB	1	309.00	4.42	0.0363	
aC	1	309.00	13.52	0.0003	
aD	1	309.00	0.0990	0.7533	
BC	1	309.00	0.1191	0.7303	
BD	1	309.00	1.58	0.2094	
CD	1	309.00	3.37	0.0674	
a²	1	309.00	47.11	< 0.0001	
B²	1	309.00	11.71	0.0007	
C²	1	309.00	0.7085	0.4006	
D²	1	309.00	1.20	0.2745	


Roughness (microns)
300
250
200
150
100
25
30
35
40
45
0.5
0.6
0.6
1
0.9
0.8
0.7
0.7
B: work speed (rpm)
a: wheel speed (rpm)


Predicted vs. Actual
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Predicted
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
Actual



Roughness (microns)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
300
250
200
150
100
B: work speed (rpm)
25
30
35
40
45
a: wheel speed (rpm)

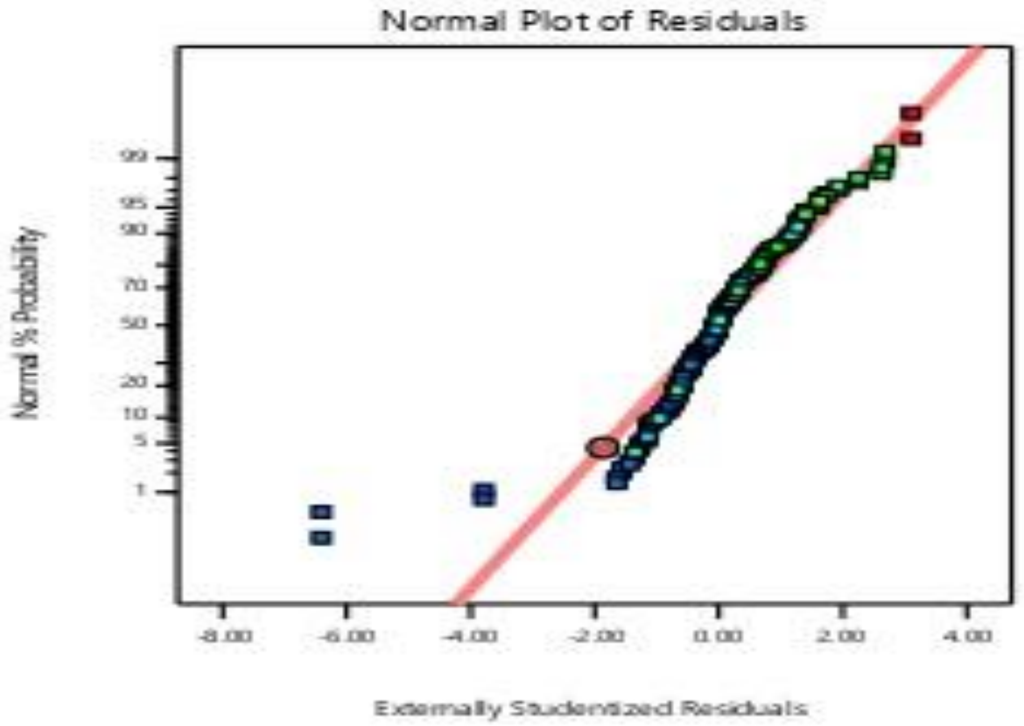


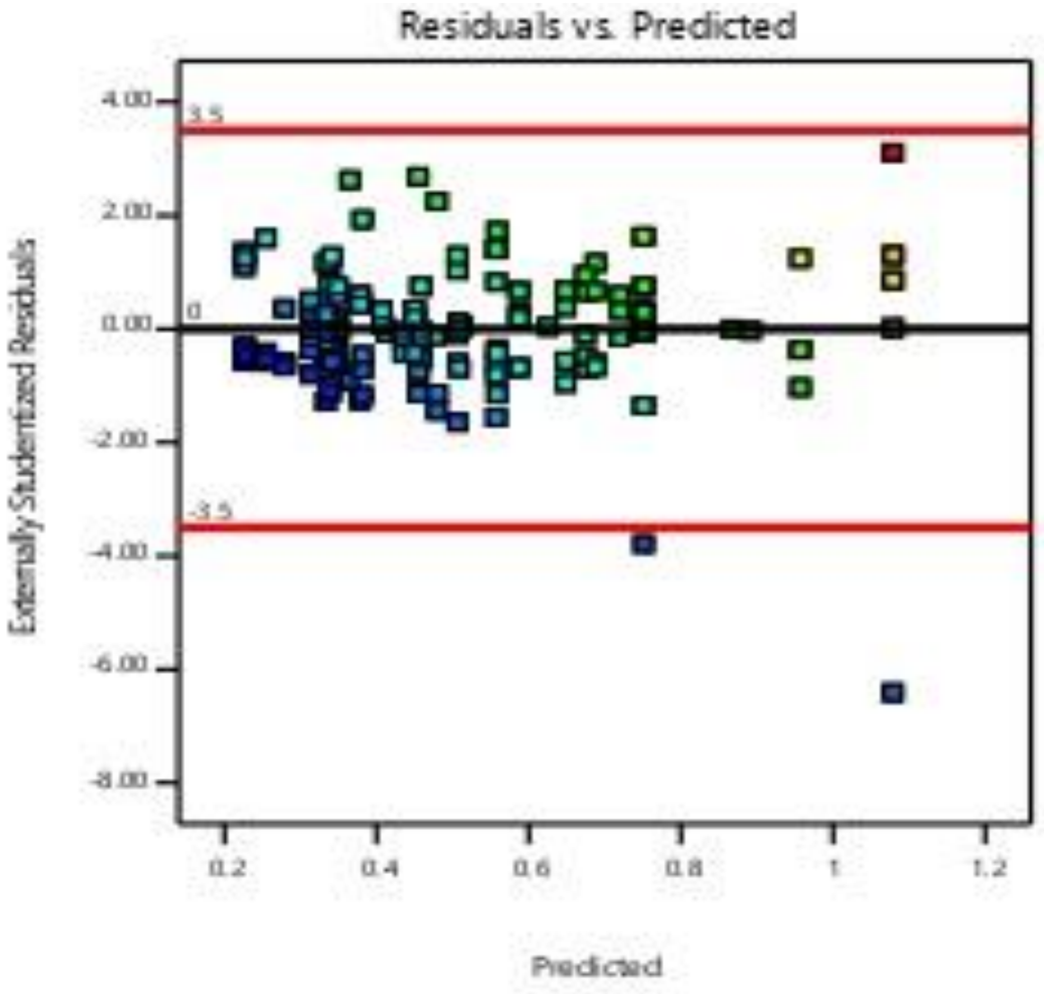
Roughness (microns)
2
1.5
1
0.5
0
1.2
0.9
0.6
0.3
0
C: feed rate (mm/rev)
25
30
35
40
45
a: wheel speed (rpm)

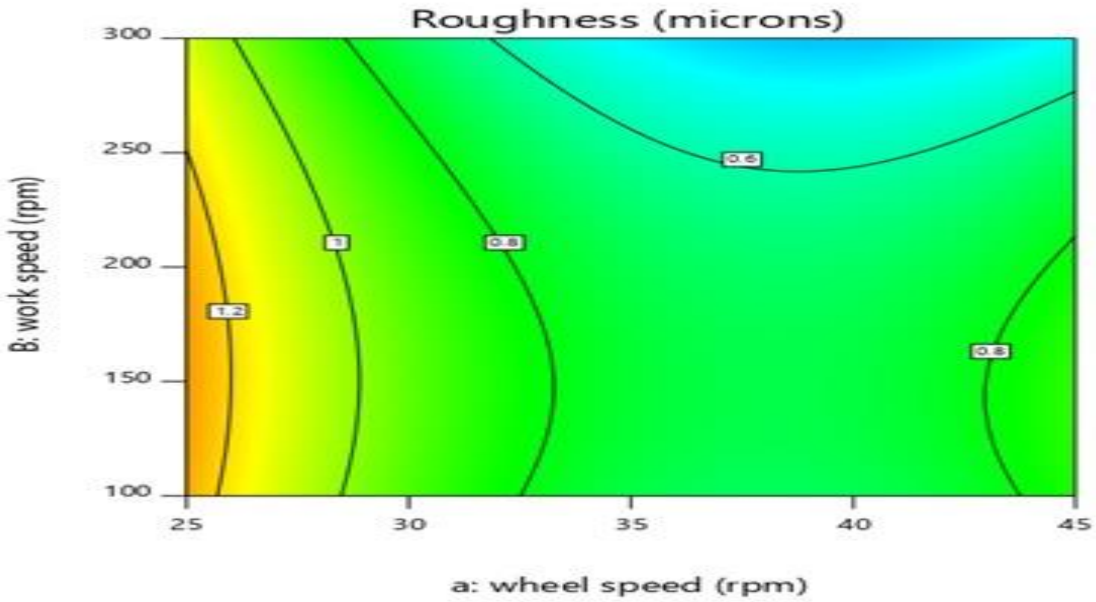
3] Cubic

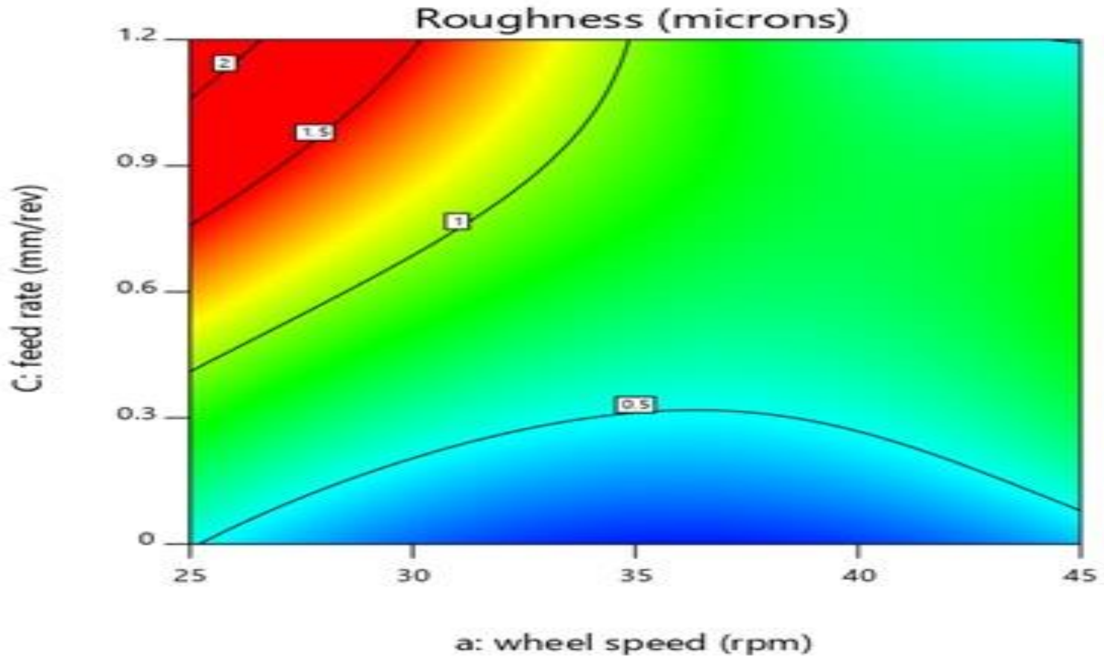
Std. Dev.	0.1363	R^2	0.7210
Mean	0.4984	Adjusted R^2	0.6914
C.V. %	27.36		

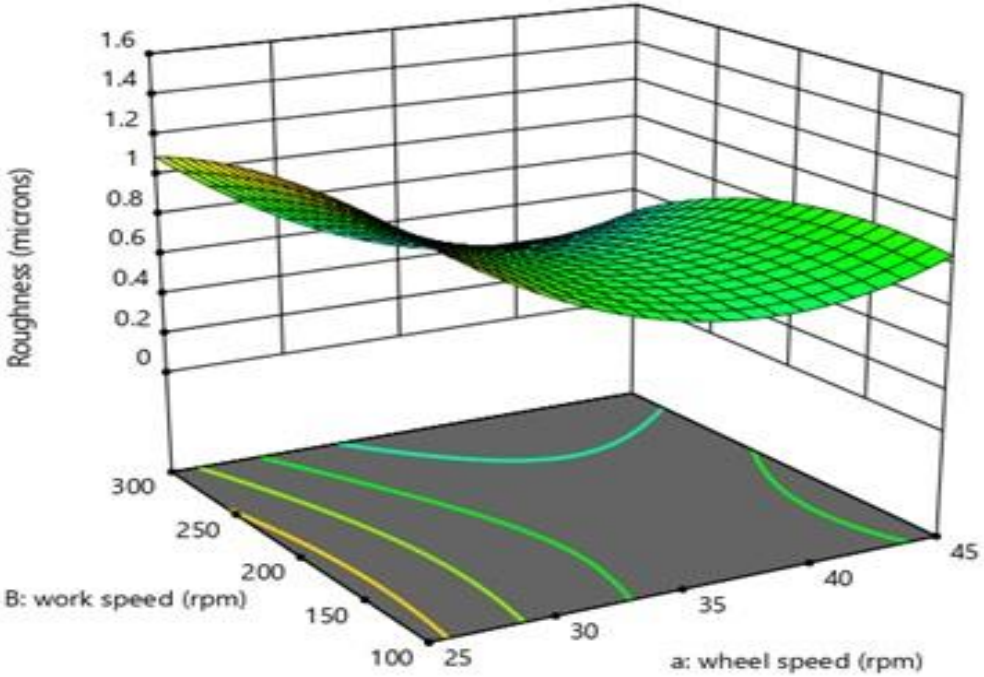
A, AB, A^2, ABC are significant

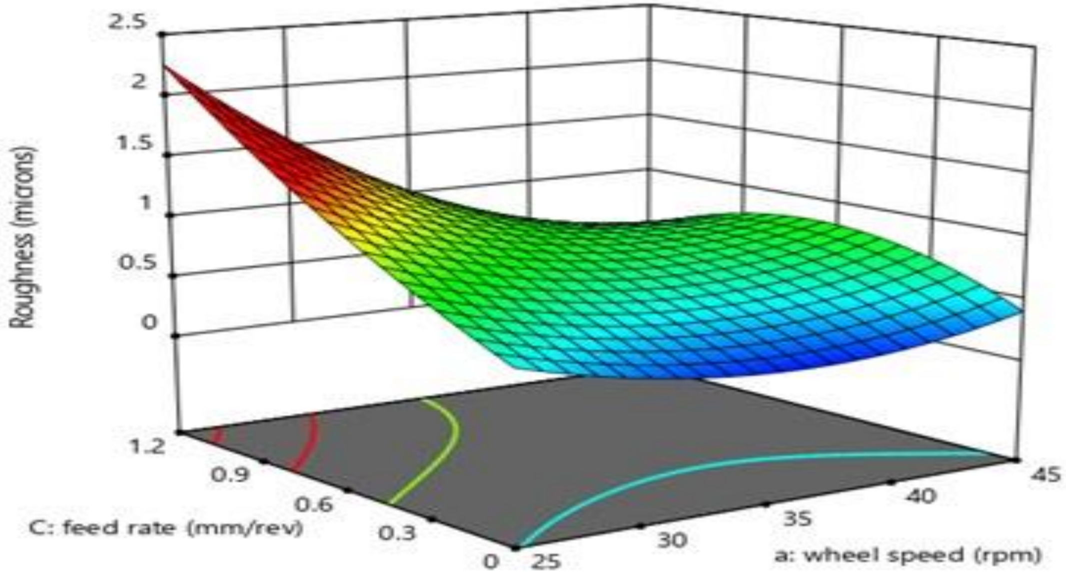





Roughness (microns)
B: work speed (rpm)
a: wheel speed (rpm)
300
250
200
150
100
25
30
35
40
45
1.2
1
0.8
0.6
0.8


Roughness (microns)
C: feed rate (mm/rev)
a: wheel speed (rpm)
1.2
0.9
0.6
0.3
0
25
30
35
40
45
2
1.5
1
0.5

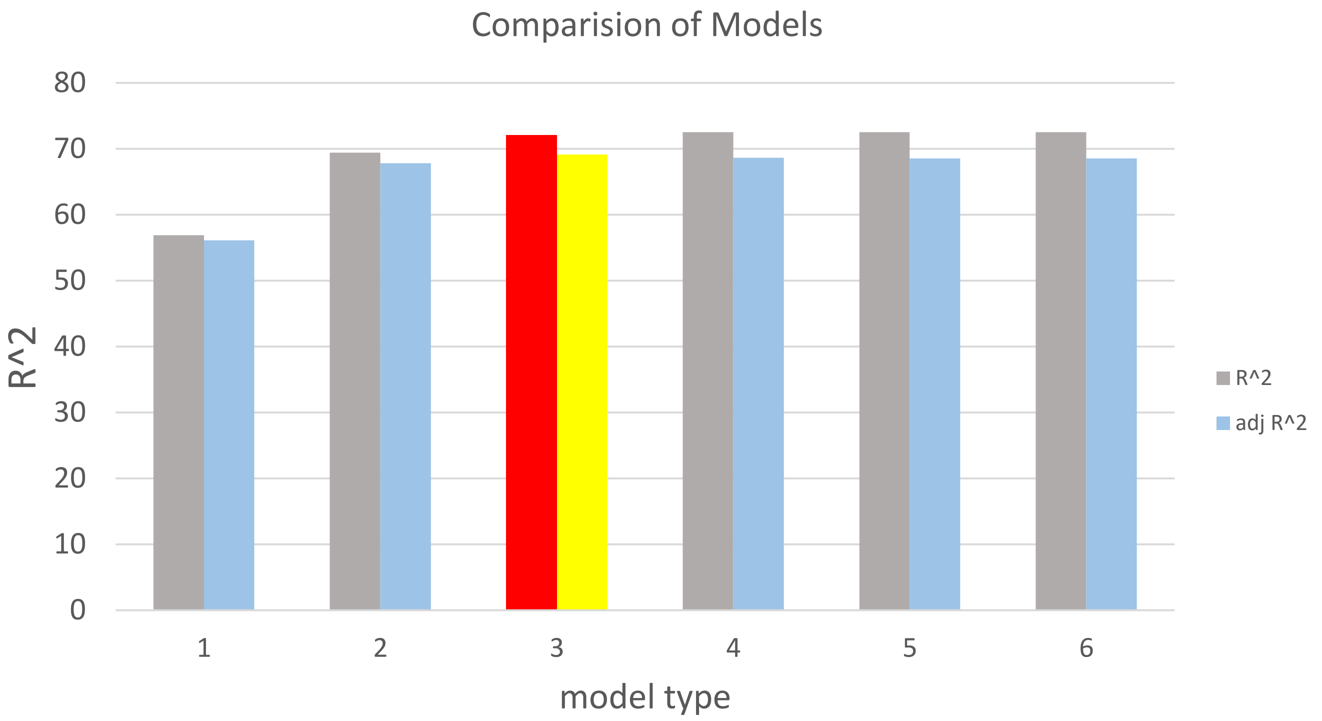

Roughness (microns)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
B: work speed (rpm)
300
250
200
150
100
a: wheel speed (rpm)
25
30
35
40
45

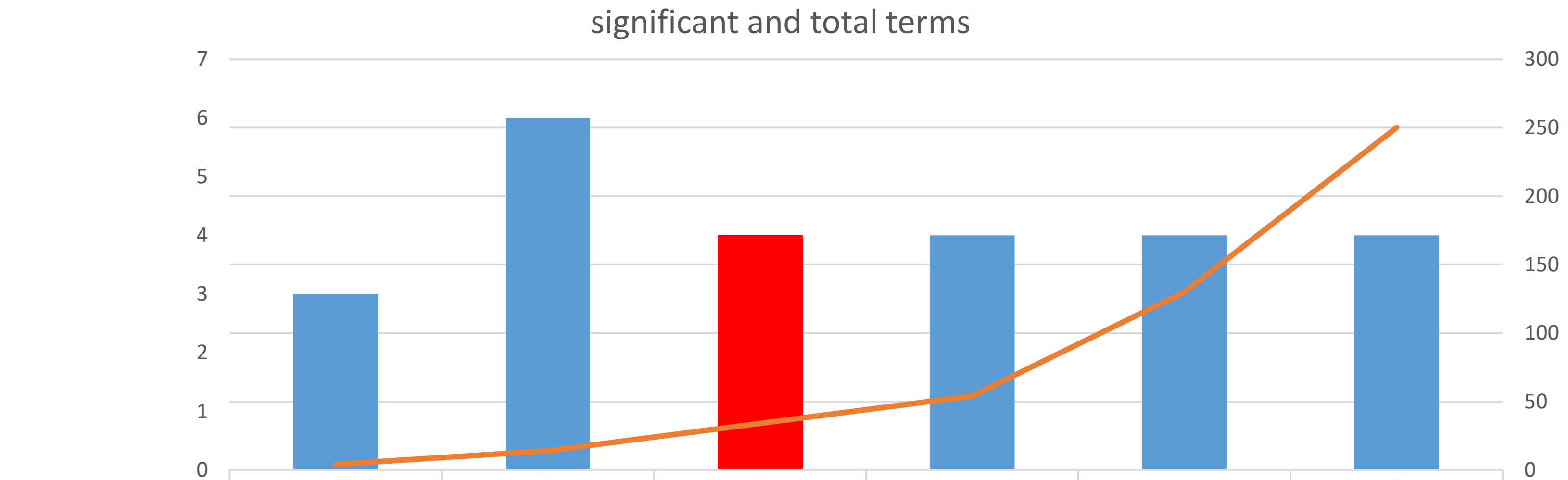

Roughness (microns)
2.5
2
1.5
1
0.5
0
C: feed rate (mm/rev)
1.2
0.9
0.6
0.3
0
a: wheel speed (rpm)
25
30
35
40
45

Comparison of various models

A	Wheel speed				
B	work speed				
C	feed rate				
D	Stage				
MODEL TYPE	**R^2 %**	**adj. R^2 %**	**Significant terms**	**Total terms**	**Significant factors**
Linear	56.91	56.1	3	4	A, C, D
Quadratic	69.41	67.82	6	14	A, C, AB, AC, A^2, B^2
Cubic	72.1	69.14	4	34	A, AB, A^2, ABC
Quartile	72.52	68.64	4	54	A, AB, A^2, ABC
fifth	72.52	68.53	4	130	A, AB, A^2, ABC
sixth	72.52	68.53	4	250	A, AB, A^2, ABC

- Cubic model is the best model as it gives best R^2 value
- As model level increases more number of insignificant terms are added with no improvement in R^2 and significant terms.

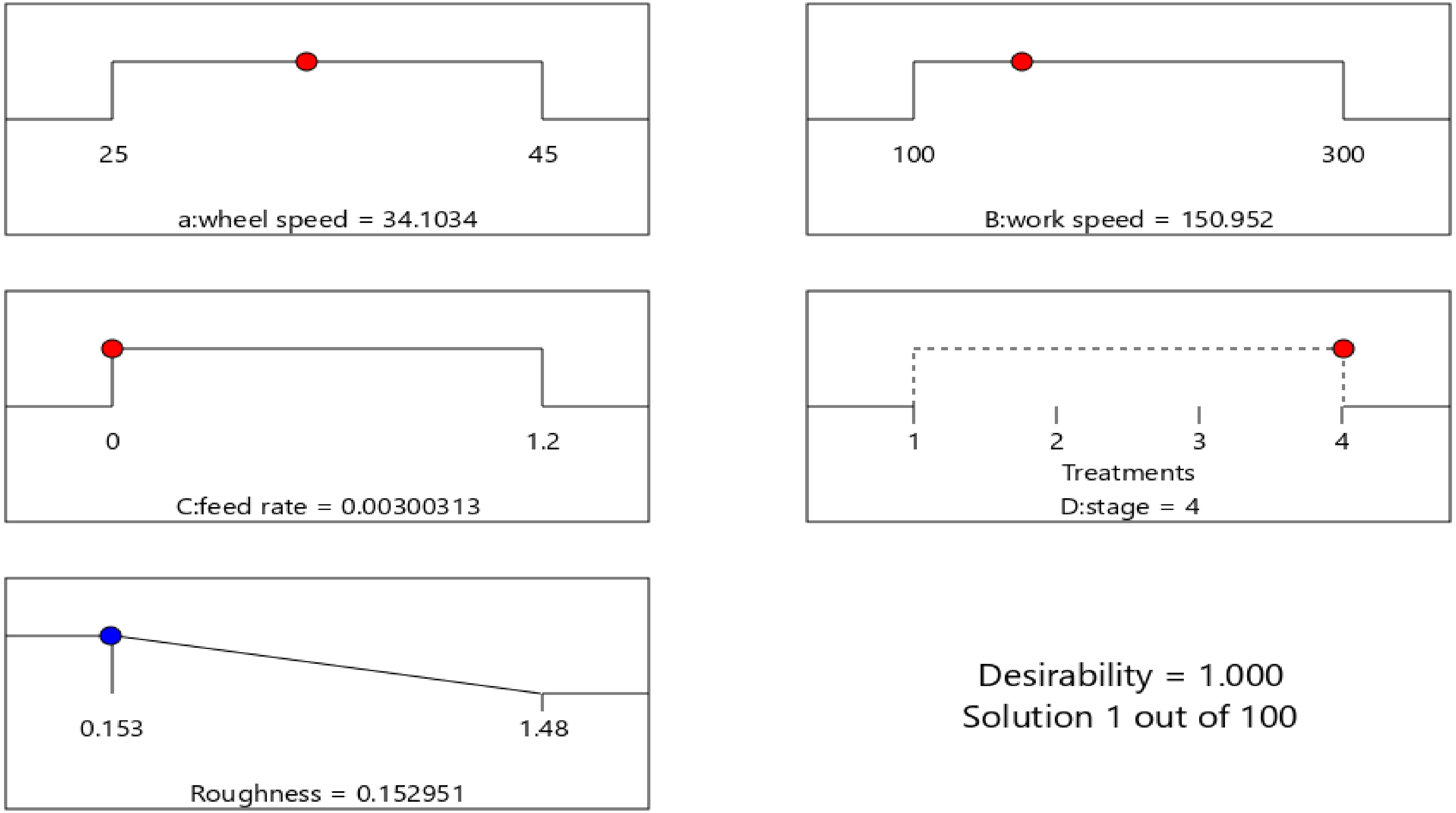


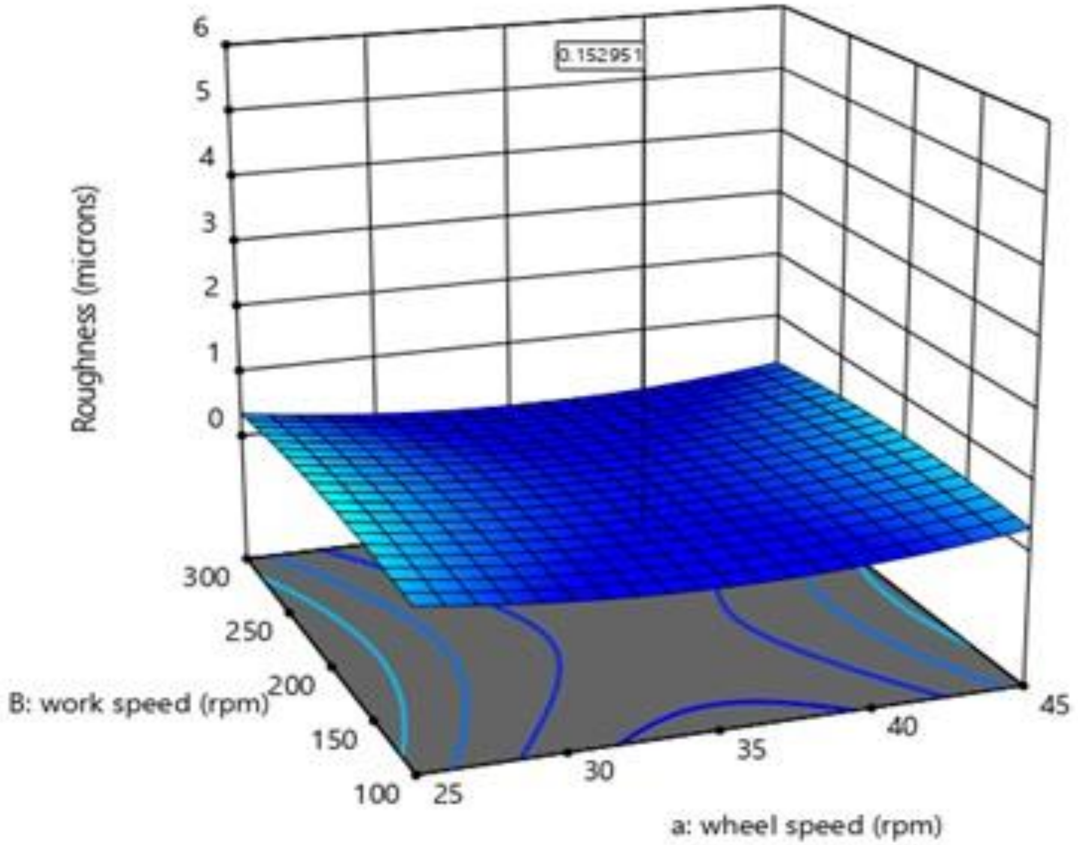


	1	2	3	4	5	6
Significant terms	3	6	4	4	4	4
total terms	4	14	34	54	130	250

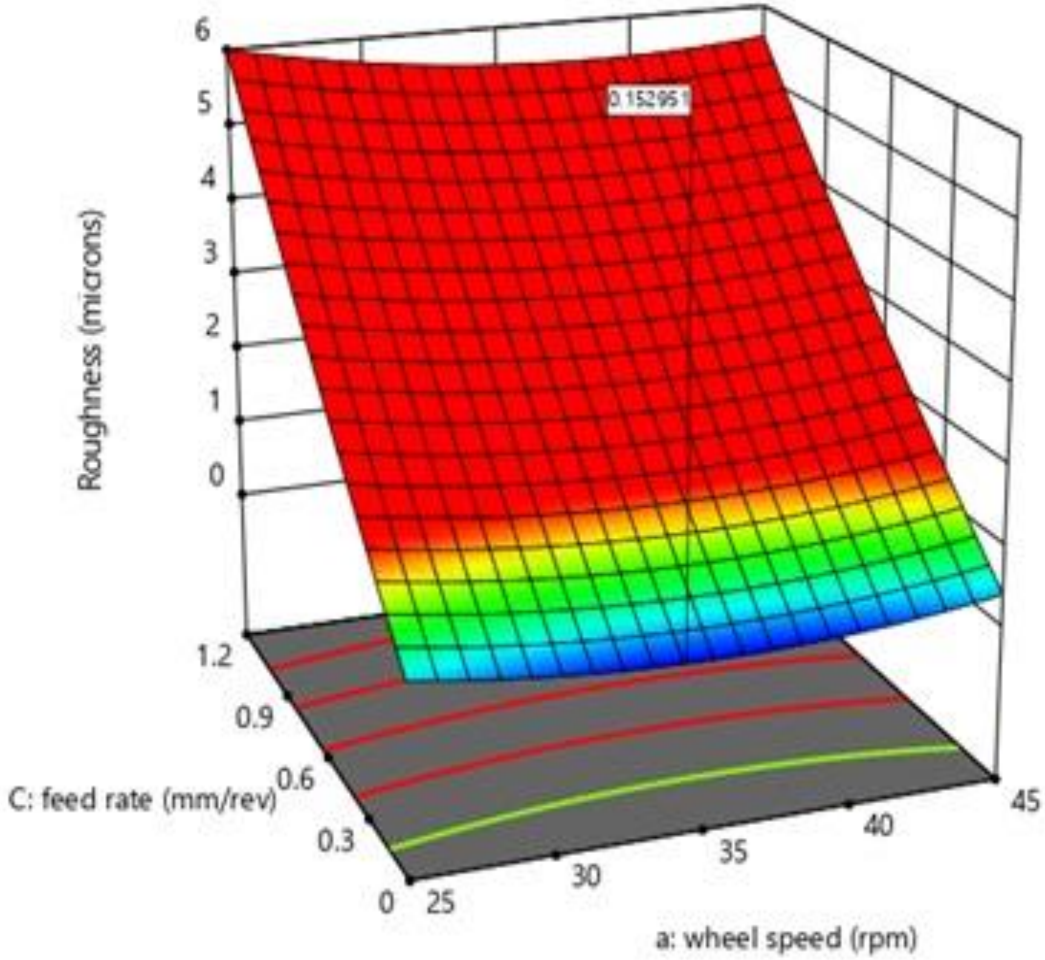
Optimization

wheel speed	work speed	feed rate	stage
40.2699	296.959	1.02722	1.71848



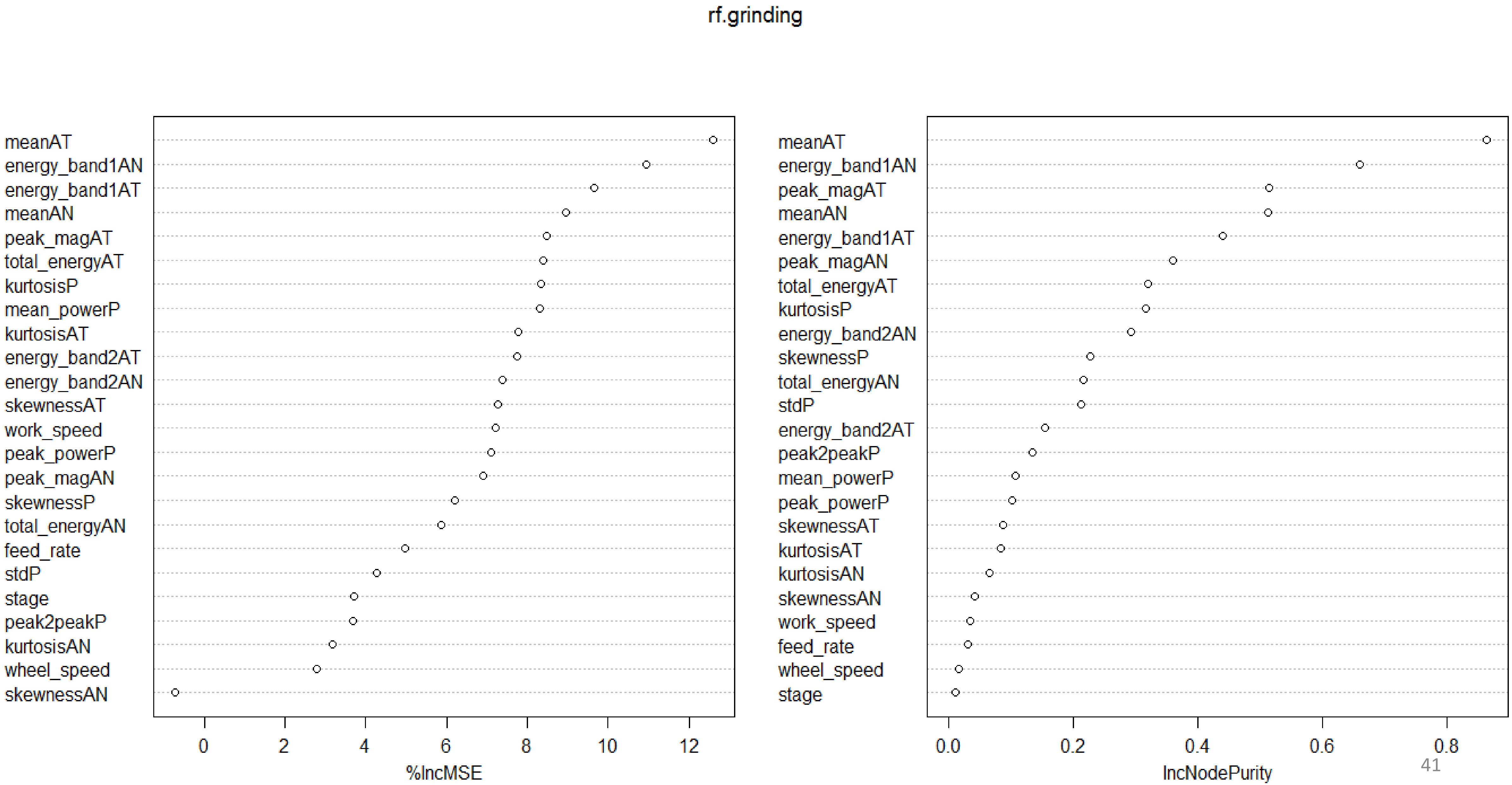


Roughness (microns)
0.152951
6
5
4
3
2
1
0
B: work speed (rpm)
300
250
200
150
100
25
30
35
40
45
a: wheel speed (rpm)



Roughness (microns)
0.152951
6
5
4
3
2
1
0
C: feed rate (mm/rev)
1.2
0.9
0.6
0.3
0
25
30
35
40
45
a: wheel speed (rpm)

Case 2


rf.grinding
meanAT
energy_band1AN
energy_band1AT
meanAN
peak_magAT
total_energyAT
kurtosisP
mean_powerP
kurtosisAT
energy_band2AT
energy_band2AN
skewnessAT
work_speed
peak_powerP
peak_magAN
skewnessP
total_energyAN
feed_rate
stdP
stage
peak2peakP
kurtosisAN
wheel_speed
skewnessAN
0
2
4
6
8
10
12
%IncMSE
meanAT
energy_band1AN
peak_magAT
meanAN
energy_band1AT
peak_magAN
total_energyAT
kurtosisP
energy_band2AN
skewnessP
total_energyAN
stdP
energy_band2AT
peak2peakP
mean_powerP
peak_powerP
skewnessAT
kurtosisAT
kurtosisAN
skewnessAN
work_speed
feed_rate
wheel_speed
stage
0.0
0.2
0.4
0.6
0.8
IncNodePurity

Conclusion

Model	Significant Parameters		
Only process parameter	Wheel speed, (wheel speed)^2		
	Wheel speed* work speed		
	Wheel speed* work speed*feed rate		
Process, Acceleration and Power Parameters	**Process**	**Acceleration**	**Power**
	Work speed	Mean AT	Kurtosis
	Feed rate	Energy band AN	Mean power
	Stage	Mean AN	
	Wheel speed	Energy band AN	

Quotes

- *Every company has big data in its future and every company will eventually be in the data business."-* **Thomas H. Davenport**
- *"Big data will replace the need for 80% of all doctors"-* **Vinod Khosla**
- *"Information is the oil of the 21st century, and analytics is the combustion engine"-* **Peter Sondergaard**, Senior Vice President at Gartner
- "Data are becoming the new raw material of business."- **Craig Mundie**, Senior Advisor to the CEO at Microsoft.

Learnings

- Implementation of algorithms in R and Python
- Modelling in design expert
- Optimization
- Approach to solve the big data problem

Thank you