

Fallstudien II

Laura Kampmann, Christian Peters, Alina Stammen

11. Dezember 2020

1. Task I - Vorhersage der Datenrate

Gradient Boosted Trees

Regression mit ARMA-Fehlern

Validierung

2. Task II - Handover Vorhersage und Link Lifetime

Lösungsansatz Task II

Task I - Vorhersage der Datenrate

Task I - Vorhersage der Datenübertragungsrate

- Ziel: Evaluation von neuen *anticipatory vehicular communication systems* durch möglichst realitätsnahe Simulationen [5]
⇒ Ansatz: *Data-Driven Network Simulation*
- Durch Machine Learning Modelle sollen möglichst realistische Vorhersagen der Datenraten generiert werden
- Hoffnung: Bessere Aussagekraft der Simulationen durch Einsatz echten Datenmaterials

Task I - Vorhersage der Datenrate

Gradient Boosted Trees

Gradient Boosted Trees

- Kann man aus vielen "schwachen" Lernern einen starken Lerner konstruieren?
 - ⇒ Ja, Boosting ist eines der mächtigsten Konzepte des Machine Learning [2]
- Kombination von einfachen CART Bäumen zu einem starken Ensemble
 - ⇒ Ähnlich zu Random Forest
- Der Unterschied zum Random Forest liegt im Training!

Training von Gradient Boosted Trees

- Bäume werden nacheinander zum Ensemble hinzugefügt
- Jeder neue Baum versucht, die Schwächen seiner Vorgänger "auszubügeln"
 - ⇒ *Additives Training*
- Je mehr Bäume aufgenommen werden, desto geringer wird der Training-Error (das Modell wird aber komplexer)
 - ⇒ Kontrolle des *Bias-Variance Tradeoffs*
 - ⇒ Zusätzlich gibt es Regularisierungs-Parameter

Task I - Vorhersage der Datenrate

Regression mit ARMA-Fehlern

Regression mit ARMA-Fehlern

Gegeben:

- Beobachtungen (y_1, \dots, y_T) der Zeitreihe $(y_t)_t$
- Beobachtungen $(x_1^{(i)}, \dots, x_T^{(i)})$ der Zeitreihen $(x_t^{(i)})_t$ für $i = 1, \dots, k$

Modellgleichung: Regression mit ARMA(p, q)-Fehlern [3]

$$y_t = c + \sum_{j=1}^k \beta_j x_t^{(j)} + \eta_t \text{ mit}$$

$$\eta_t = \underbrace{\sum_{k=1}^p \phi_p \eta_{t-p}}_{\text{vergangene Fehler: LM}} + \underbrace{\sum_{l=1}^q \theta_l \epsilon_{t-q}}_{\text{vergangene Fehler: ARMA}} + \epsilon_t$$

Vorarbeit:

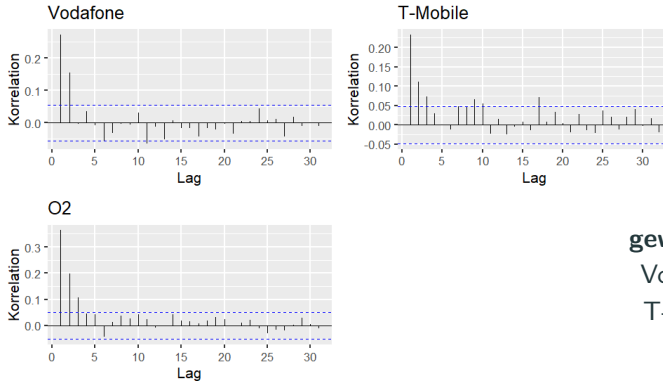
- Überprüfung Autokorrelation der Zielvariablen (Acf, pAcf)
- Standardisierung Train, Skalierung Test

Überprüfung der Voraussetzungen:

- Stationarität aller Variablen (Augmented Dickey-Fuller Test)
- keine Multikollinearität vorhanden (VIF)
- Normalverteilung der Residuen (Scatterplot, Histogramm, QQ-Plot)

Bestimmung des Grids für die AR-Ordnung - Uplink

partielle Autokorrelationsfunktionen der Residuen - Uplink



gewählte Grids:

Vodafone: 0 - 2

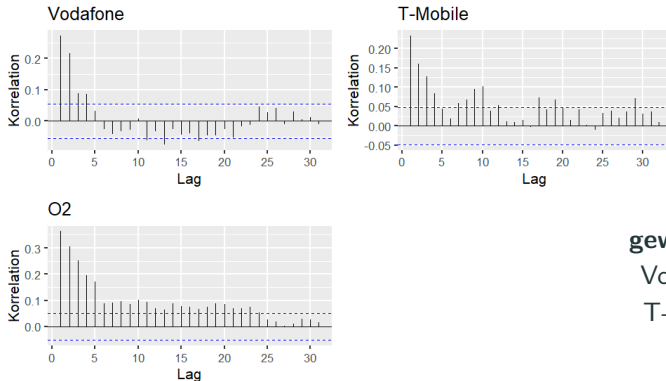
T-Mobile: 0 - 3

O2: 0 - 3

Figure 1: Partielle Autokorrelationsfunktion der Residuen des linearen Modells in Richtung Uplink.

Bestimmung des Grids für die MA-Ordnung - Uplink

Autokorrelationsfunktionen der Residuen - Uplink



gewählte Grids:

Vodafone: 0 - 4

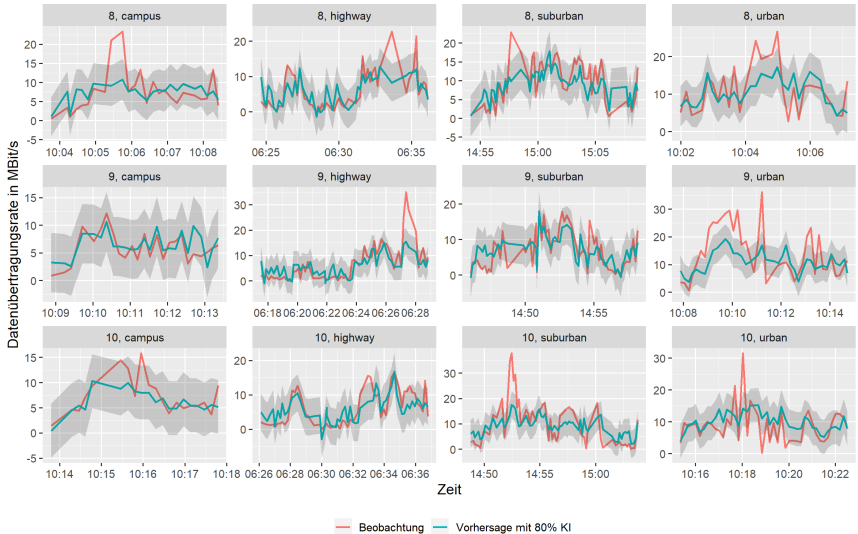
T-Mobile: 0 - 4

O2: 0 - 5

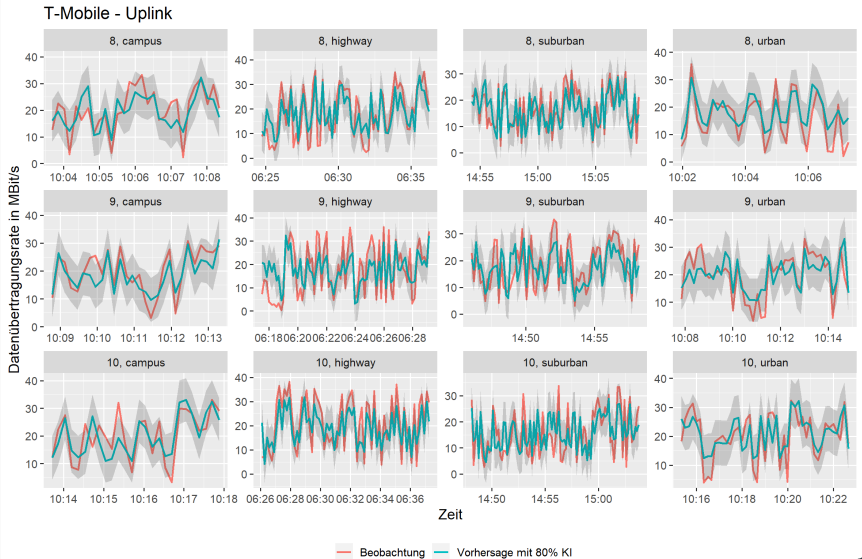
Figure 2: Autokorrelationsfunktion der Residuen des linearen Modells in Richtung Uplink.

Regression mit ARMA-Fehlern: Ergebnisse (Uplink)

Vodafone - Uplink

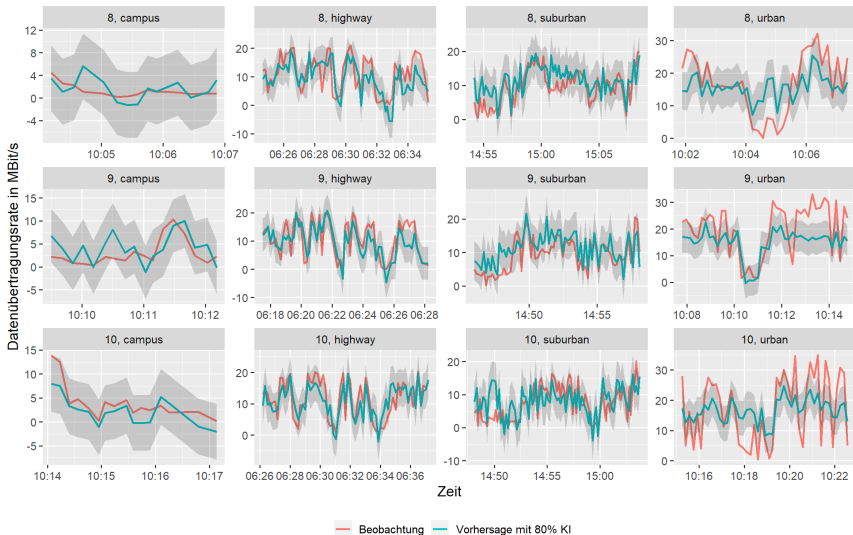


Regression mit ARMA-Fehlern: Ergebnisse (Uplink)



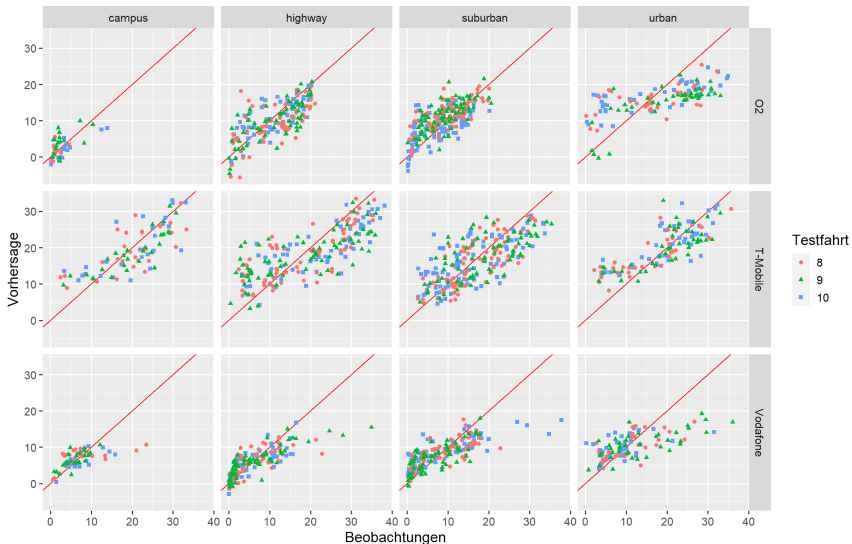
Regression mit ARMA-Fehlern: Ergebnisse (Uplink)

O2 - Uplink

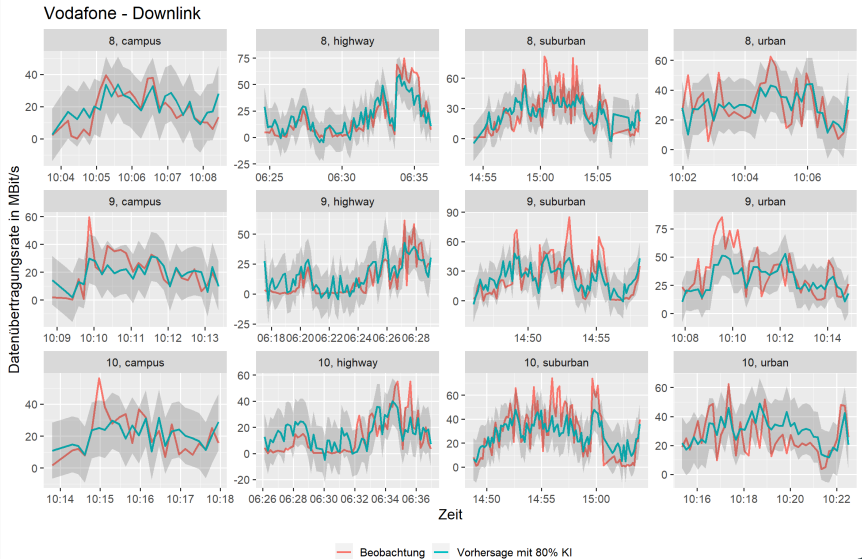


Regression mit ARMA-Fehlern: Ergebnisse (Uplink)

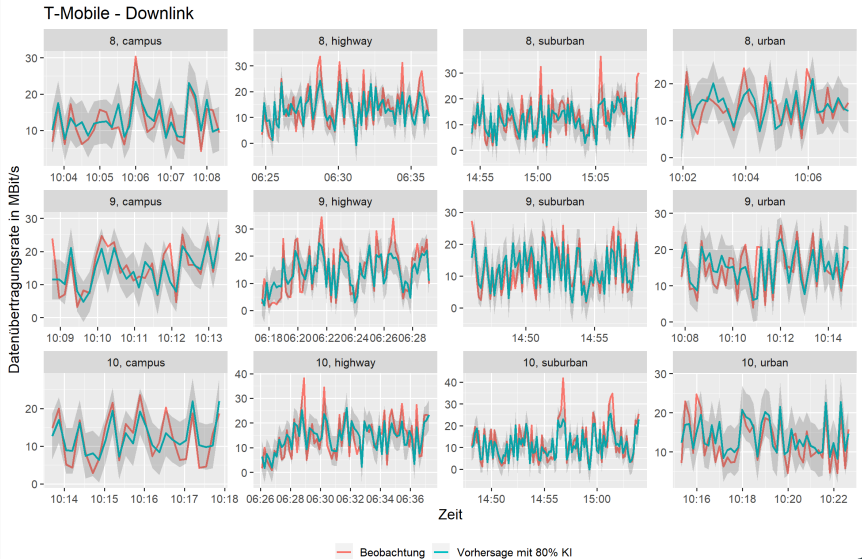
Scatterplot der Beobachtungen und der Vorhersagen: Uplink



Regression mit ARMA-Fehlern: Ergebnisse (Downlink)

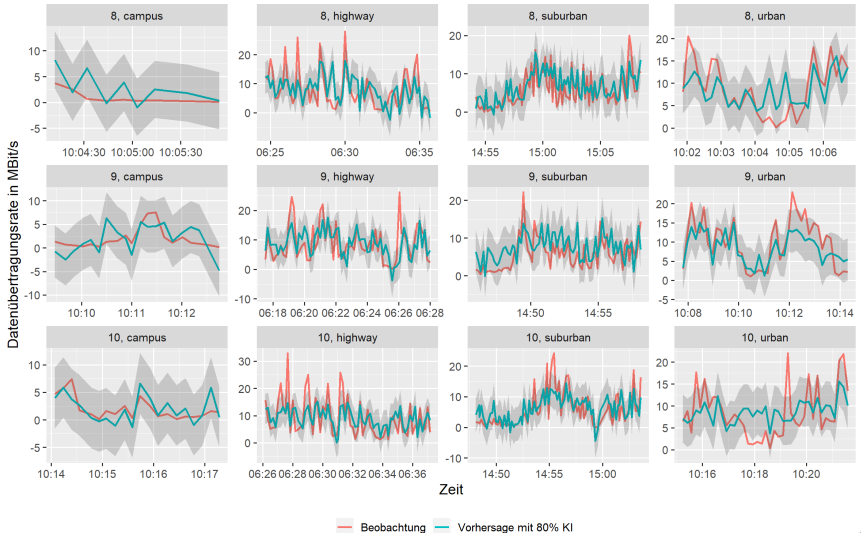


Regression mit ARMA-Fehlern: Ergebnisse (Downlink)

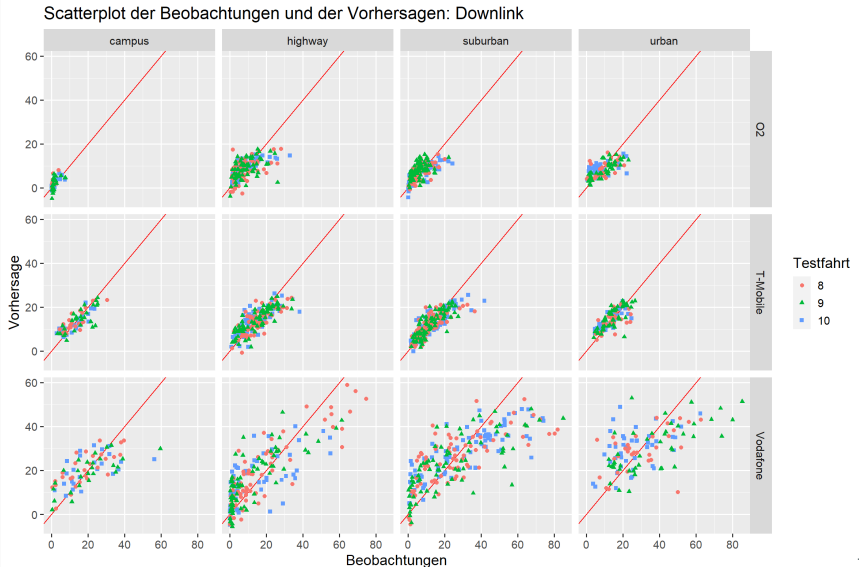


Regression mit ARMA-Fehlern: Ergebnisse (Downlink)

O2 - Downlink



Regression mit ARMA-Fehlern: Ergebnisse (Downlink)



Task I - Vorhersage der Datenrate

Validierung

k-fache Kreuzvalidierung

- beachtet Abhängigkeit der Datenpunkte nicht
- zerstört zeitliche Komponente
- verwendet eventuell zukünftige Beobachtungen für Prognose der Gegenwart

⇒ **Kreuzvalidierung für Zeitreihen**

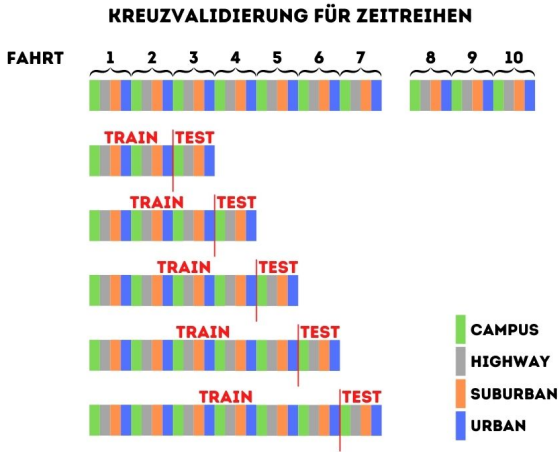
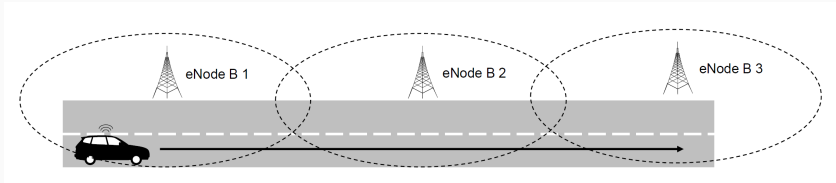


Figure 11: Einteilungen in Trainings- und Testdatensätze bei der Kreuzvalidierung für Zeitreihen.

Task II - Handover Vorhersage und Link Lifetime

Vorhersage des Handovers und Link Lifetime

- Vergleich des RSRP Wertes zur verbundenen Zelle sowie zu den Nachbarzellen
- Vorhersage des Handovers durch Angabe der Link Lifetime



Task II - Handover Vorhersage und Link Lifetime

Lösungsansatz Task II

Lösungsansatz Task II

Idee: Prädiktionsmodell für Link Lifetime mit Einfluss des RSRP der verbunden sowie der Nachbarzellen

→ Datentransformation nötig

- Anpassen der RSRP Messwerte in "Cells" an RSRP Werte in "Context"
- Cell Id → eNodeB
- eNodeB Wechsel → Response Variable Link Lifetime

time_s	rsrp_dbm	ci	scenario	provider	enodeb	drive_id	rsrp_neighbor	link_lifetime
0.06	-98	13828122	campus	o2	54016	1	-99	18.01
1.07	-101	13828122	campus	o2	54016	1	-104	17.00
2.07	-101	13828122	campus	o2	54016	1	-104	16.00
3.07	-94	13828122	campus	o2	54016	1	-100	15.00
4.07	-94	13828122	campus	o2	54016	1	-100	14.00

- Anwendung des Prädiktionsmodells XGBoost um Link Lifetime vorherzusagen
- Validierung analog zu Task I mit Zeitreihenkreuzvalidierung



T. Hastie, R. Tibshirani, and J. Friedman.

The elements of statistical learning: data mining, inference and prediction.

Springer, 2 edition, 2009.



R. Hyndman and G. Athanasopoulos.

Forecasting: principles and practice, 2018.



W. Palma.

Time Series Analysis.

John Wiley and Sons, Inc., Hoboken, New Jersey, 2016.



B. Sliwa and C. Wietfeld.

Data-driven network simulation for performance analysis of anticipatory vehicular communication systems.

IEEE Access, 7:172638–172653, 2019.

Autokorrelation der Datenübertragungsrate (Uplink)

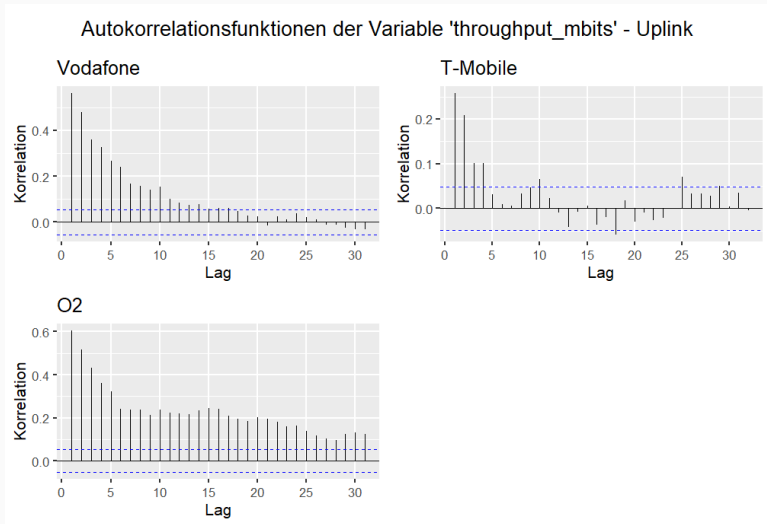


Figure 12: Autokorrelationsfunktion der Datenübertragungsrate in Richtung Uplink.

partielle Autokorrelation der Datenübertragungsrate(Uplink)

partielle Autokorrelationsfunktionen der Variable 'throughput_mbits' - Uplink

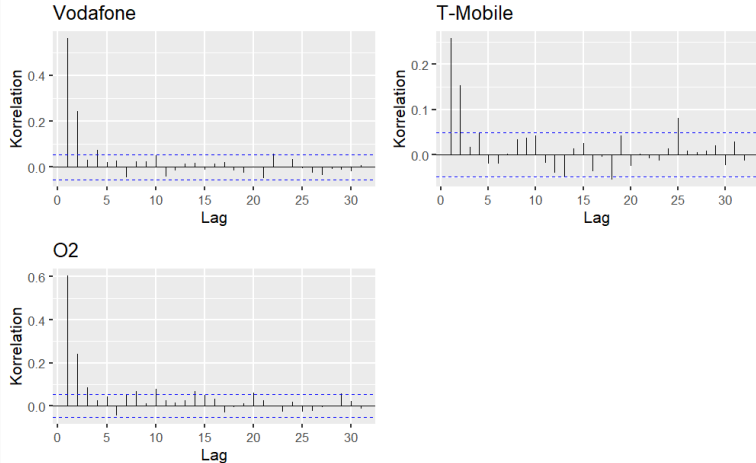


Figure 13: partielle Autokorrelationsfunktion der Datenübertragungsrate in Richtung Uplink.

Test auf Stationarität (Uplink)

Augmented Dickey-Fuller Test [4]:

H_0 : Zeitreihe hat Einheitswurzel \Rightarrow Zeitreihe ist nicht stationär

H_1 : Zeitreihe hat keine Einheitswurzel \Rightarrow Zeitreihe ist stationär

Feature	Vodafone	T-Mobile	O2
throughput_mbits	0,01	0,01	0,01
payload_mb	0,01	0,01	0,01
f_mhz	0,01	0,045	0,01
rsrp_dbm	0,01	0,01	0,01
rsrq_db	0,01	0,01	0,01
rssnr_db	0,01	0,01	0,01
cqi	0,01	0,01	0,01
ta	0,01	0,01	0,01
velocity_mps	0,01	0,01	0,01
enodeb	0,01	0,01	0,01

Figure 14: Ergebnisse des Augmented Dickey-Fuller Tests auf Stationarität für alle Variablen in Richtung Uplink.

Überprüfung der Multikollinearität (Uplink)

Varianzinflationsfaktor (VIF) [1]:

$VIF = \frac{1}{1-R_j^2}$ gibt an, um welchen Faktor die Varianz von β_j durch lineare Abhängigkeit vergrößert wird. Faustregel: $VIF < 10$

Feature	Vodafone	T-Mobile	O2
payload_mb	1,01	1,00	1,00
f_mhz	1,45	1,26	1,50
rsrp_dbm	2,65	2,02	1,81
rsrq_db	2,39	2,21	2,81
rssnr_db	2,78	2,62	3,44
cqi	2,05	1,84	2,71
ta	1,38	1,27	1,23
velocity_mps	1,13	1,27	1,21
enodeb	1,20	1,29	1,05

Figure 15: Varianzinflationsfaktor aller Einflussvariablen in Richtung Uplink.

Normalverteilung der Residuen (Uplink)

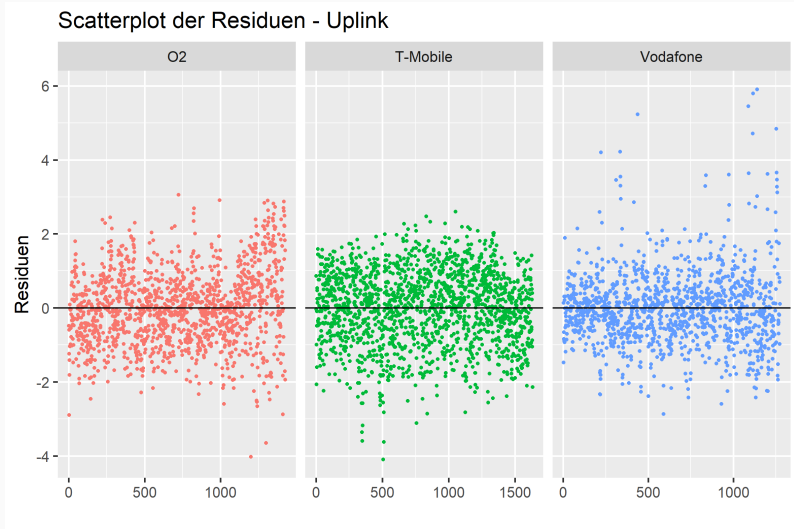


Figure 16: Scatterplots der Residuen der linearen Modelle mit Daten der Richtung Uplink.

Normalverteilung der Residuen (Uplink)

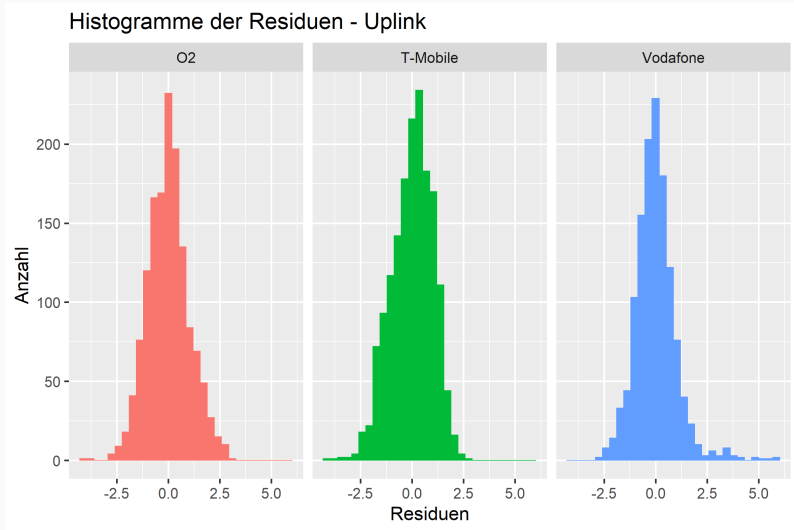


Figure 17: Histogramme der Residuen der linearen Modelle mit Daten der Richtung Uplink.

Normalverteilung der Residuen (Uplink)

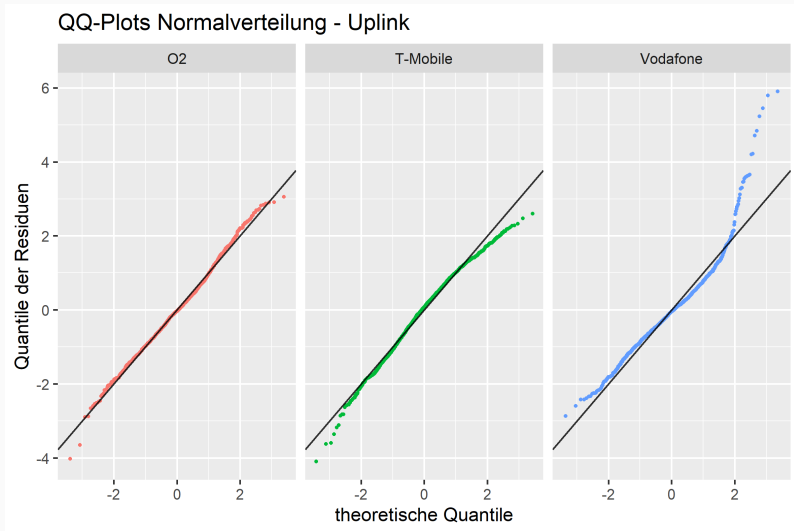


Figure 18: qq-Plots der Residuen der linearen Modelle mit Daten der Richtung Uplink.

Bestimmung des Grids für die AR-Ordnung - Downlink



Figure 19: Autokorrelationsfunktion der Residuen des linearen Modells in Richtung Downlink.

Bestimmung des Grids für die MA-Ordnung - Downlink

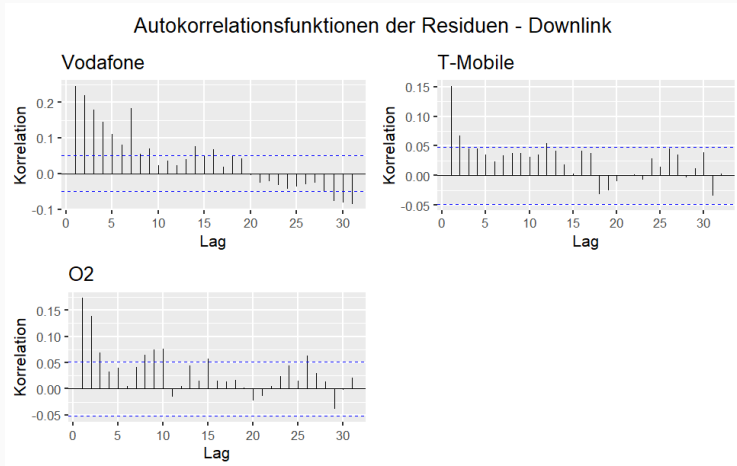


Figure 20: Autokorrelationsfunktion der Residuen des linearen Modells in Richtung Downlink.