

Project 2

Samuel Kelly

03 May, 2021

Introduction



Figure 1: Two distinct regions

- ▶ Discovered in 1671 by Giovanni Cassini
- ▶ Has a unique color dichotomy
- ▶ Ridge
- ▶ Explore the relationship between the regions
- ▶ See if there is a SLR

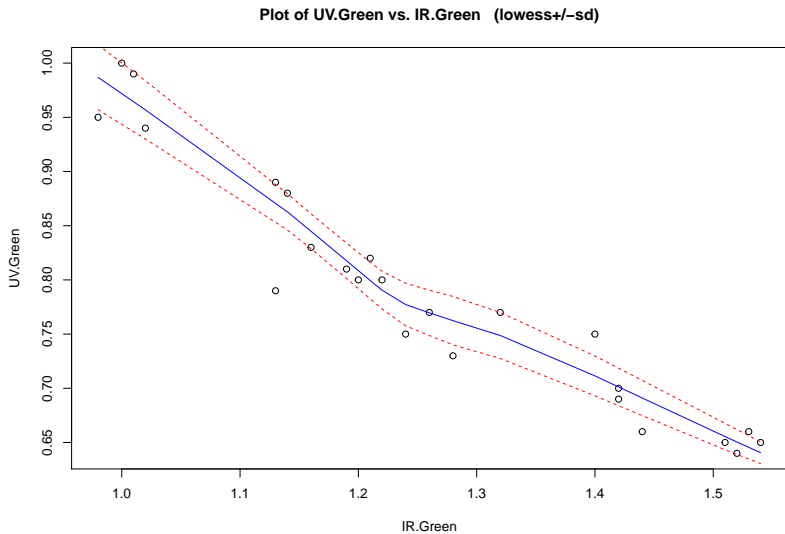
Data

##	Region	IR.Green	UV.Green
## 1	CassiniRegio	1.52	0.64
## 2	CassiniRegio	1.51	0.65
## 3	CassiniRegio	1.54	0.65
## 4	CassiniRegio	1.53	0.66
## 5	TransitionZone	1.44	0.66
## 6	TransitionZone	1.42	0.69

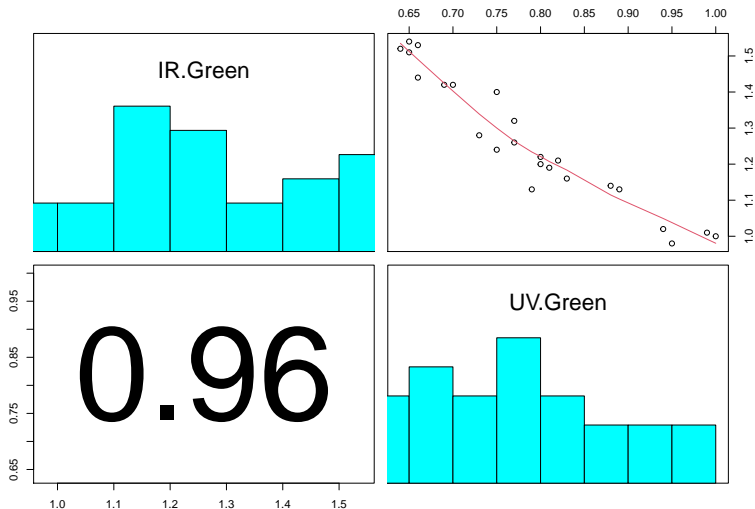
[1] "UV/Green = 0.5951"

[1] "IR/Green = 1.6373"

Initial Graphs



Initial Graphs



```
## [1] "Correlation: -0.9551 Determination: 0.9122"
```

Theory

- ▶ Independent Variable IR/Green
- ▶ Dependent Variable UV/Green
- ▶ Probabilistic vs Deterministic
- ▶ What sort of trend did we observe?
- ▶ $0 \leq R^2 \leq 1$
- ▶ $y = \beta_0 + \beta_1 x + \epsilon$

Assumptions

$$\epsilon_i \sim N(0, \sigma^2)$$

- ▶ 1: Mean of $\epsilon = 0$
- ▶ 2: Variance of ϵ is constant for the model
- ▶ 3: Distributed Normally
- ▶ 4: Independent errors

SLR Linear

```
##
```

```
## Call:
```

```
## lm(formula = UV.Green ~ IR.Green)
```

```
##
```

```
## Residuals:
```

##	Min	1Q	Median	3Q	Max
##	-0.075426	-0.019347	0.002606	0.021480	0.058216

```
##
```

```
## Coefficients:
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	1.52915	0.04947	30.91	< 2e-16 ***
##	IR.Green	-0.58737	0.03885	-15.12	4.2e-13 ***

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
##
```

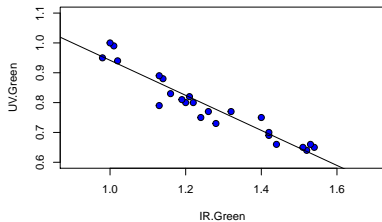
```
## Residual standard error: 0.0331 on 22 degrees of freedom
```

```
## Multiple R-squared:  0.9122, Adjusted R-squared:  0.9082
```

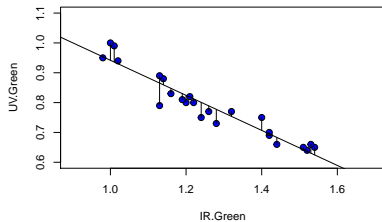
```
## F-statistic: 228.5 on 1 and 22 DF,  p-value: 4.2e-13
```


Verifying Assumptions

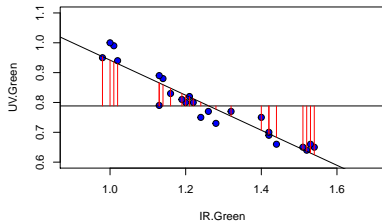
UV.Green vs. IR.Green



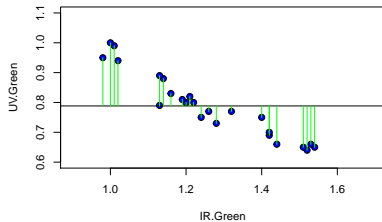
UV.Green vs. IR.Green Residuals



UV.Green vs. IR.Green Model

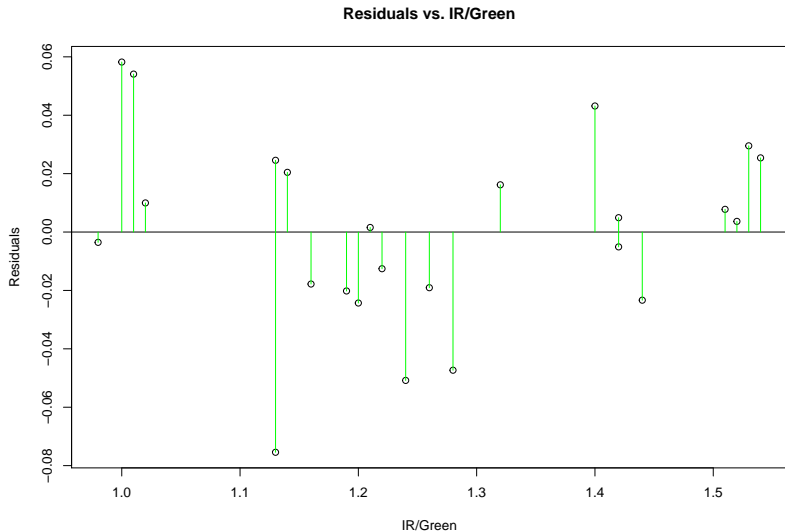


UV.Green vs. IR.Green Total

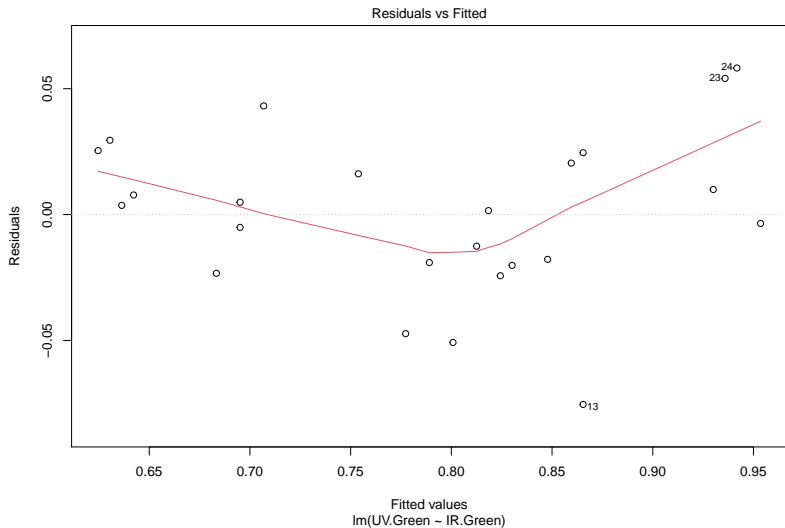


Residuals and IR/Green

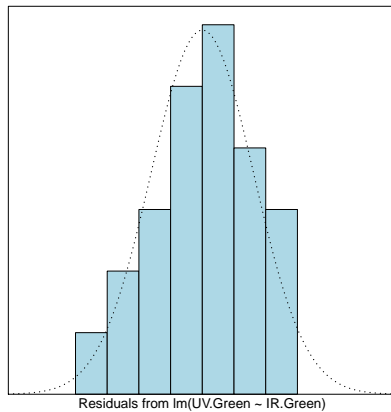
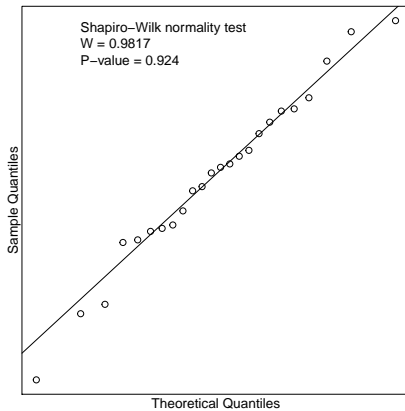
Plot Residuals vs. IR/Green



Residuals and Fitted



Normality



Quadratic

```
##
```

```
## Call:
```

```
## lm(formula = UV.Green ~ IR.Green + I(IR.Green^2), data =
```

```
##
```

```
## Residuals:
```

```
##           Min           1Q       Median           3Q           Max
## -0.068917 -0.010960 -0.000647  0.020407  0.051199
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.4822     0.3273   7.583 1.92e-07 ***
## IR.Green         -2.1174     0.5221  -4.056 0.000569 ***
## I(IR.Green^2)     0.6025     0.2052   2.937 0.007883 **
```

```
## ---
```

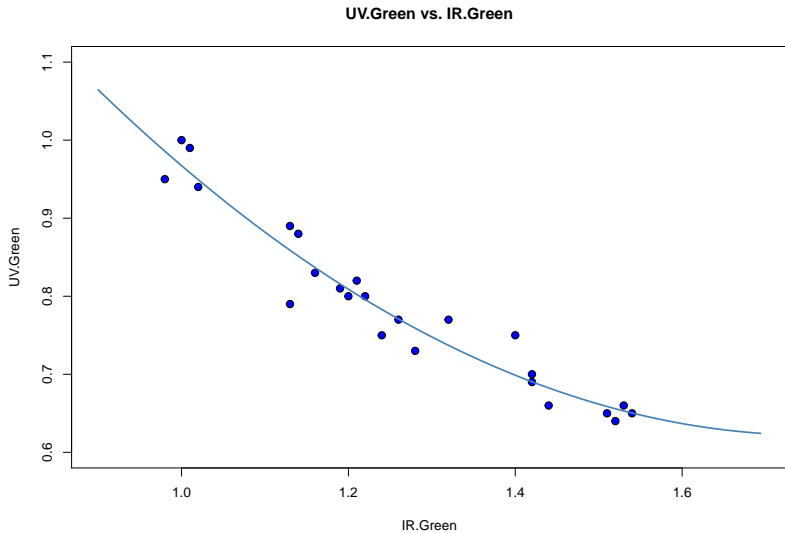
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
##
```

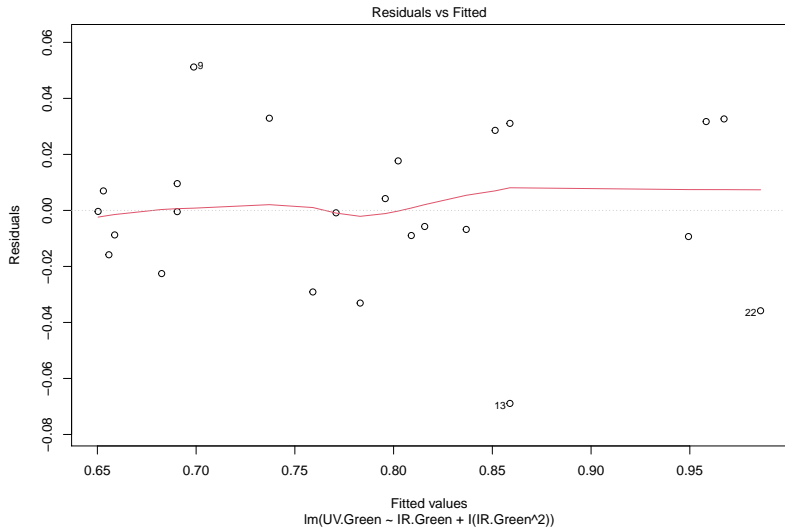
```
## Residual standard error: 0.02853 on 21 degrees of freedom
```

```
## Multiple R-squared:  0.9377, Adjusted R-squared:  0.9318
```

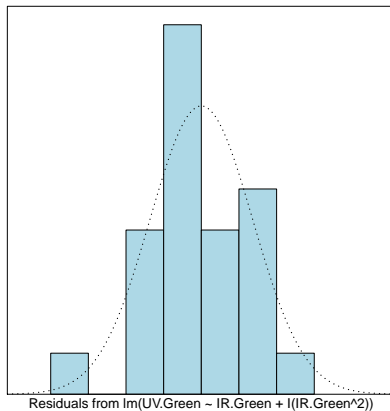
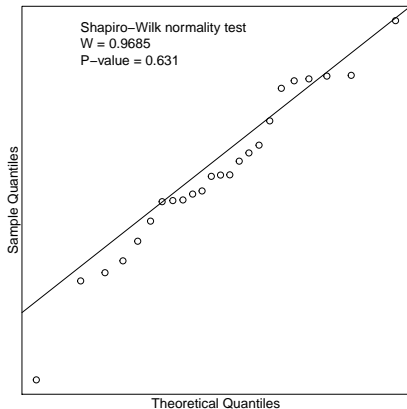
Plot Quadratic



Residuals and Fitted



Normality



Predictors

```
##           1           2           3           4           5
## 0.3544136 0.6363513 0.8184361 0.9417838 1.2354689
```

```
## [1] "Mean is: 0.797290728845569"
```

```
##           1           2           3           4           5
## 0.6575019 0.6558367 0.8023171 0.9673404 1.5741458
```

```
## [1] "Mean is: 0.931428405630695"
```

Estimate 2, 1.52, 1.21, 1, .5.

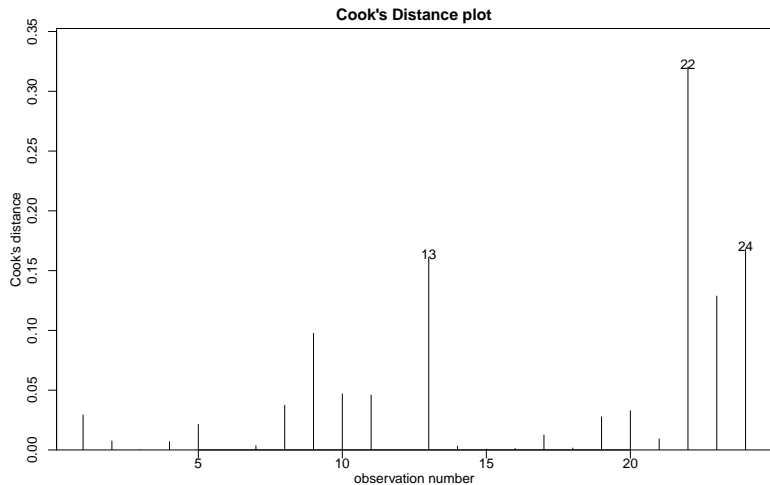
Compare R squared values

- ▶ Linear Model:
- ▶ Multiple R^2 : 0.9122
- ▶ Adjusted R^2 : 0.9082
- ▶ Quadratic Model:
- ▶ Multiple R^2 : 0.9377
- ▶ Adjusted R^2 : 0.9318

Anova

```
## Analysis of Variance Table
##
## Model 1: UV.Green ~ IR.Green
## Model 2: UV.Green ~ IR.Green + I(IR.Green^2)
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1      22 0.024108
## 2      21 0.017090  1 0.0070178 8.6233 0.007883 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Cook's Distance



Remove 22

```
##
## Call:
## lm(formula = UV.Green ~ IR.Green + I(IR.Green^2), data = 
##     ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.072116 -0.013636  0.001271  0.018835  0.053708
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.7336     0.3560   7.678 2.19e-07 ***
## IR.Green       -2.4972     0.5618  -4.445 0.000249 ***
## I(IR.Green^2)   0.7443     0.2187   3.403 0.002821 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 0.02762 on 20 degrees of freedom
```

Compare R squared

- ▶ With 22
- ▶ Multiple R^2 : 0.9377
- ▶ Adjusted R^2 : 0.9318
- ▶ Without 22
- ▶ Multiple R^2 : 0.9383
- ▶ Adjusted R^2 : 0.9321

Research Question

- ▶ Which model best fitted our data?
- ▶ Was there a relationship between UV/Green and IR/Green?
- ▶ As one variable increases, what does the other do?

Suggestions

- ▶ Distance between measurements
- ▶ More measurements