UNIVERSITY OF CAPE TOWN
2019

Course: STA2007F

Student details: Kyle Du Plessis (DPLKYL002)
Malikah Hardenberg (HRDMAL001)
Tanweer Beckett (BCKTAN001)

Project 1: Regression

Land degradation: Factors driving vegetation productivity in South Africa

Due Date: 26th March 2019

Abstract:

The main objective of this report is to determine the key factors driving vegetation productivity in South Africa. Gross primary productivity (GPP) is thought to be influenced by a number of factors. We aim to find which of these factors has the greatest effect on the gross primary productivity and how exactly these factors and the GPP are related. To understand the main factors driving vegetation productivity, a seven-step hypothesis-driven statistical approach for statistical modelling had been followed: models to be fitted and motivation, exploratory data analysis, adaptation of models using exploratory data analysis, fitting of the adapted models, model checking, presentation of results, and interpretation of the final model. Our results suggest that the factors that had the most influence on the gross primary productivity (GPP) were livestock units and seasonality as well as an interaction between soil nutrients and rainfall. We found that this interaction was in fact positive and not negative as predicted in our hypothesis. Using Akaike's Information Criterion (AIC), we found that the model with the interaction term was the best in our set and had an adjusted R^2 of 90.55%.

Introduction:

Gross primary productivity (GPP) is the rate at which ecosystems producers store and capture provided amount of energy as biomass in a given length of time and is measured in grams of carbon / m^2/ year. It is thought to be influenced by a number of factors. GPP provides highly synthesized, quantitative information for sustained resource management. It is an important component of the biospheres carbon cycle, which is important aspect of climate change studies. This indicates its importance to understand what factors, but also how these factors influence GPP, as well as monitor GPP levels. We should therefore question, how do environmental and climatic factors such as rainfall, temperature, livestock units, land size, soil nutrients, seasonality and land use affect gross primary productivity? Also, which are the factors that has the greatest effect on the gross primary productivity and how exactly are these factors and the GPP related?

We hypothesized: (1) Factors predicted to influence the gross primary productivity (GPP) were soil nutrients, rainfall, livestock units and seasonality. (2) The factors predicted to influence the gross primary productivity (GPP) were livestock units and seasonality as well as a negative interaction between soil nutrients and rainfall.

We gathered information from previous studies to justify these hypotheses and choice of covariates. Davis *et al*. (2003) found that the best model depended on temperature and precipitation in spring and summer. It was also noted that models depending on the bioclimatic variables were less well supported than their temperature and precipitation-based explanatory variables. Furthermore, EL-Vilaly *et al*. (2018) found that the best model depended on elevation, aspect, slope, seasonal precipitation, seasonal temperature and vegetation communities' explanatory variables. Previous studies done by Nemani *et al.* (2003), also demonstrated that global change in climate and well as land use had resulted in 6% increase in net primary productivity globally. Matthews (2006) found that the livestock sector generates 18% more greenhouse gas emissions than the transport sector and is a substantial source of land and water degradation.

In the second model we predicted a negative interaction between soil nutrients and rainfall – the effect of soil nutrients decreases with increasing values of rainfall. Kline (2012) found that areas that get more than 25 inches of rain per year suffer loss of soil nutrition.

We will be following a hypothesis-driven statistical approach for selecting the best model.

Data and Methods:

To understand the main factors driving vegetation productivity, a seven-step hypothesis-driven statistical approach for statistical modelling had been followed: models to be fitted and motivation, exploratory data analysis, adaptation of models using exploratory data analysis, fitting of the adapted models, model checking, presentation of results, and interpretation of the final model.
The statistical software package was used was R. We will assess the goodness-of-fit of our models to the data by looking at the adjusted R^2 of each model. The assumptions of the models are linearity, constant error variance, independence as well as normality, and each of these will be assessed using model checking for each model. We used Akaike's Information Criterion (AIC) to select the best model.

As mentioned previously, we came up with the following two hypotheses:

1. The factors predicted to influence the gross primary productivity (GPP) were soil nutrients, rainfall, livestock units and seasonality.
2. The factors predicted to influence the gross primary productivity (GPP) were livestock units and seasonality as well as a negative interaction between soil nutrients and rainfall.

We then formulated a model that can structurally represent each of these hypotheses. The models we have fitted are the following, where 'H_i' represents the response variable we are interested in – i.e. the measured the gross primary productivity (GPP) $[\text{grams of carbon} / m^2 / \text{year}]$ of the amount of photosynthesis taking place i. The explanatory variables include:

Variable	Description	Codes/Values
soil.nutrients (continuous)	soil nutrient richness index	10-95
rainfall (continuous)	mean annual rainfall	53-918 (mm)
livestock.units (continuous)	livestock unit	3-47 (per ha)
seasonality (categorial)	predominant type of rainfall	Summer rain / Winter rain

The β's are parameters to be estimated and the ϵ_i are normally distributed errors following $N(\mathbf{0},\ \boldsymbol{\sigma}^2)$:

1. The linear relationship had been described using a linear regression model, also adding a quadratic term to the linear regression model to allow the fitted relationship to assume an optimum:

$$H_i = \beta_0 + \beta_1 \times \text{soil.nutrients}_i + \beta_2 \times \text{rainfall}_i + \beta_3 \times \text{livestock.units}_i + \beta_4 \times (\text{livestock.units})^2 + \beta_{\text{seasonalityWinter rain}} + \epsilon_i$$

This model has seven parameters.

2. The linear relationship had been described using a linear regression model, also adding a quadratic term to the linear regression model to allow the fitted relationship to assume an optimum as well as an interaction term:

$$H_i = \beta_0 + \beta_1 \times \text{soil.nutrients}_i + \beta_2 \times \text{rainfall}_i + \beta_3 \times \text{livestock.units}_i + \beta_4 \times (\text{livestock.units})^2 + \beta_{\text{seasonalityWinter rain}} + \beta_{1:2}(\text{soil.nutrients}_i \times \text{rainfall}_i) + \epsilon_i$$

This model has eight parameters.

In R, we fitted these models using the following code:

```
m1 <- lm(GPP ~ soil.nutrients + rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)

m2 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)
```

Both models are linear models and we can fit them using function lm().

Results:

Exploratory data analysis:

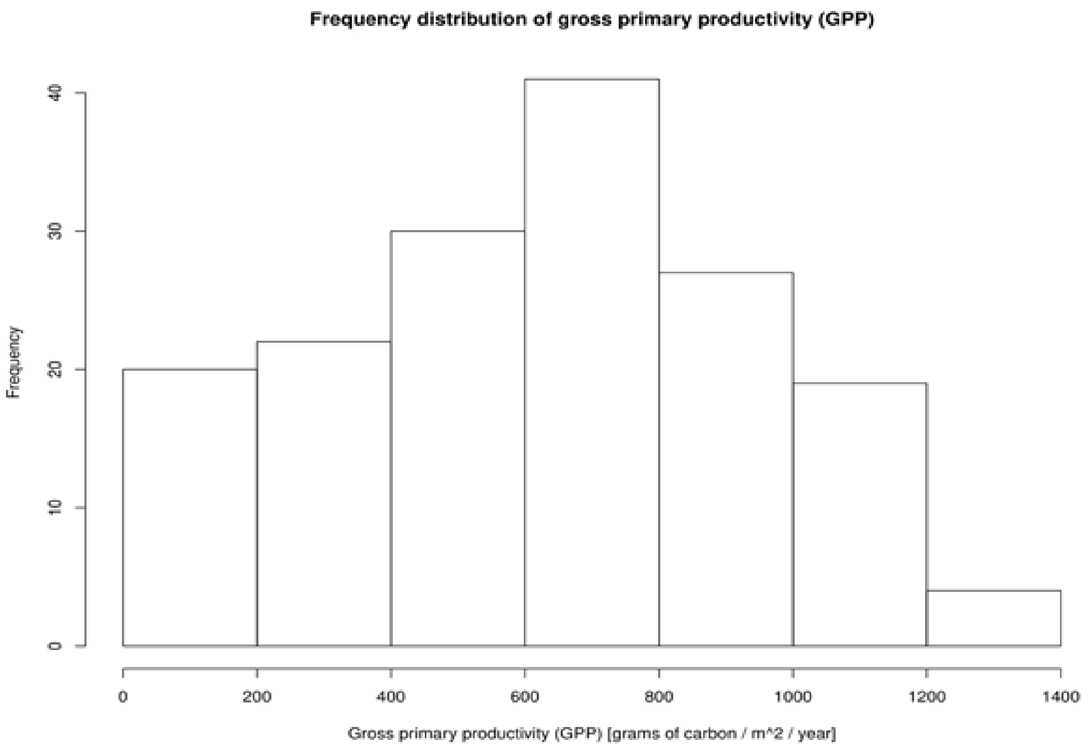

Figure 1: Histogram illustrating the distribution of GPP

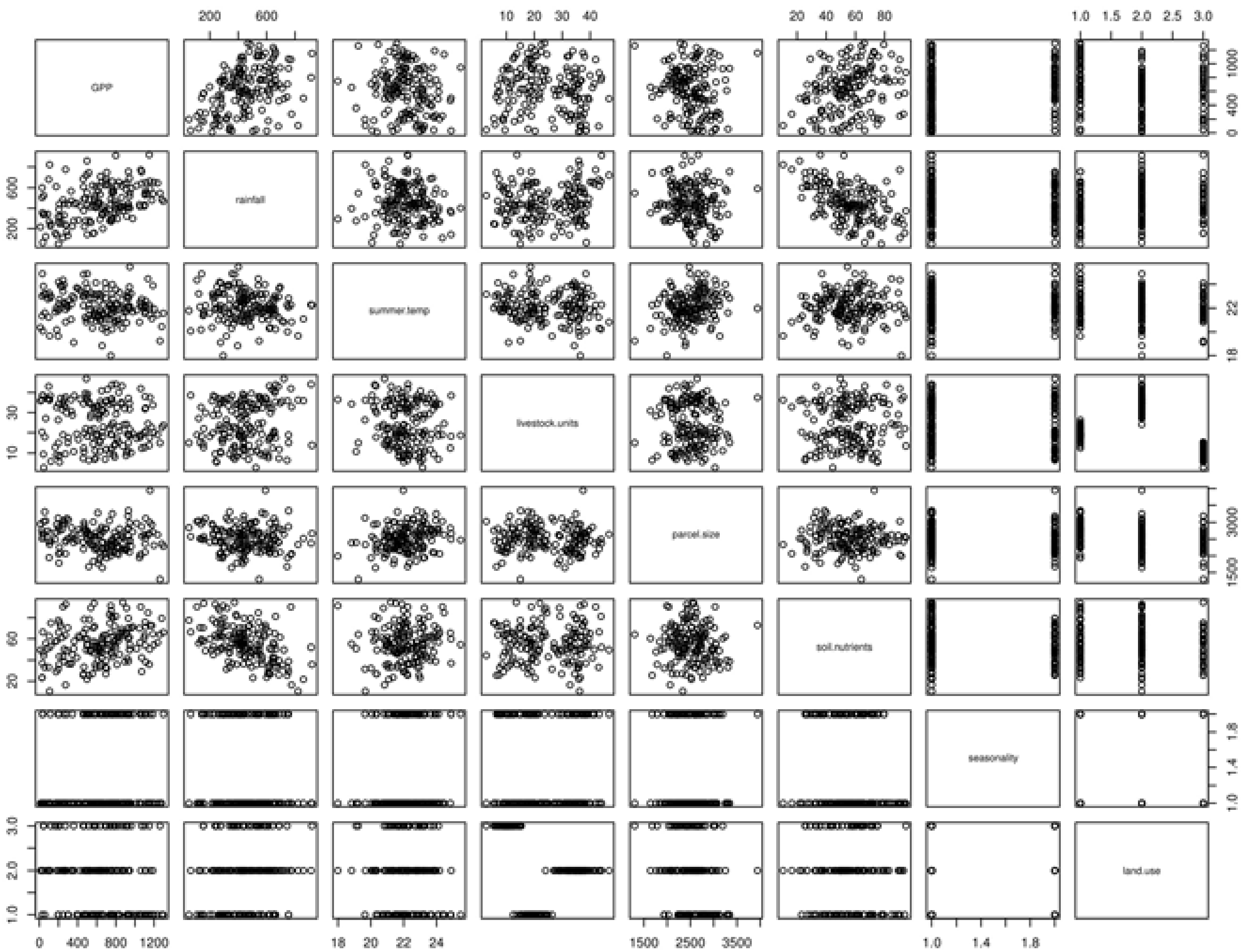

Figure 2: Pairs plot illustrating the relationships between GPP and the covariates

<u>Fitting the models:</u>

Model 1:

$$H_i = \beta_0 + \beta_1 \times \text{soil.nutrients}_i + \beta_2 \times \text{rainfall}_i + \beta_3 \times \text{livestock.units}_i + \beta_4 \times (\text{livestock.units})^2 + \beta_{winter} + \epsilon_i$$

Motivation for fitting this model:

- summer.temp and parcel.size not included since there's no clear pattern in the data and this cannot be captured appropriately by neither a linear nor non-linear model
- land.use not included since it is correlated with livestock.units
- only livestock.units is included since it is more directly related and meaningful for modelling the response

Model 2: $H_i = \beta_0 + \beta_1 \times \text{soil.nutrients}_i + \beta_2 \times \text{rainfall}_i + \beta_3 \times \text{livestock.units}_i + \beta_4 \times (\text{livestock.units})^2 + \beta_{winter} + \beta_{1:2}(\text{soil.nutrients}_i \times \text{rainfall}_i) + \epsilon_i$

Motivation for fitting this model:

- summer.temp and parcel.size not included since there's no clear pattern in the data and this cannot be captured appropriately by neither a linear nor non-linear model
- land.use not included since it is correlated with livestock.units
- only livestock.units is included since it is more directly related and meaningful for modelling the response

In R, we fitted these models using the following code:

```
m1 <- lm(GPP ~ soil.nutrients + rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)

m2 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)
```

<u>Multicollinearity:</u>

Checking for multicollinearity using test models t1 and t2:

Test model t1:

t1 <- lm(gpp$GPP ~ rainfall + summer.temp + livestock.units + parcel.size + soil.nutrients + seasonality + land.use, data=gpp)

vif(t1)

Model output:

livestock.units	11.090290
land.use	11.417444

<u>Table 1: Model output for test model t1</u>

Test model t2:

t2 <- lm(GPP ~ soil.nutrients + rainfall + seasonality + livestock.units, data=gpp)

vif(t2)

Model output:

Soil.nutrients	**rainfall**	**seasonality**	**livestock.units**
1.411833	1.411885	1.047019	1.041133

<u>Table 2: Model output for test model t2</u>

Checking for multicollinearity using models m1 and m2:

m1 <- lm(GPP ~ soil.nutrients + rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)

vif(m1)

Model output:

Soil.nutrients	**rainfall**	**livestock.units**	**I(livestock.units^2)**	**seasonality**
1.451314	1.557309	29.685631	30.439368	1.075000

<u>Table 3: VIF model output for model m1</u>

m2 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)

vif(m2)

Model output:

soil.nutrients	**rainfall**	**livestock.units**	**I(livestock.units^2)**	**seasonality**	**soil.nutrients:rainfall**
7.936598	10.006172	29.693300	30.466244	1.086224	8.580993

<u>Table 4: VIF model output for model m2</u>

```
termite.lm4 <- lm(density ~ height*map + soil, data=termites)

vif(termite.lm4)
```

Model output:

Variable	**GVIF**	**Df**	**GVIF^(1/(2*Df))**
Height	486.228578	1	22.050591
Map	42.165773	1	6.493518
Soil	3.494321	4	1.169285
height:map	402.633740	1	20.065735

Table 5: VIF output for model termite.lm4

Model checking for m1:

Linearity:

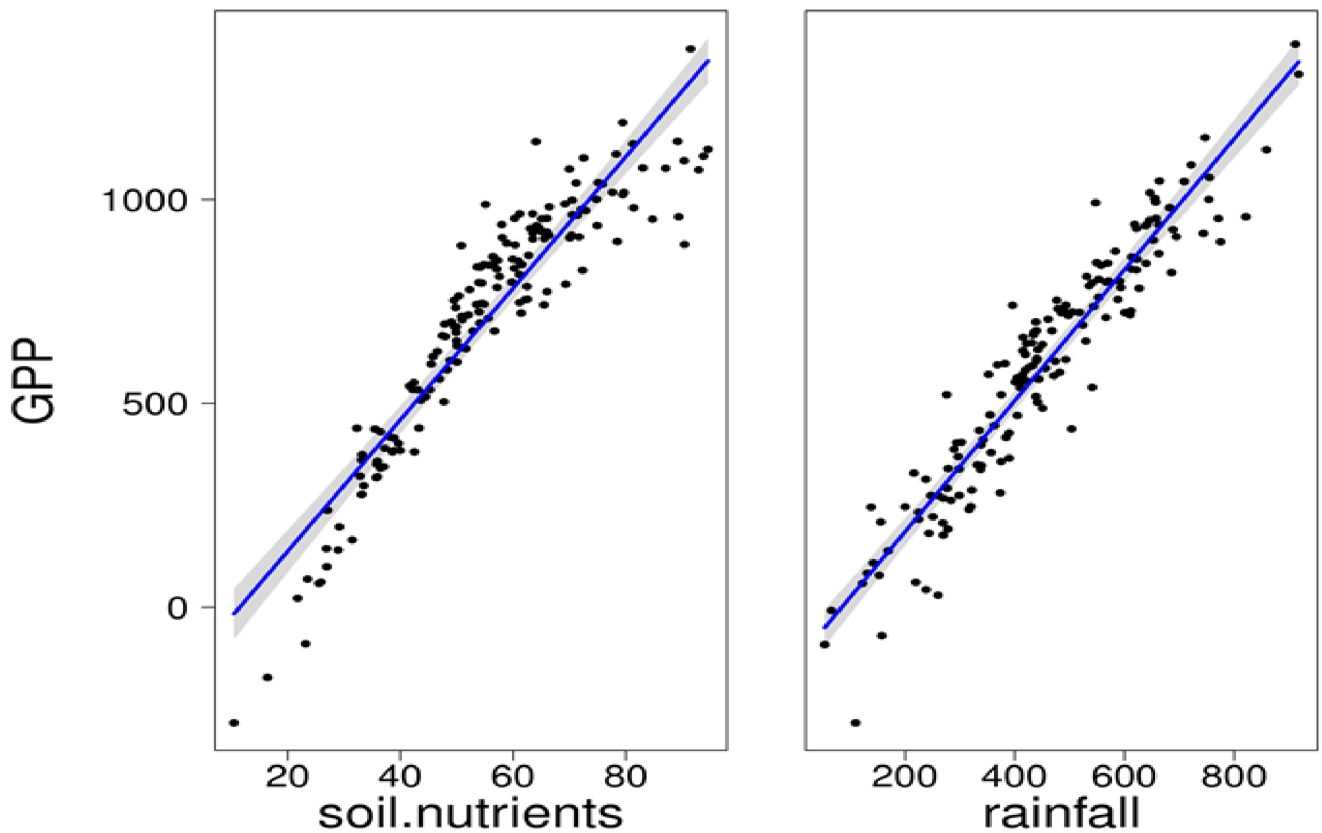

Figure 3: Partial relationship plots between GPP and soil.nutrients and between GPP and rainfall, holding everything else constant

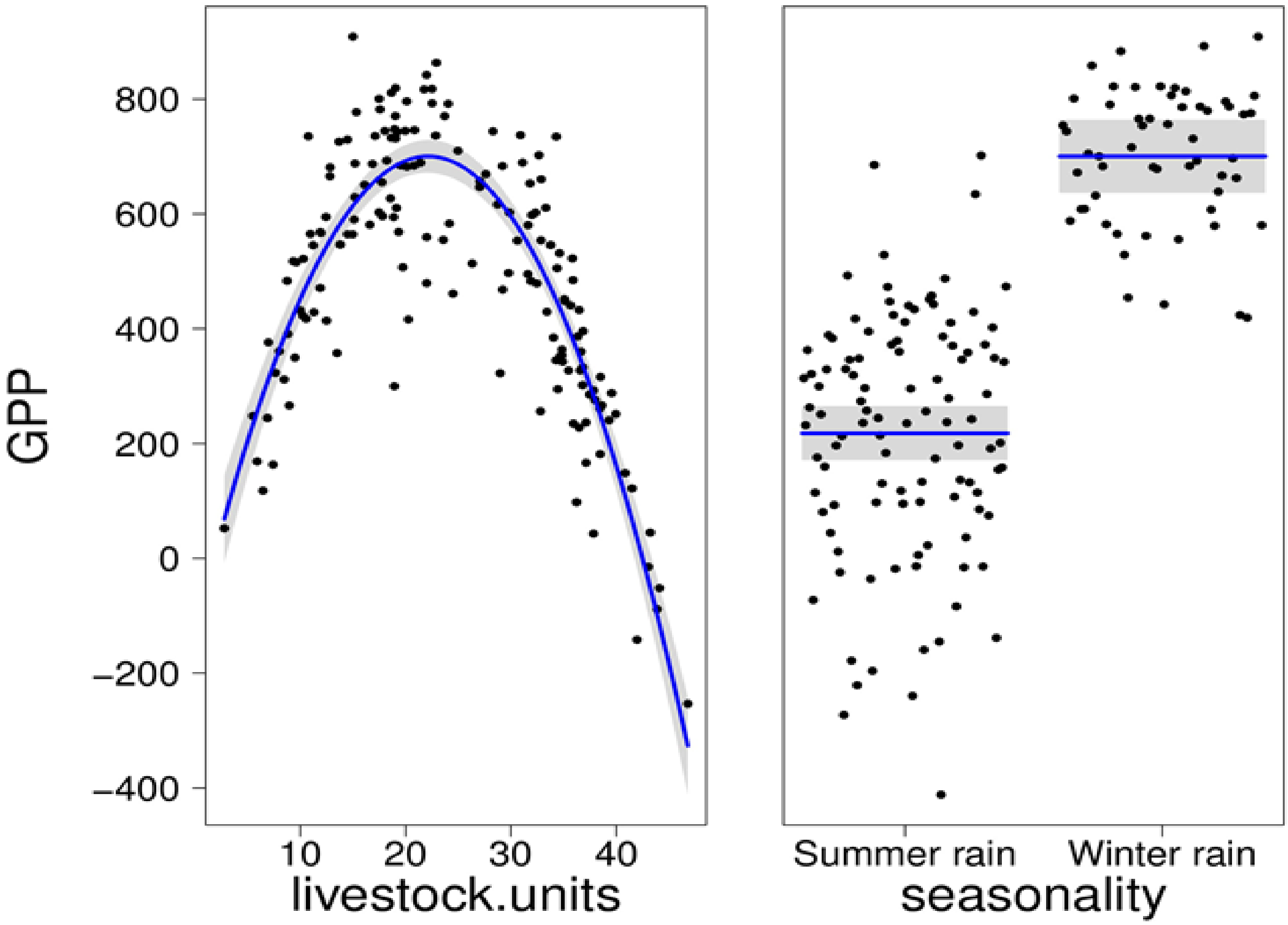

Figure 4: Partial relationship plots between GPP and livestock.units and between GPP and seasonality, holding everything else constant

Constant error variance:

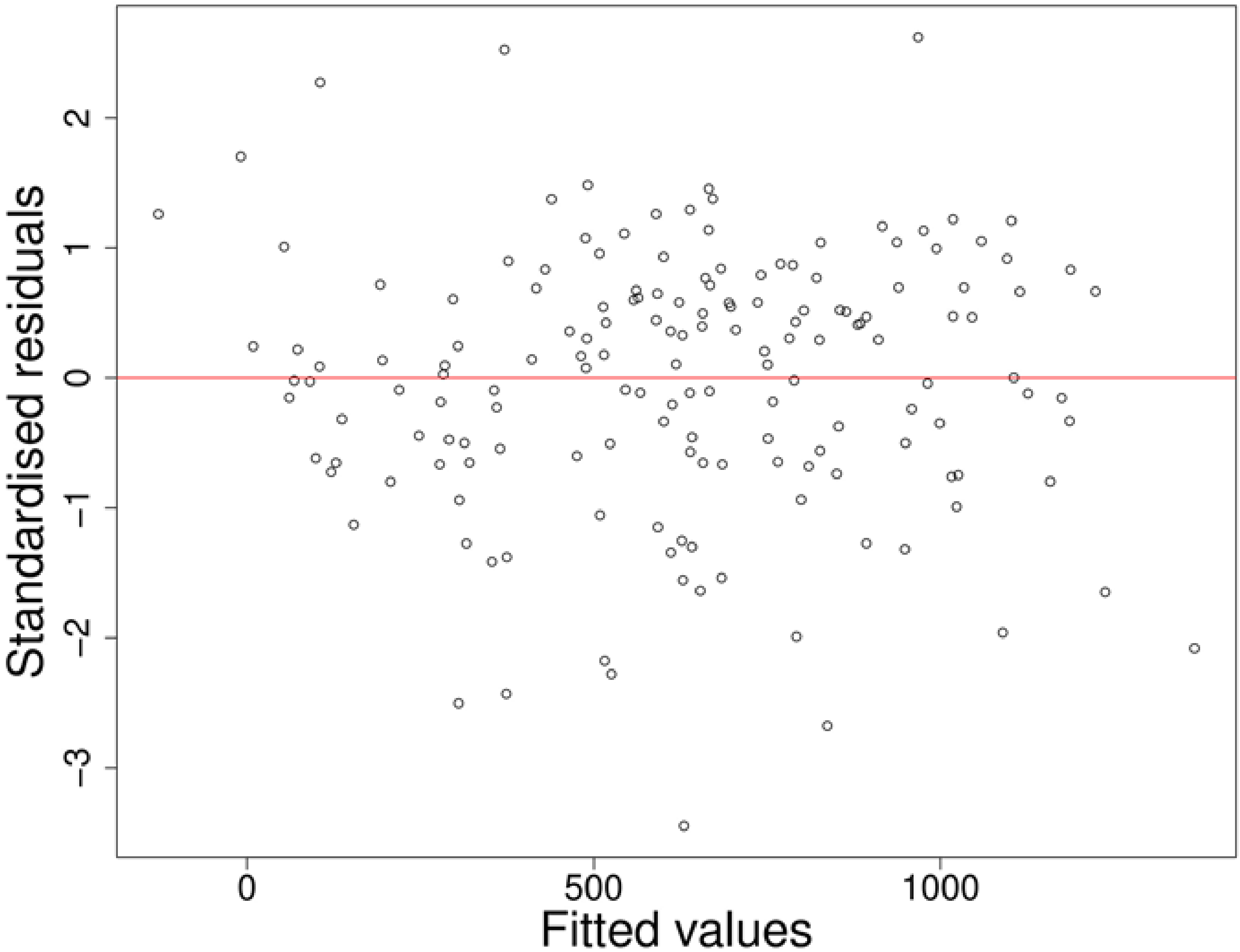

Figure 5: Standardised residuals vs fitted values

Independence:

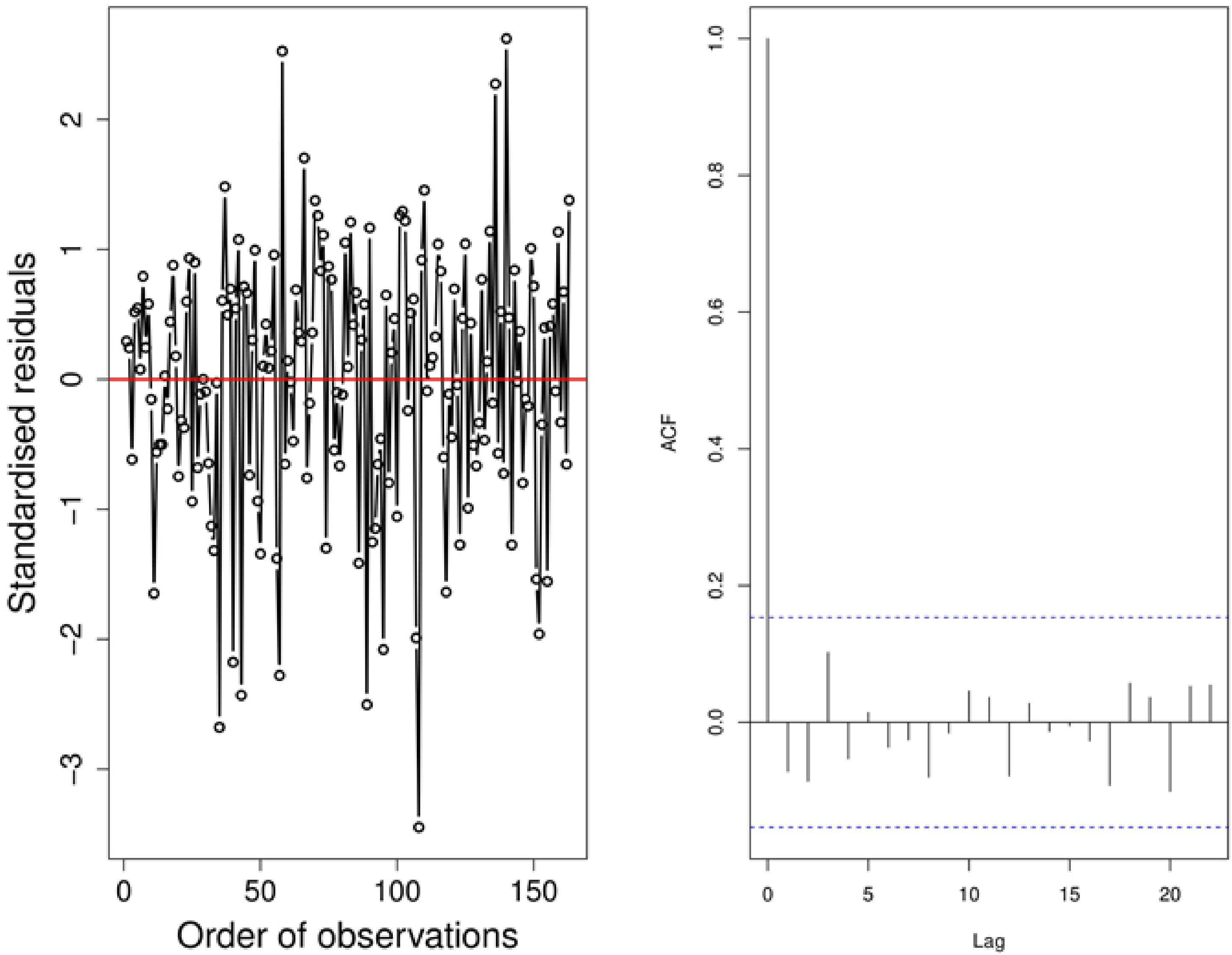

Figure 6: Standardised residuals vs order of observations

Model output:

Durbin-Watson test

data	m1
DW	2.1319
p-value	0.4605

Table 6: Output for the Durbin-Watson test for model m1

Normality:

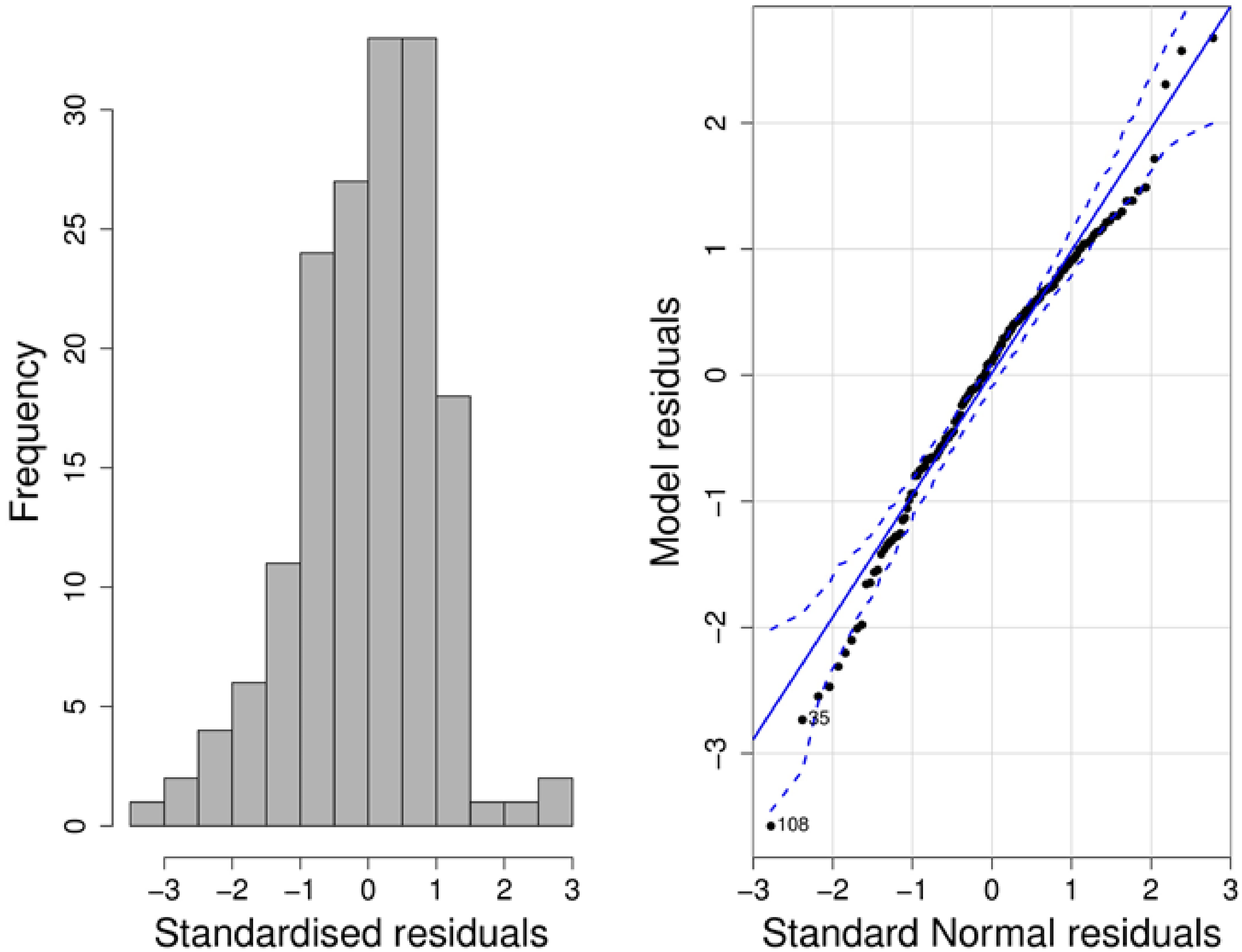

Figure 7 Standardised residual plots measuring symmetry of residuals for model m1

shapiro.test(rstandard(m1))

Model output:

Shapiro-Wilk normality test

data	rstandard(m1)
W	0.97637
p-value	0.006838

Table 7: Output of the Shapiro-Wilk normality test for model m1

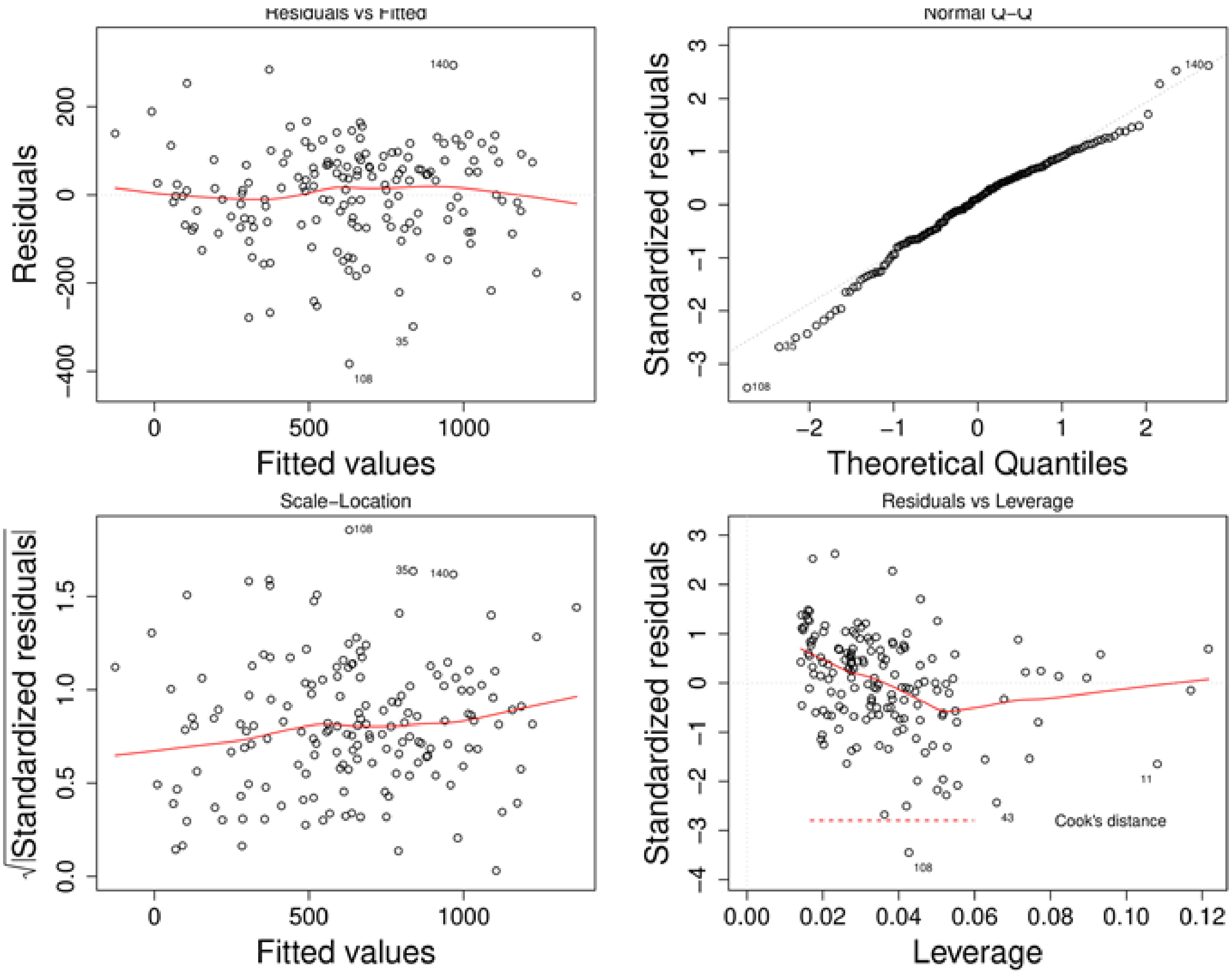

Figure 8: Residual plots for model m1

This shows four residual plots for m1. These plots are the most frequently used plots – but it is more useful to look at more things as we have done above.

Model checking for m2:

Linearity:

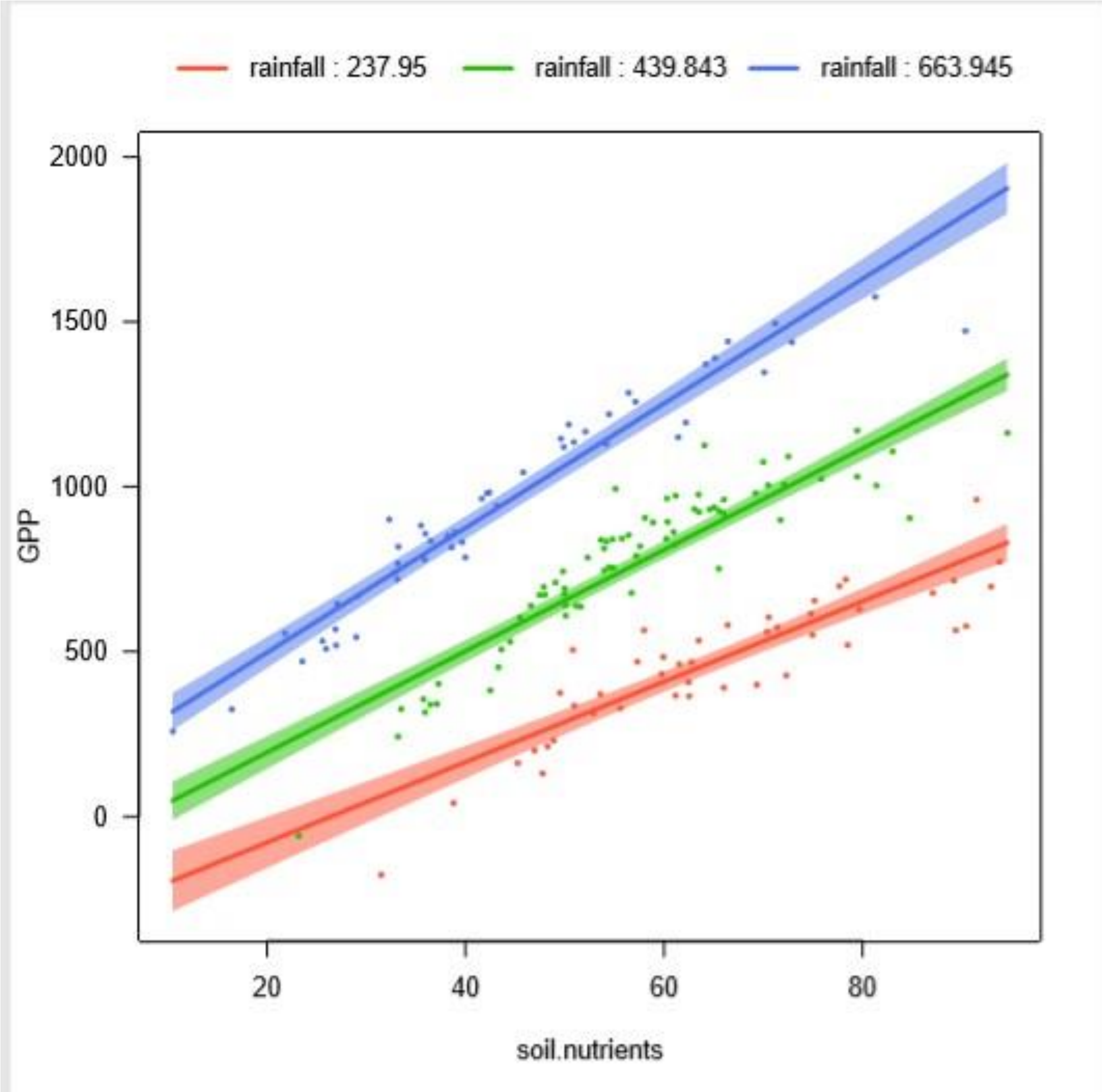

Figure 9: Interaction plot for model m2

This plot in figure 9 indicates the interaction between soil nutrients and rainfall. Rainfall is broken into 3 classes: low, medium and high rainfall and we show the effect of soil.nutrients on each of these rainfall classes.

rainfall (low)	237.95
rainfall (medium)	439.843
rainfall (high)	663.945

At low rainfall, gpp increases with increasing soil.nutrients.

At medium rainfall, gpp increases stronger with increasing soil.nutrients.

At high rainfall, gpp increases even stronger with increasing soil.nutrients.

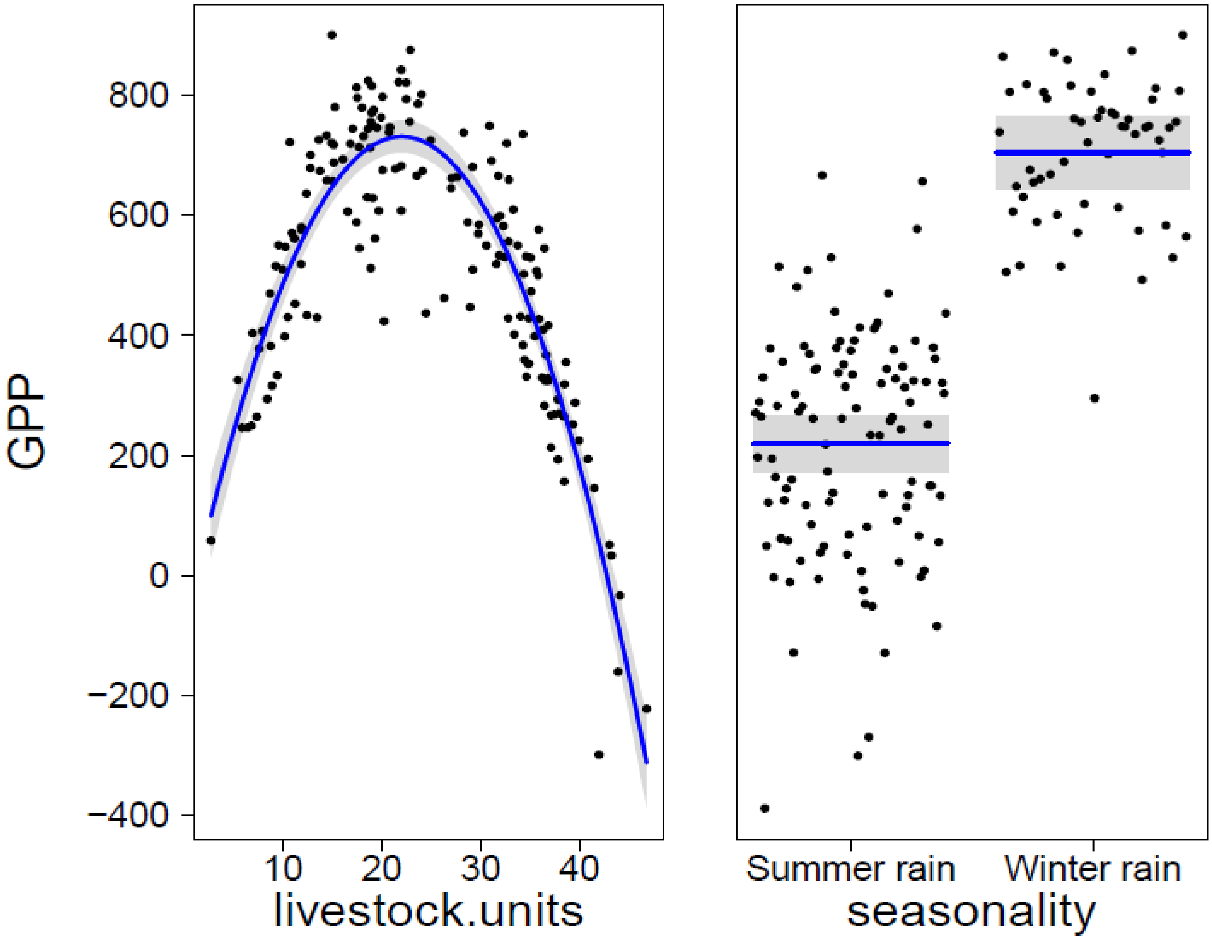

Figure 10: Partial relationship plots between GPP and livestock.units and between GPP and seasonality for model m2

The partial relationship plot in figure 10 suggests that the relationship between GPP and livestock.units is not linear but better described by a quadratic function.

Constant error variance:

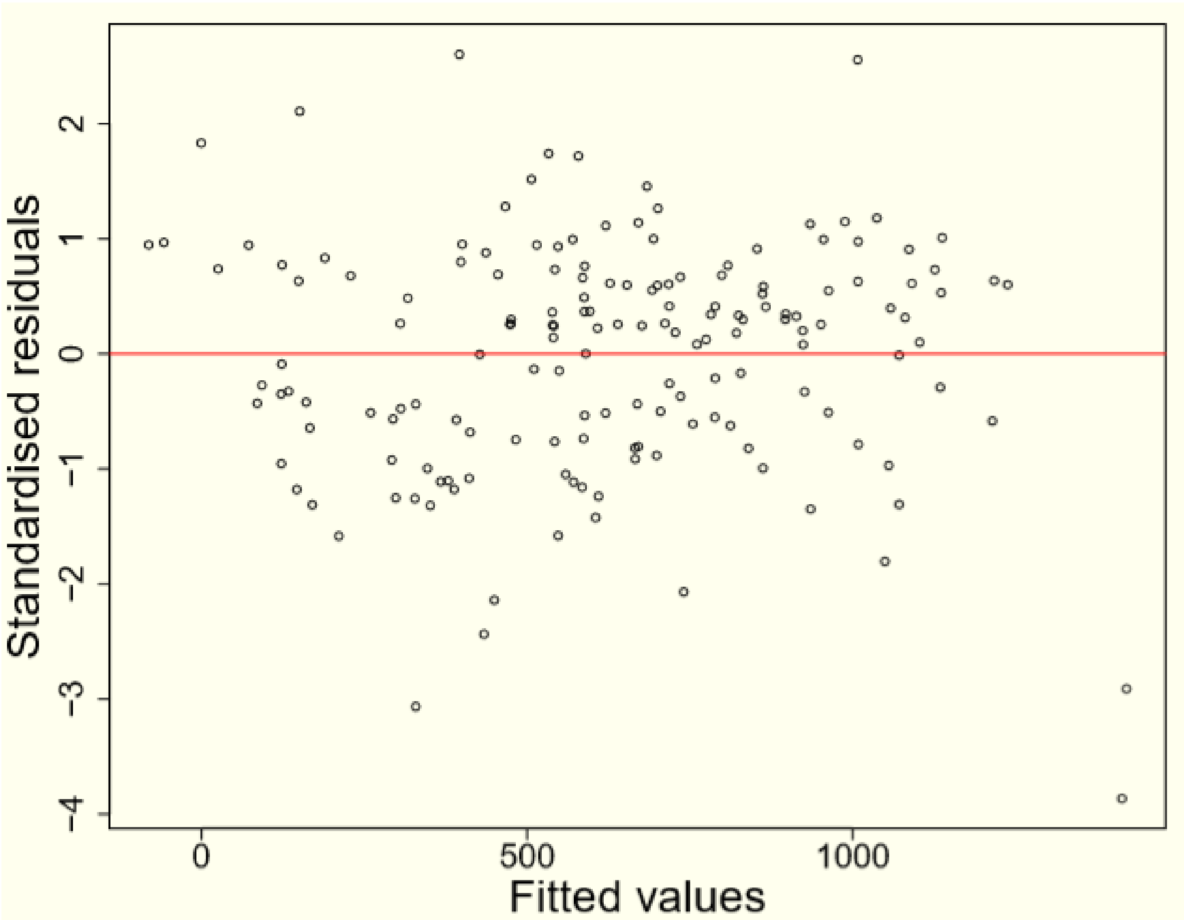

Figure 11: Standardised residual vs fitted values for model m2

The plot in figure 11 shows well behaved residuals, randomly scattered around the zero line with no obvious patterns.

Independence:

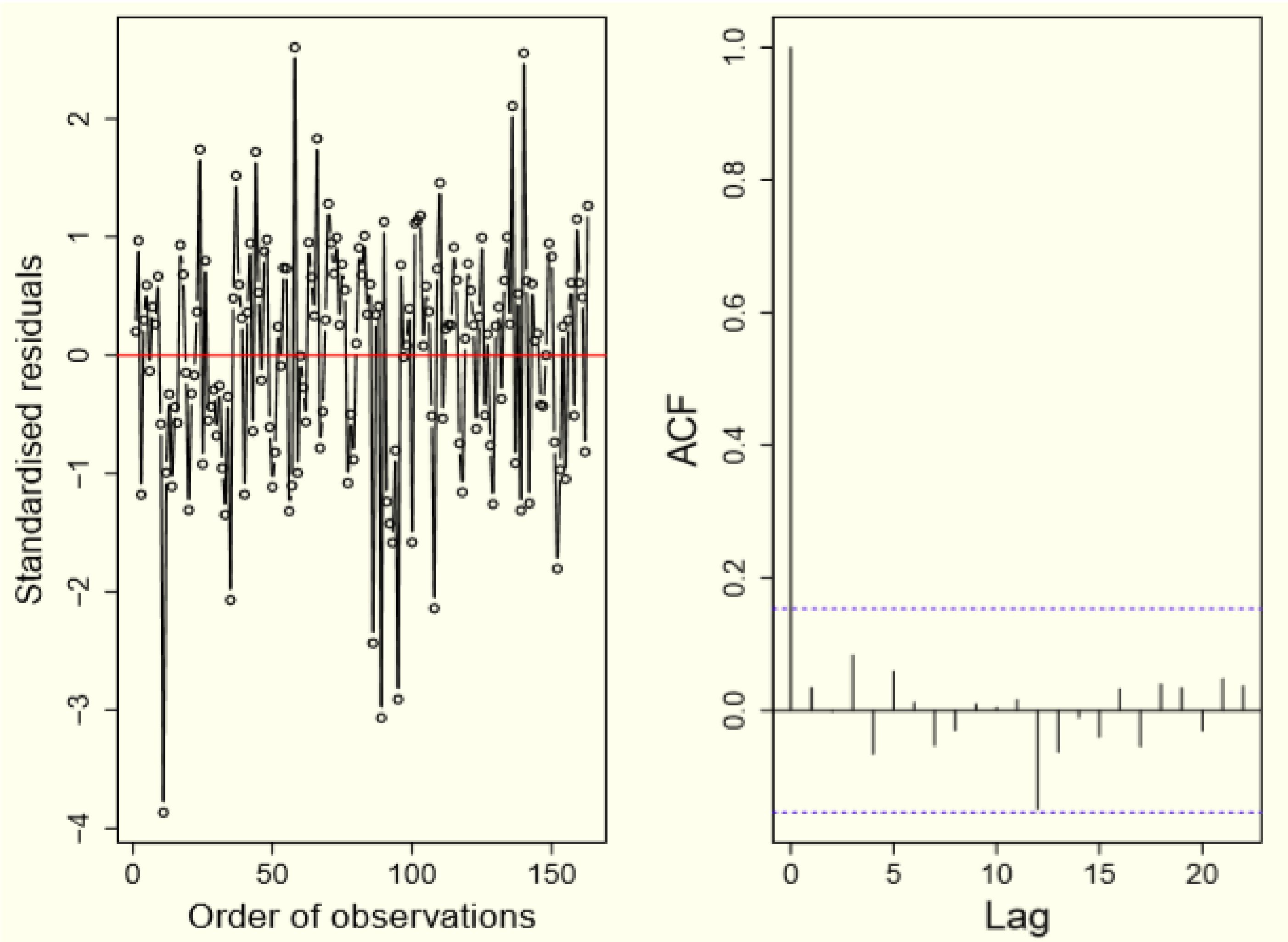

Figure 12: Standardised residuals vs order of observations for model m2

Model output:

Durbin-Watson test

data	m2
DW	1.9236
p-value	0.5422

Table 8: Output of Durbin-Watson test for model m2

Normality:

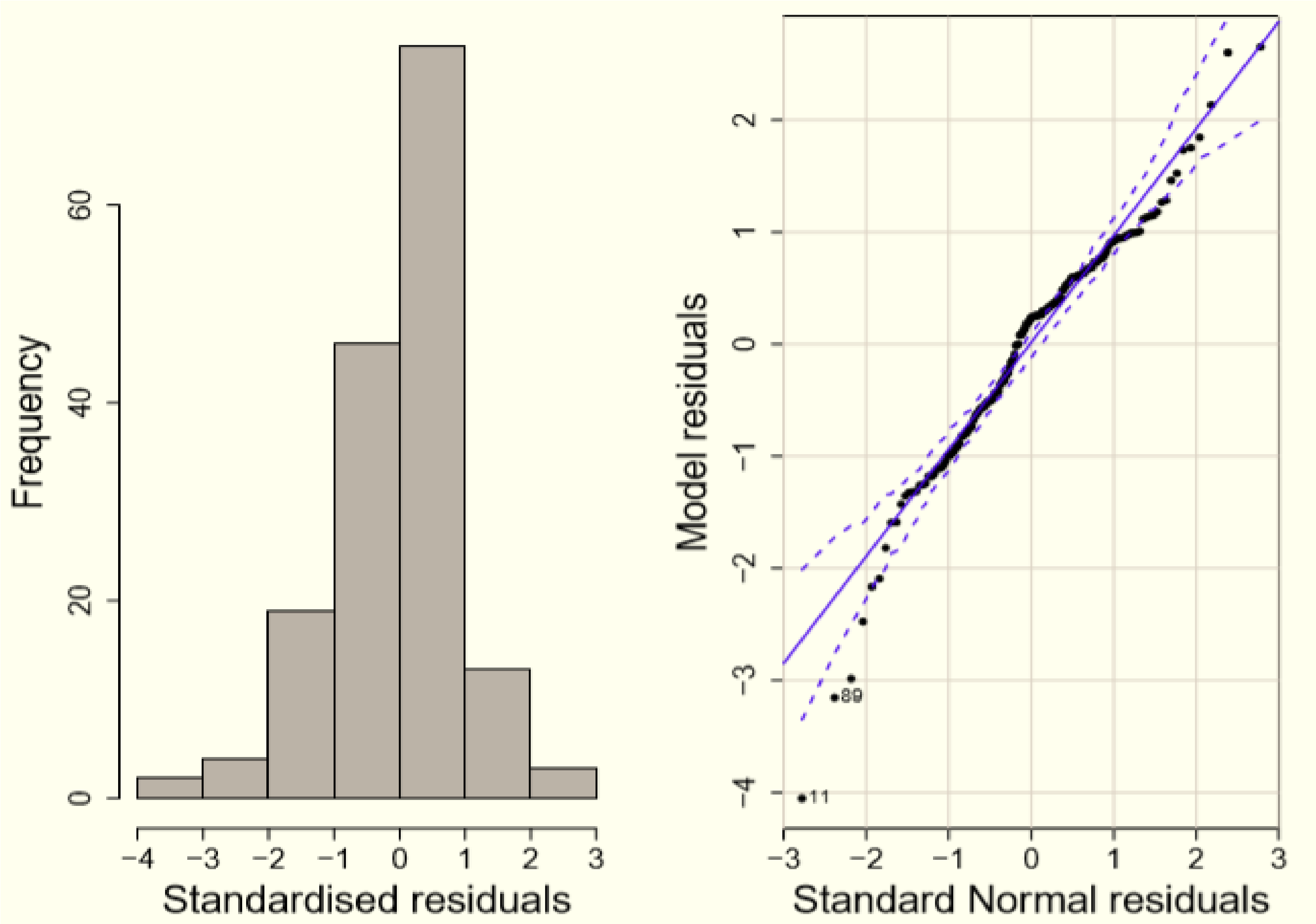

Figure 13: Standardised residuals frequency plots for model m2

Model output:

Shapiro-Wilk normality test

data	rstandard(m2)
W	0.96965
p-value	0.001202

Table 9: Output of Shapiro-Wilk normality test for model m2

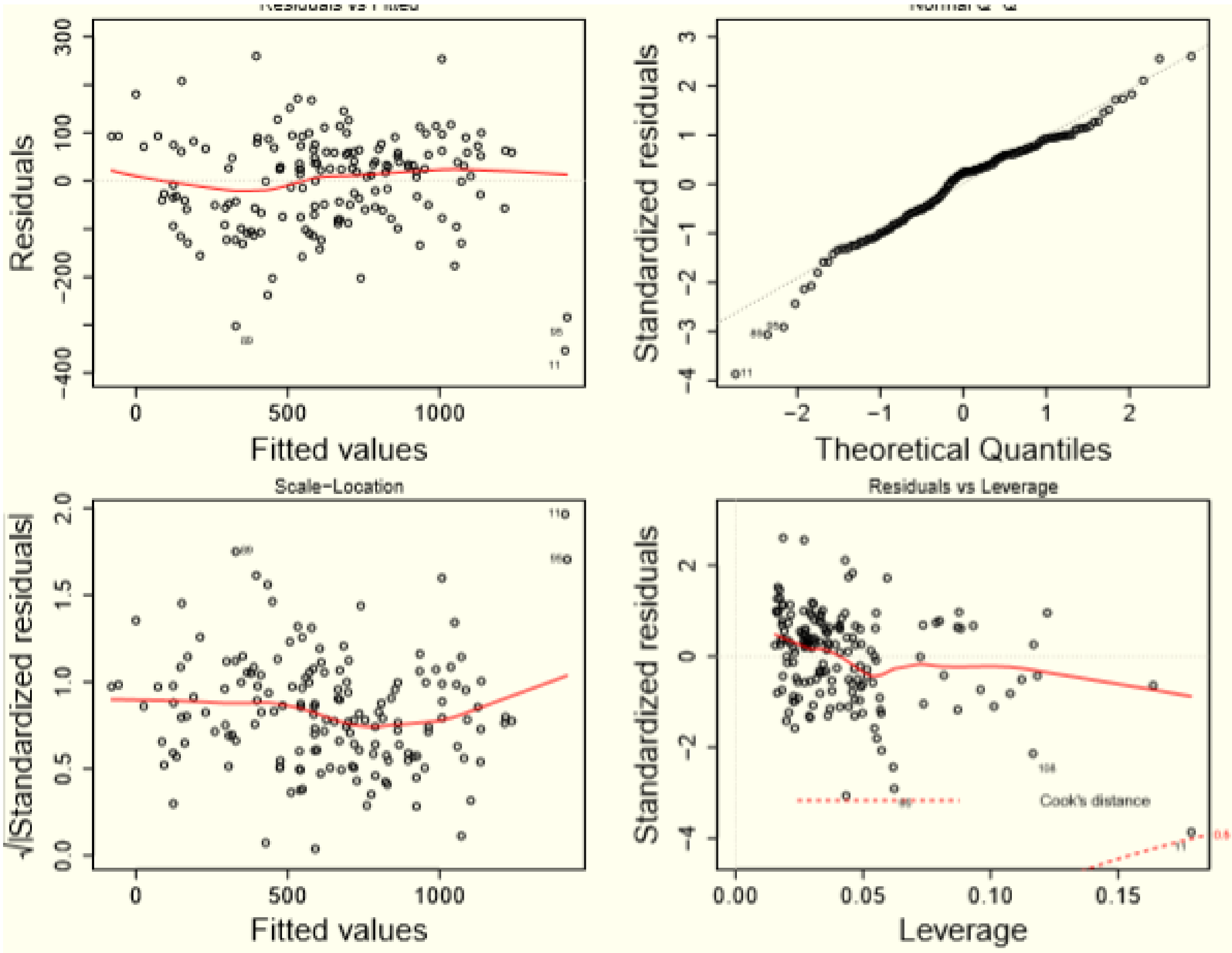

Figure 14: Residual plots for model m2

Finding the best model: model comparison and selection

We used Akaike's Information Criterion (AIC) to select the best model.

Model number	Model name	-2xloglik	K	AIC	Delta AIC	w
1	m1	1999.231079	7	2013.231	37.9829152	5.65E-09
2	m2	1959.248164	8	1975.248	0	1.00E+00

Table 10: Output for AIC for model m2

Presentation of results of best model (m2):

lm(formula = GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality, data = gpp)

Model output:

Residuals:

Min	1Q	Median	3Q	Max
-352.99	-60.62	24.24	64.56	259.84

Coefficients:

Variable	Estimate	Std. Error	t-value	Pr(>|t|)
Intercept	-1.396e+03	9.902e+01	-14.095	< 2e-16***
soil.nutrients	8.454e+00	1.289e+00	6.558	7.58e-10***
rainfall	1.039e+00	1.421e-01	7.307	1.33e-11***
livestock.units	7.510e+01	3.969e+00	18.920	< 2e-16***
I(livestock.units^2)	-1.703e+00	8.058e-02	-21.133	< 2e-16***
seasonalityWinter rain	2.817e+02	1.732e+01	16.262	< 2e-16***
soil.nutrients:rainfall	1.569e-02	2.382e-03	6.585	6.56e-10***

Residual standard error	100.8 on 156 degrees of freedom
Multiple R-squared	0.909
Adjusted R-squared	0.9055
F-statistic	259.7 on 6 and 156 DF
p-value	< 2.2e-16

Table 11: Output of regression analysis of m2

t0 <- lm(GPP ~ 1, data=gpp)

t1 <- lm(GPP ~ soil.nutrients * rainfall, data=gpp)

t2 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units, data=gpp)

t3 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2), data=gpp)

t4 <- lm(GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality, data=gpp)

anova(t0, t1, t2, t3, t4)

Model Output:

ANOVA TABLE	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model 1: GPP ~ 1	162	17413299				
Model 2: GPP ~ soil.nutrients * rainfall	159	8808850	3	8604449	282.374	< 2.2e-16 ***
Model 2: GPP ~ soil.nutrients * rainfall+livestock.units	158	7863633	1	945217	93.058	< 2.2e-16 ***
Model 4: GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2)	157	4270670	1	3592964	353.733	< 2.2e-16 ***
Model 5: GPP ~ soil.nutrients * rainfall + livestock.units + I(livestock.units^2) + seasonality	156	1584533	1	1584533	264.455	< 2.2e-16 ***

Table 12: ANOVA Table for test models 1 to 5

Confidence intervals for regression coefficients:

confint(m2)

Model output:

Variable	**2.5%**	**97.5%**
Intercept	-1.591290e+03	-1.200107e+03
soil.nutrients	5.908008e+00	1.100095e+01
rainfall	7.577554e-01	1.319251e+00
livestock.units	6.725508e+01	8.293557e+01
I(livestock.units^2)	-1.862037e+00	-1.543706e+00
seasonalityWinter rain	2.475085e+02	3.159495e+02
soil.nutrients:rainfall	1.098281e-02	2.039437e-02

Table 13: Output for confidence intervals for regression coefficients for model m2

Conclusions and Discussion:

Figure 1 shows that the spread of the response variable is approximately symmetrical, so the response does not require a log transformation. There are also a few large values but it is possible that the explanatory variables may explain these.

The pairs plot in figure 2 suggests a positive linear relationship between GPP and soil nutrients as well as GPP and rainfall.

Interpretation of model output in table 1:

In table 1 it can be seen that livestock.units and land.use are correlated (variance inflation factors (vif) above 4 or 5).

Interpretation of model output in table 2:

In table 2 can be seen that there is no multicollinearity (variance inflation factors (vif) under 4 or 5).

Interpretation of model outputs in tables 3 and 4:

In tables 3 and 4 we can see that there is no multicollinearity - no inflated standard errors and consequently no large p-values even though variance inflation factors (vif) for the explanatory variables are now larger than 4 or 5.

The high variance inflation factors (vif) doesn't matter in this case, since in the termites.R example code after adding the interaction term, the vif is also well above 4 or 5 (as seen in table 5).

Model checking for m1:

Linearity:

The partial relationship plots in figure 3 suggest a positive linear relationship between GPP and soil nutrients as well as GPP and rainfall. The partial relationship plot in figure 4 suggests that the relationship between GPP and livestock.units is not linear but better described by a quadratic function.

Constant error variance:

There is constant error variance / homoscedasticity.

The plot in figure 5 shows well behaved residuals, randomly scattered around the zero line with no obvious patterns.

Independence:

Residuals are independent.

The first plot in figure 6 shows that there are no consecutive big residuals or small residuals.

The second plot shows no autocorrelation – values for lags >= 1 falls within the dotted lines.

Interpretation of model output in table 6:

The null hypothesis is that autocorrelation is 0.

alternative hypothesis: true autocorrelation is not 0

The large p-value (0.4605) indicates that there is not much evidence that autocorrelation is not 0.

Normality:

Residuals are normal.

The first plot in figure 7 shows that the residuals are symmetrical within 2 standard errors of zero.

The second plot shows that the residuals appear approximately along a straight line within 2 standard errors of zero.

Interpretation of model output in table 7:

The null hypothesis is that the sample distribution is normal. No sample will be perfectly normal an one does not have to prove that residuals are normal, only check that they do not seriously violate this assumption.

This assumption is approximately met. The approximately large p-value (0.006838) indicates that there is not much evidence that the sample distribution is not normal.

Figure 8 shows four residual plots for m1. These plots are the most frequently used plots – but it is more useful to look at more things as we have done above.

Model checking for m2:

Linearity:

The plot in figure 9 indicates the interaction between soil nutrients and rainfall. Rainfall is broken into 3 classes: low, medium and high rainfall and we show the effect of soil.nutrients on each of these rainfall classes.

rainfall (low)	237.95
rainfall (medium)	439.843
rainfall (high)	663.945

At low rainfall, gpp increases with increasing soil.nutrients.

At medium rainfall, gpp increases stronger with increasing soil.nutrients.

At high rainfall, gpp increases even stronger with increasing soil.nutrients.

The partial relationship plot in figure 10 suggests that the relationship between GPP and livestock.units is not linear but better described by a quadratic function.

Constant error variance:

There is constant error variance / homoscedasticity.

The plot in figure 11 shows well behaved residuals, randomly scattered around the zero line with no obvious patterns.

Independence:

Residuals are independent.

The first plot in figure 12 shows that there are no consecutive big residuals or small residuals.

The second plot shows no autocorrelation – values for lags >= 1 falls within the dotted lines.

Interpretation of model output in table 8:

The null hypothesis is that autocorrelation is 0.

alternative hypothesis: true autocorrelation is not 0

The large p-value (0.5422) indicates that there is not much evidence that autocorrelation is not 0.

Normality:

Residuals are normal.

The first plot in figure 13 shows that the residuals are symmetrical within 2 standard errors of zero.

The second plot shows that the residuals appear approximately along a straight line within 2 standard errors of zero.

Interpretation of model output in table 9:

The null hypothesis is that the sample distribution is normal. No sample will be perfectly normal an one does not have to prove that residuals are normal, only check that they do not seriously violate this assumption.

This assumption is approximately met. The approximately large p-value (0.001202) indicates that there is not much evidence that the sample distribution is not normal.

Figure 14 shows four residual plots for m2. These plots are the most frequently used plots – but it is more useful to look at more things as we have done above.

Interpretation of model output in table 10:

We used Akaike's Information Criterion (AIC) to select the best model.

Model m2 has the smallest AIC value and is therefore the best model in our set. Looking at the ΔAIC makes it clear that models 1 and 2 are close competitors for being the best model in the set. We calculated the Akaike weights, ,ω_i, which are interpreted as strength of evidence for a particular model to be the best, relative to the other models in the set. These Akaike weights sum up to 1 across the models in the set. It can be seen that model m2 had approximately 100% of the support relative to m1. We also calculated evidence ratios, i.e. the odds that one model is in fact the best, compared to another model.

Comparing model m2 to model m1, the evidence ratio is:

$$\frac{m_2}{m_1} = \frac{1.000000e + 00}{5.650863e - 09} = 176964132.1$$

Thus model m2 was 176964132.1 more likely than m1 to be the best in the set.

Interpretation of model output in table 11:

$$H_i = \beta_0 + \beta_1 \times \text{soil.nutrients}_i + \beta_2 \times \text{rainfall}_i + \beta_3 \times \text{livestock.units}_i + \beta_4 \times (\text{livestock.units})^2 + \beta_{\text{seasonalityWinter rain}} + \beta_{1:2}(\text{soil.nutrients}_i \times \text{rainfall}_i) + \epsilon_i$$

The estimated regression coefficients are:

$\beta_0 =$ -1.396e+03

$\beta_1 =$ 8.454e+00

- The p-value is very small (7.58e-10) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of zero slope / no relationship)
- The small standard error (1.289e+00) indicates that the estimate is close the true value.

$\beta_2 =$ 1.039e+00

- The p-value is very small (1.33e-11) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of zero slope / no relationship)
- The small standard error (1.421e-01) indicates that the estimate is close the true value.

$\beta_3 =$ 7.510e+01

- The p-value is very small (< 2e-16) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of zero slope / no relationship)
- The small standard error (3.969e+00) indicates that the estimate is close the true value.

$\beta_4 =$ -1.703e+00

- The p-value is very small (< 2e-16) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of zero slope / no relationship)
- The small standard error (8.058e-02) indicates that the estimate is close the true value.

$\beta_{winter} =$ 2.817e+02

- The p-value is very small (< 2e-16) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of being relative to the baseline level)
- The small standard error (1.732e+01) indicates that the estimate is close the true value.

$\beta_{1:2} =$ 1.569e-02

- The p-value is very small (6.56e-10) which indicates strong evidence against the null hypothesis (i.e. null hypothesis of no interaction)
- The small standard error (2.382e-03) indicates that the estimate is close the true value.

The fitted regression equation can be written as follows:

$$\hat{y} = (-1.396\text{e} + 03) + (8.454\text{e} + 00) \times \text{soil.nutrients}_i + (1.039\text{e} + 00) \times \text{rainfall}_i + (7.510\text{e} + 01) \times \text{livestock.units}_i + (-1.703\text{e} + 00) \times (\text{livestock.units})^2 + (2.817\text{e} + 02) \times (0/1) + (1.569\text{e} - 02)(\text{soil.nutrients}_i \times \text{rainfall}_i)$$

The intercept estimate β_0 tells us that the gross primary productivity (GPP) when the value of all the explanatory variables is zero is -1.396e+03. The interpretation of the intercept is not sensible in this case as the gross primary productivity (GPP) studied here does not measure of the amount of photosynthesis taking place with zero soil nutrients, rainfall and livestock units.

If soil.nutrients increases by 1 unit, GPP increases by 8.454e+00 units holding constant the value of all other explanatory variables. If rainfall increases by 1 unit, GPP increases by 1.039e+00 units holding constant the value of all other explanatory variables. If livestock.units increases by 1 unit, GPP increases by 7.510e+01 units holding constant the value of all other explanatory variables. If increases by 1 unit, GPP decreases by 1.703e+00 units holding constant the value of all other explanatory variables. The coefficient for the level of the categorical variable (here "Winter rain") 2.817e+02 estimate the effect of Winter rain – measures the change in mean response relative to the baseline category. The mean response for the baseline level (here "seasonalitySummer rain") -1.396e+03, is estimated by the intercept.If $(\text{livestock.units})^2$ increases by 1 unit, GPP decreases by 1.703e+00 units holding constant the value of all other explanatory variables. The coefficient for the level of the categorical variable (here "Winter rain") $\beta_{\text{seasonalityWinter rain}} =$ 2.817e+02 estimate the effect of Winter rain – measures the change in mean response relative to the baseline category. The mean response for the baseline level (here "seasonalitySummer rain") $\beta_0 =$ -1.396e+03, is estimated by the intercept. The positive interaction means that if rainfall increases by 1 unit, soil.nutrients (effect) increases by 1.569e-02 units.

The adjusted R^2 is 90.55% which indicates most of the variance in the response can be explained by the model. For the F-test, the p-value is very small (< 2.2e-16) which indicates strong evidence against the null hypothesis (i.e. null hypothesis that the explanatory variables have no explanatory power).

Our results suggest that the factors that had the most influence on the gross primary productivity (GPP) were livestock units and seasonality as well as an interaction between soil nutrients and rainfall. We found that this interaction was in fact positive and not negative as predicted in our hypothesis.

We used analysis of variance / deviance to understand which are the variables that had the greatest effect on the response.

Interpretation of model output in table 12:

The p-values are very small (< 2.2e-16) which indicates strong evidence against the null hypothesis (i.e. null hypothesis that the extra parameters are all zero), and therefore there is evidence that the extra parameters improves the model (improves the amount of variation explained). The addition of the terms *soil.nutrients * rainfall* followed by *I(livestock.units^2)* caused the most reduction in RSS (residual sum of squares). Hence, the variables that had the greatest effect on the response were *soil.nutrients, rainfall* and *livestock.units.*

Interpretation of model output in table 13:

The confidence intervals for regression coefficients above are quite narrow which conveys high precision in the estimates (i.e. low degree of uncertainty).

References:

The work presented is not our own, we have borrowed from many sources:

1. Davis, Miranda, Lynn (2013) *Projecting the impact of climate change on ungulate population dynamics: the importance of trophic interactions*, Durham theses, Durham University.
2. EL-Vilaly, M., Didan, K., Marsh, S., Crimmins. M and Munoz, A. 2018. *Characterizing Drought Effects on Vegetation Productivity in the Four Corners Region of the US Southwest,* The University of Arizona.
3. Kline, G. 2012. *Effects of Rain on Soil Fertility*. <http://sanctuarysoil.com/effects-of-rain-on-soil-fertility-guest-post-by-gary-kline>
4. Mathews, C. 2006. *Livestock a major threat to environment* <http://www.fao.org/newsroom/en/news/2006/1000448/index.html>
5. Nemani, R., Keeling, D., Hashimoto, H., Jolly, M., Pipers, C., Tucker, J., Mynenis, B. and Running, W. 2003. *Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999.* Science 06 Jun 2003: Vol. 300, Issue 5625, pp. 1560-1563.
6. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. < https://www.R-project.org/>