

TECHNISCHE UNIVERSITÄT BERLIN

Anwendung von Gauß- Prozessen zur Vorhersage des Euro – Dollar Wechselkurses

Studienarbeit am Lehrstuhl „Methoden der Künstlichen
Intelligenz“ von Professor Manfred Opper

Alexander Kagoshima

26. März 2012

Die selbständige und eigenhändige Anfertigung versichere ich an Eides statt.

Berlin, den

Unterschrift

Inhalt

1	Einführung.....	4
1.1	Zielsetzung	4
1.2	Devisenmarkt.....	4
1.3	Finanzmarktdaten	5
2	Gauß-Prozesse	7
2.1	Grundlagen	7
2.2	Kovarianzfunktionen	7
2.3	Lernen mit Gauß-Prozessen.....	11
2.4	Optimierung der Hyperparameter	11
3	Vorhersage mit Gauß-Prozessen.....	13
3.1	Datenmaterial	13
3.2	Abbildung des Zeitpunkts	14
3.3	Abbildung des Zustandsvektors.....	17
4	Resultate	19
4.1	Abbildung des Zeitpunkts	19
4.1.1	1-Schritt Vorhersage.....	19
4.1.2	10-Schritt Vorhersage.....	24
4.1.3	40-Schritt Vorhersage.....	27
4.2	Abbildung des Zustandsvektors.....	31
4.2.1	1-Schritt Vorhersage.....	31
4.2.2	10-Schritt Vorhersage.....	34
4.2.3	40-Schritt Vorhersage.....	37
5	Fazit	39

Abbildungen

Abbildung 1: Darstellung eines GP Priors mit SE-Kovarianzfunktion, $\sigma^2 = 1, l$ variabel.....	8
Abbildung 2: Darstellung eines GP Priors mit RQ-Kovarianzfunktion, $\sigma^2 = 1, l = 1, \alpha$ variabel	9
Abbildung 3: Darstellung eines GP Priors mit Matérn Kovarianzfunktion, $\sigma^2 = 1, l$ und v variabel ...	10
Abbildung 4: Tägliche Schlusskurse des Euro-US Dollar Wechselkurs vom 03.01.2010 bis zum 21.10.2011	13
Abbildung 5: Veranschaulichung der Normalisierungsmethode	15
Abbildung 6: Ablaufdiagramm des Lern- und Vorhersagealgorithmus' zur Abbildung des Zeitpunkts..	16
Abbildung 7: Ablaufdiagramm des Lern- und Vorhersagealgorithmus' zur Abbildung des Zustandsvektors.....	18
Abbildung 8: MSE in Abhängigkeit der Breite des Trainingsfensters für 1-Schritt Vorhersagen	21
Abbildung 9: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 1-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.	23
Abbildung 10: MSE in Abhängigkeit der Breite des Trainingsfensters für 10-Schritt Vorhersagen	25
Abbildung 11: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 10-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.	26
Abbildung 12: MSE in Abhängigkeit der Breite des Trainingsfensters für 40-Schritt Vorhersagen	28
Abbildung 13: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 40-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.	29
Abbildung 14: Darstellung der Kauf und Verkaufssignale eines Moving-Average Crossover Systems. Die schwarzen Pfeile geben die Richtung an, die durch ein Kreuzen des Exponential-Smoothings vorhergesagt wird. Dies entspricht der Umkehrung der untersuchten Exponential-Smoothing Vorhersage.....	30
Abbildung 15: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSE für 1-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Dimension des Zustandsvektors und die Länge des Trainingsfensters angegeben.....	33
Abbildung 16: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSE für 10-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters und die Dimension des Zustandsvektors angegeben.....	35
Abbildung 17: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSEs für 40-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters und die Dimension des Zustandsvektors angegeben.....	38

1 Einführung

1.1 Zielsetzung

In der vorliegenden Studienarbeit wird untersucht, inwieweit mit Hilfe von *Gauß-Prozessen* Vorhersagen über zukünftige Schlusskurse des Euro-Dollar Wechselkurs getroffen werden können. Hierzu werden zwei unterschiedliche Ansätze verfolgt: Einerseits wird durch den Gauß-Prozess eine Abbildung des Zeitpunkts auf die Kursdaten, andererseits eine Abbildung von vergangenen täglichen Änderungen auf zukünftige tägliche Änderungen trainiert. Die Vorhersagen der Gauß-Prozesse werden mit denen verschiedener anderer Methoden verglichen. Hierzu werden *Exponential Smoothings*, *ARIMA-Modelle* und *Neuronale Netze*, sowie auch eine *naïve* Vorhersage genutzt.

In Abschnitt 1 werden kurz ökonomische Hintergründe des Devisenmarkts erläutert und eine kurze Einführung gegeben, wie Finanzmarktkurse häufig in den Wirtschaftswissenschaften betrachtet werden. Abschnitt 2 gibt anschließend eine kurze Einführung in die Theorie der Gauß-Prozesse, was zum Verständnis der folgenden Kapitel wichtig ist. Die Vorhersagemethodik wird in Abschnitt 3 ausführlich erklärt und die entstandenen Resultate dieser Vorhersagemethodik in Abschnitt 4 erläutert. Eine kurze Zusammenfassung der Erkenntnisse dieser Studienarbeit befindet sich in Abschnitt 5.

1.2 Devisenmarkt

Der Euro-Dollar Wechselkurs (auch Dollarkurs) ist die Menge an Dollar, die für einen Euro bezahlt werden muss oder der Preis eines Euros in Dollar. Die Wechselkurse der Währungen entstehen am Devisenmarkt (auch *Forex-Markt*, Abk. für „*Foreign Exchange*“), welcher im Gegensatz zu Aktien- oder Derivatemarkten über keinen zentralen Börsenplatz verfügt, sondern weitgehend zwischen den Marktteilnehmern direkt abgewickelt wird. Für den Handel zwischen den größten Marktteilnehmern, den Banken, hat sich der Begriff *Interbankenmarkt* entwickelt. Am Interbankenmarkt werden die größten Devisentransaktionen durchgeführt, weshalb er de facto als zentral für den Devisenmarkt angesehen werden kann.

Um auch kleineren Marktteilnehmern die Möglichkeit zu geben, sich gegen Devisenkursschwankungen abzusichern und auf Devisenkurse zu spekulieren, gewähren viele Banken als *Forex-Broker* Zugang zum Devisenmarkt. Meist treten die Banken hier als *Market Maker* auf, was bedeutet, dass bei Abschluss eines Handels durch den Kunden kein anderer Kunde die entsprechende Gegentransaktion durchführt, sondern die Bank diese übernimmt und die kumulierten entstandenen Risiken am Interbankenmarkt absichert. Ein *Market Maker* übernimmt aber nicht nur die Gegentransaktion für den Kunden, er stellt auch den handelbaren Preis.

Da dieses Geschäftsmodell zu einem Interessenkonflikt seitens der Bank führt, werben viele *Forex-Broker* damit, als sogenannte *ECN-Broker* (*Electronic Communication Network*) zu agieren. Sie bieten Zugang zu einem elektronischen Handelssystem in dem eine große Zahl an anderen Banken und Händlern eine hohe Liquidität garantieren sollen. In diesem Fall werden die Transaktionen wie auf dem Interbankenmarkt zwischen den Kunden abgewickelt und der *Forex-Broker* tritt nur als Mittler auf.

Da weltweit verschiedene, konkurrierende *ECNs* existieren und der Interbankenmarkt über keinen zentralen Marktplatz verfügt, existieren keine „offiziellen“ Devisenkurse. Die Devisenkurse der einzelnen *ECNs*, dem Interbankenmarkt und der *Market Maker* können sich kurzfristig voneinander unterscheiden, jedoch werden diese Unterschiede durch Arbitragetätigkeiten auf den jeweiligen Märkten minimiert.

Auf den weltweiten Devisenmärkten wird hauptsächlich mit der weltweiten Leitwährung, dem US-Dollar, gehandelt. Dies resultiert aus der Tatsache, dass alle Währungen bis zum Zusammenbruch des *Bretton-Woods-Systems* an den Dollar gekoppelt waren und sich diese Tradition bis heute auf den Devisenmärkten gehalten hat. Es ist auch bis heute üblich, dass eine Transaktion mit einem Devisenpaar, das den US-Dollar nicht enthält, wie beispielsweise der Euro-Yen Wechselkurs, im Hintergrund über zwei Transaktionen der Devisenpaare Euro-Dollar und Dollar-Yen abgewickelt wird.

Von den US-Dollar Devisenpaaren ist der Euro-Dollar Wechselkurs das meistgehandelte Währungspaar überhaupt [1]. Dies liegt an den weitreichenden transatlantischen Handels- und Investitionsbeziehungen, welche die umfangreichsten bilateralen Wirtschaftsbeziehungen weltweit darstellen: Das Handelsvolumen an Waren und Dienstleistungen 2010 betrug 668 Mrd. € während die wechselseitigen Investitionsbestände 2009 bei über 2,1 Bio. € lagen und im selben Jahr sogar noch um 180 Mrd. € stiegen [2]. Diese weitreichenden Verflechtungen sorgen für eine sehr hohe Liquidität in diesem Devisenpaar, im April 2010 wurden täglich im Durchschnitt 1,1 Bio. € Umsatz allein im Euro-Dollar Währungspaar erreicht [1]. Aus diesen Gründen ist der Devisenmarkt mit insgesamt 4 Bio. € Umsatz im täglichen Mittel der mit Abstand größte Finanzmarkt der Welt.

1.3 Finanzmarktdaten

Finanzmarktdaten sind aufgrund ihrer guten Verfügbarkeit häufig Gegenstand von Forschung in den Wirtschaftswissenschaften sehr. Bereits um 1900 versuchte Louis Bachelier Kursbewegungen stochastisch auszudrücken [3] und ist damit einer der Begründer der Finanzmathematik. Traditionell bedient sich diese häufig an physikalischen Modellen, so basiert die Optionsbewertungsmethode nach Black-Scholes von 1973 [4] auf der Annahme, dass Kursbewegungen durch einen *Wiener-Prozess* modelliert werden können, der auch zur Modellierung der Bewegung von Elementarteilchen genutzt wird.

Die Modellierung durch den Wiener-Prozess ist in der Ökonomie heute immer noch weit verbreitet. Sie impliziert, dass die Kursänderungen normalverteilt und unabhängig voneinander sind (*random walk hypothesis*). In der Ökonomie wurde die Annahme dieser unabhängigen Richtungsänderungen mit der Theorie der *effizienten* Märkte begründet. Effizienz bedeutet in diesem Zusammenhang, dass neue Informationen durch perfekt informierte und rationale Investoren augenblicklich in den Kurs eingepreist werden [5]. Hierdurch sind Kursbewegungen nur von den zufälligen Ereignissen in der Welt beeinflusst und damit nicht vorhersagbar. Hieraus folgt, dass beim Spekulieren am Finanzmarkt eine Rendite über die Markttrendite hinaus nur möglich ist, wenn auch ein höheres Risiko eingegangen wird, da am Finanzmarkt die jeweiligen Finanzprodukte immer korrekt bepreist sind und der Spekulant somit keinen Vorteil gegenüber anderen erlangen kann.

Die *random walk hypothesis* wurden durch die Arbeiten von Benoît Mandelbrot und anderen [6] in zahlreichen Untersuchungen widerlegt. Mandelbrot fand in seiner Forschung, dass die Kursänderungen keiner Normalverteilung unterliegen können, da extreme Kursänderungen wie der Börsencrash von 1987 [7] unter dieser Annahme zu häufig auftreten. Er unterstellt für die Kursänderungen eine *Lévy-stabile* Verteilung mit $\alpha < 2$, eine Verteilung mit „*fat tails*“. Als ein weiteres charakteristisches Merkmal fand Mandelbrot, dass in den Kursänderungen eine langfristige positive Korrelation besteht, dass in den Kursen also beständige *Trends* existieren. Ebenso konnte beobachtet werden, dass die Varianz der Kursänderungen nicht konstant zu sein scheint, sondern Cluster bildet, in denen die Varianz hoch, und ausgedehnte Phasen, in denen die Varianz gering ist.

Auch im Umfeld des maschinellen Lernens finden sich zahlreiche Ansätze, Methoden des maschinellen Lernens auf Finanzmärkte anzuwenden. Dies beschränkt sich nicht nur auf die

Vorhersage von Wechselkursen oder anderen Finanzmarktdaten (siehe[8], [9]), sondern erstreckt sich auf viele wichtige Bereiche des Finanzwesens, wie beispielsweise die Bepreisung von Derivaten und Optionen [10], die Modellierung von Risiken bestimmter Finanzprodukte[11], oder das optimale Hedging offener Positionen[12]. Zur Vorhersage von Finanzmarktkursen sei angemerkt, dass viele Ansätze aus der Literatur sich ausschließlich auf die Vorhersagbarkeit von Aktienkursen beziehen. Nach allgemeinem Verständnis dürfte dies einfacher sein als die Vorhersage eines Wechselkurses, da Aktienmärkte über lange Perioden generell in Richtung höherer Kursniveaus streben.

2 Gauß-Prozesse

2.1 Grundlagen

In diesem Abschnitt sollen die Grundlagen von *Gauß-Prozessen* nur kurz rekapituliert werden, für eine ausführlichere Beschreibung soll hier auf [13], Kapitel 2 und 3, verwiesen sein.

Ein stochastischer Prozess $f(x)$ ist ein Gauß-Prozess mit der Mittelwertfunktion $m(x)$ und der Kovarianzfunktion $k(x, x')$, wenn die Werte einer endlichen Untermenge an Punkten $f(x_1) \dots f(x_n)$ die Komponenten eines multivariat normalverteilten Vektors mit Mittelwert $\mathbf{m} = (m(x_1) \dots m(x_n))^T$ und Kovarianzmatrix $\Sigma_{i,j} = k(x_i, x_j)$ mit $1 \leq i, j \leq n$ bilden. Dies kann geschrieben werden als

$$f \sim GP(\mathbf{m}, \Sigma_{i,j}).$$

Einfacher ausgedrückt definiert ein Gauß-Prozess eine multivariate Normalverteilung mit Korrelationen über eine Funktion. GPs sind als Verfahren im maschinellen Lernen nützlich, da ein GP mit einer gegebenen Kovarianzfunktion, welche seine Charakteristik kontrolliert, auch als *Prior*-Vermutung über die Eigenschaften einer unbekannteren Funktion verstanden werden kann. Diese *Prior*-Vermutung kann anhand von Daten effizient konditioniert werden, wodurch eine *Posterior* Verteilung entsteht, die zur Vorhersage unbekannter Datenpunkte genutzt werden kann. Gauß-Prozesse bieten damit ein komplettes Bayes'sches Framework für Inferenz auf Funktionen.

2.2 Kovarianzfunktionen

Die Kovarianzfunktion stellt ein Skalarprodukt in einem hochdimensionalen Merkmalsraum (*feature space*) dar. Jedoch wird dieses Skalarprodukt nicht im hochdimensionalen Raum berechnet, sondern nur implizit im niedrigdimensionalen Raum der Input-Daten, dies ist der sogenannte *kernel trick* [14].

Der Merkmalsraum ist abhängig von der verwendeten Kovarianzfunktion und ist potentiell unendlich-dimensional, was mit einer hohen Flexibilität von GPs zur Anpassung an vorhandene Daten einhergeht. Grundsätzlich kann jede *positiv definite* Funktion eine Kovarianzfunktion sein. Dies stellt sicher, dass auch die resultierende Kovarianzmatrix positiv definit ist.

Bei der Anwendung von Gauß-Prozessen auf einen Datensatz spiegeln sich die Annahmen über den zugrundeliegenden Prozess in der Kovarianzfunktion und den gewählten *Hyperparametern* wider. Hyperparameter sind die Parameter der Kovarianzfunktion. Wegen ihrer leichten Interpretierbarkeit werden für GPs häufig *stationäre* Kovarianzfunktionen verwendet, dies sind Kovarianzfunktionen, bei denen der Funktionswert nur vom Abstand der Datenpunkte abhängt.

Wenn beispielsweise *a priori* nur „sanfte“ Veränderungen des Prozesses vermutet werden, oder anders ausgedrückt, die Funktion unendlich oft differenzierbar sein soll, bietet sich die *Squared Exponential* Kovarianzfunktion cov_{SE} an. Hyperparameter sind bei dieser Kovarianzfunktion $\Theta = \{\sigma_f^2, l\}$. l steht für die sogenannte *length-scale*, welche bestimmt, ab welchem Abstand in den Input-Daten große Veränderungen in den Output-Daten möglich sind. σ_f^2 bestimmt die Varianz der Funktion.

$$\text{cov}_{\text{SE}}(x_i, x_j) = \sigma_f^2 * \exp\left(-\frac{|x_i - x_j|^2}{2l^2}\right)$$

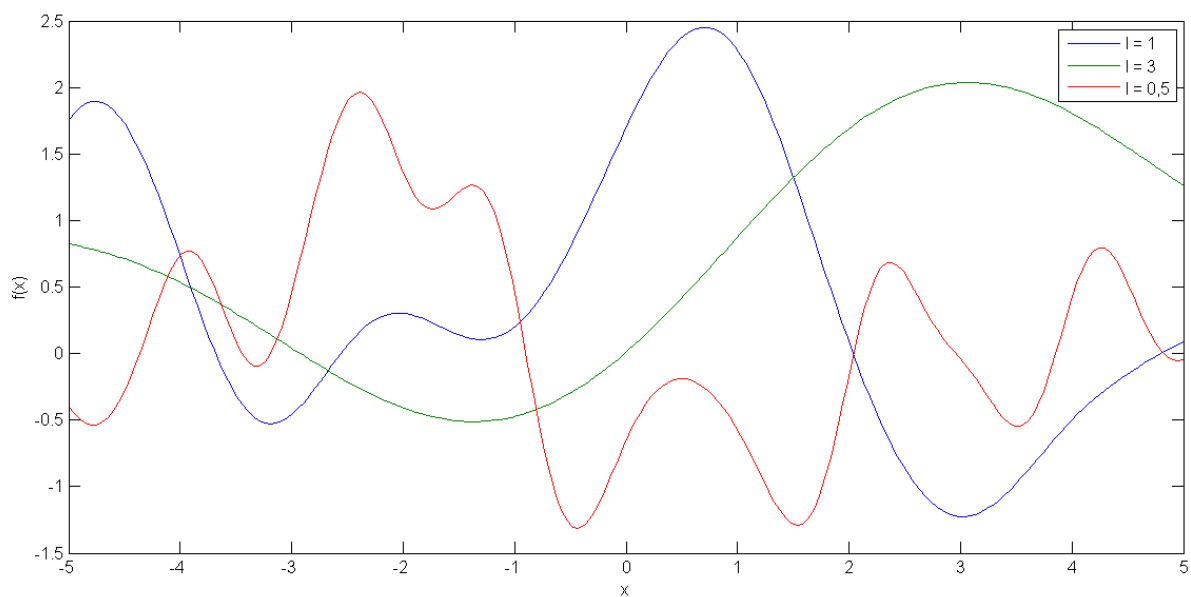


Abbildung 1: Darstellung eines GP Priors mit SE-Kovarianzfunktion, $\sigma_f^2 = 1, l$ variabel

Da für diese Kovarianzfunktion jedoch die Korrelationen für größer werdende Abstände der Input-Daten schnell verschwinden, können langfristige Abhängigkeiten mit ihr nicht modelliert werden. Um auch längerfristige Abhängigkeiten in den Daten auszudrücken ist die *Rational-Quadratic* Kovarianzfunktion cov_{RQ} mit $\Theta = \{\sigma_f^2, l, \alpha\}$ besser geeignet. σ_f^2 und l haben ähnliche Funktionen wie bei der Squared-Exponential Kovarianzfunktion. α kontrolliert, wie schnell die Korrelationen zwischen unterschiedlichen Punkten verschwinden. Wird es klein, können sehr langfristige Korrelationen modelliert werden.

$$\text{cov}_{\text{RQ}}(x_i, x_j) = \sigma_f^2 * \left(1 - \frac{|x_i - x_j|^2}{2\alpha l^2}\right)^{-\alpha}$$

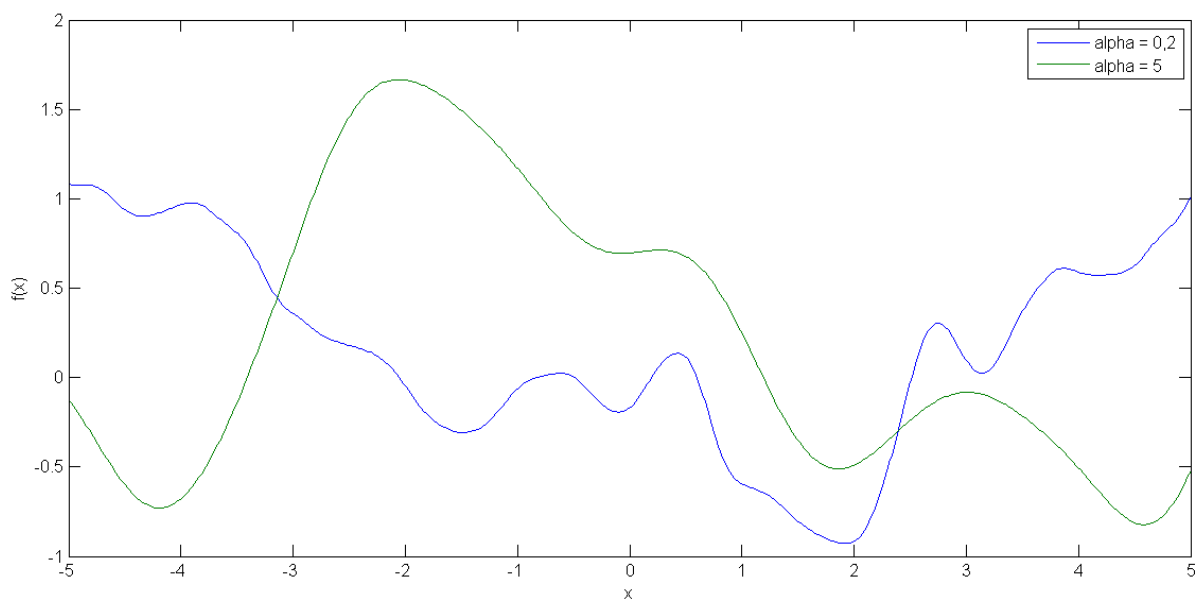


Abbildung 2: Darstellung eines GP Priors mit RQ-Kovarianzfunktion, $\sigma_f^2 = 1, l = 1, \alpha$ variabel

Da „sanfte“ Veränderungen nicht für jede Art von Daten passend sind, bietet sich für Prozesse mit abrupten Änderungen die *Matérn* Kovarianzfunktion cov_{Mat} an. Der Ausdruck dieser Kovarianzfunktion wird vereinfacht, wenn der Parameter $\nu = p + \frac{1}{2}$ ist, mit $p \in \mathbb{N}$. Die Hyperparameter, wenn ν gegeben ist, sind $\Theta = \{\sigma_f^2, l\}$. σ_f^2 und l kontrollieren dieselben Charakteristika wie bei der SE Kovarianzfunktion, der Parameter ν kontrolliert die Differenzierbarkeit der Funktion, geht $\nu \rightarrow \infty$ so wird die Matérn-Kovarianzfunktion zur Squared-Exponential Kovarianzfunktion. Für $\nu = \frac{1}{2}$ entsteht ein Ornstein-Uhlenbeck Prozess.

$$\text{cov}_{\text{Mat}, \nu=p+\frac{1}{2}}(x_i, x_j) = \exp\left(-\frac{\sqrt{2\nu} * |x_i - x_j|}{l}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^p \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu} * |x_i - x_j|}{l}\right)^2$$

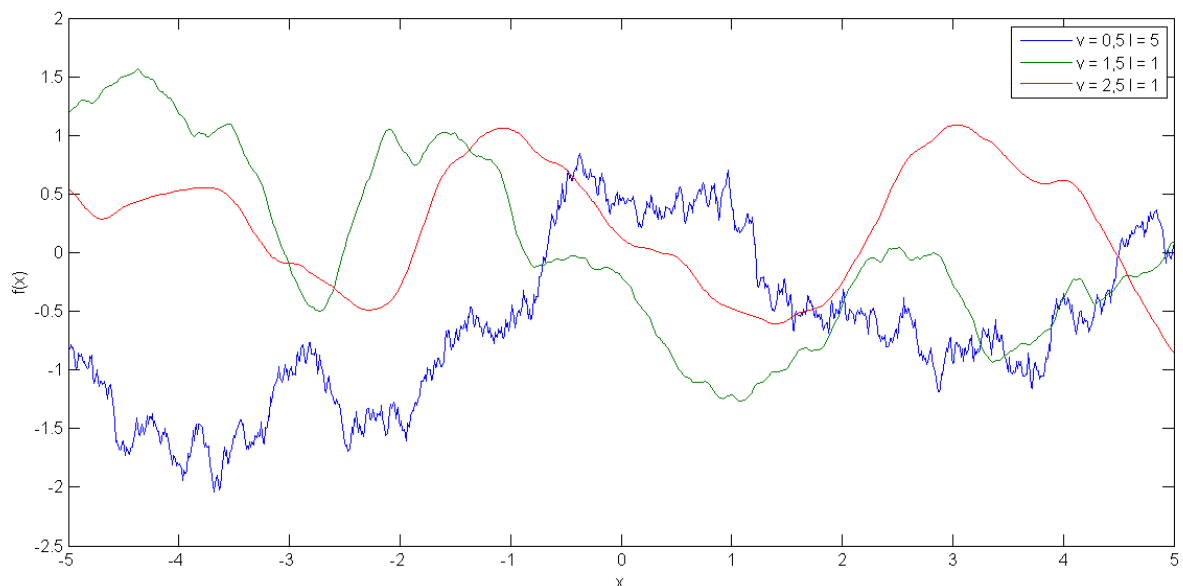


Abbildung 3: Darstellung eines GP Priors mit Matérn Kovarianzfunktion, $\sigma_f^2 = 1, l$ und ν variabel

Kovarianzfunktionen lassen sich zu komplexeren Kovarianzfunktionen kombinieren, da beispielsweise Summen und Produkte von Kovarianzfunktionen ebenso wieder Kovarianzfunktionen sind, siehe [13], Kapitel 4 (siehe dort ebenso für eine weitergehende Übersicht zu Kovarianzfunktionen)

In den vorhergehenden Abbildungen zu Kovarianzfunktionen wurde jeweils die Mittelwertsfunktion (auch *Mean-Funktion*) $m(x) = 0$ gesetzt, es sind jedoch zur Modellierung mit GPs alle denkbaren Mittelwertsfunktionen möglich. Diese werden jedoch beim Einsatz von Gauß-Prozessen häufig gleich Null gesetzt. Die Vorhersage funktioniert trotzdem recht gut, wenn die Kovarianzfunktion auch außerhalb des *Stützbereichs*¹ genügend Korrelationen mit den beobachteten Daten liefert um die Zeitreihe vorherzusagen, ohne die Mittelwertsfunktion zu nutzen.

¹ Der Bereich des Input-Raums, für den Daten beobachtet wurden

2.3 Lernen mit Gauß-Prozessen

Um Gauß-Prozesse als ein Verfahren des maschinellen Lernens zu verwenden, wird ein GP mit gegebener Kovarianzfunktion als *Prior* über die Verteilung einer Funktion genutzt. Um diese und die folgenden Darstellungen zu vereinfachen wurde im Folgenden stets $m(x) = 0$ angenommen. Diese Priorverteilung über die (unbeobachtbaren) Funktionswerte $\mathbf{f}^* = (f(x_1) \dots f(x_n))^T$ für zukünftige Input-Daten $\mathbf{X}^* = (x_1 \dots x_n)^T$ sei gemäß den Grundlagen über Gauß-Prozesse aus 2.1 gegeben durch

$$\mathbf{f}^* \sim N(\mathbf{0}, K(\mathbf{X}^*, \mathbf{X}^*)).$$

$K(\mathbf{X}^*, \mathbf{X}^*)$ entspricht hier der Matrix der Kovarianzen zwischen den Punkten \mathbf{X}^* , gegeben durch die Kovarianzfunktion $k(x, x')$.

Diese Prior-Verteilung basiert nicht auf bereits beobachteten Daten, enthält aber durch die Spezifikation der Kovarianzfunktion bereits Annahmen über den zugrundeliegenden Prozess. Wenn für die beobachteten Output-Daten $\mathbf{y} = (y_1 \dots y_n)^T$ (zugehörig zu den Input-Daten \mathbf{X}) ein zusätzliches Rauschen ε angenommen wird, so dass

$$\mathbf{y} = \mathbf{f}(\mathbf{X}) + \varepsilon, \text{ mit } \varepsilon \sim N(\mathbf{0}, \sigma_n^2 * \mathbf{I})$$

resultiert dies in der gemeinsamen Verteilung von beobachteten und zukünftigen Daten

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 * \mathbf{I} & K(\mathbf{X}, \mathbf{X}^*) \\ K(\mathbf{X}^*, \mathbf{X}) & K(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix}\right).$$

Um den *GP Posterior* zu erhalten, wird der Prior auf den gegebenen Beobachtungen konditioniert. Damit entsteht folgende Posterior-Verteilung:

$$\begin{aligned} \mathbf{f}^* | \mathbf{X}^*, \mathbf{y}, \mathbf{X} &\sim N(\bar{\mathbf{f}}^*, \text{cov}(\mathbf{f}^*)) \\ \text{cov}(\mathbf{f}^*) &= K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 * \mathbf{I})^{-1}K(\mathbf{X}, \mathbf{X}^*) \\ \bar{\mathbf{f}}^* &= K(\mathbf{X}^*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 * \mathbf{I})^{-1}\mathbf{y} \end{aligned}$$

Falls statt der Verteilung der Funktionswerte \mathbf{f}^* die Verteilung der Beobachtungen \mathbf{y}^* benötigt wird, gilt:

$$\begin{aligned} \text{cov}(\mathbf{y}^*) &= \text{cov}(\mathbf{f}^*) + \sigma_n^2 * \mathbf{I} \\ \bar{\mathbf{y}}^* &= \bar{\mathbf{f}}^* \end{aligned}$$

Mit Hilfe dieser Verteilung können nun probabilistische Vorhersagen für neue Datenpunkte \mathbf{X}^* mit Hilfe gegebener Datenpunkte \mathbf{X} und \mathbf{y} getroffen werden.

2.4 Optimierung der Hyperparameter

Der GP ist ein parameterfreies Verfahren, da über die Parameter integriert wird, vergleiche hierzu [13], Kapitel 2. Jedoch wird das Verhalten des GP wesentlich durch die Wahl und Struktur der Kovarianzfunktion und ihrer Hyperparameter gesteuert. Das *Trainieren* beziehungsweise *Lernen* eines GP entspricht daher dem Anpassen der Hyperparameter auf einen Datensatz.

Da bei GPs innerhalb eines Bayes'schen Frameworks gearbeitet wird, ist es grundsätzlich möglich ein hierarchisches Bayes-Modell zu erstellen, in dem über die Hyperparameter integriert wird, so dass ihre Optimierung nicht mehr nötig ist. Jedoch müssen dafür meist

Integrale berechnet werden, die analytisch nicht lösbar sind. Um Approximationen und *Markov-Chain-Monte-Carlo* Methoden zur Lösung dieser Integrale zu vermeiden, kann stattdessen zur Optimierung der Hyperparameter die *Marginal Likelihood* maximiert werden.

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{X}, \boldsymbol{\theta})p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta})d\mathbf{f}$$

Diese Likelihood wird marginal genannt, da über die latente Funktion \mathbf{f} integriert wird. Da beide Verteilungen im Integral Gauss-Verteilungen sind, ist die Berechnung der Marginal Likelihood analytisch möglich. Es ergibt sich

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}_y| - \frac{n}{2} \log(2\pi).$$

Da die Gradienten der Marginal Likelihood in Abhängigkeit der Hyperparameter ebenso analytisch berechenbar sind, bietet sich zur Optimierung der Hyperparameter eine (vom Rechenaufwand effiziente) Maximierung der Marginal Likelihood mittels einer gradientenbasierten Methode an. Dieses Verfahren wurde auch in dieser Studienarbeit angewendet.

Bei vielen Methoden des maschinellen Lernens muss bei der Optimierung der Parameter speziell darauf geachtet werden, dass ein Modell nicht zu stark an die Daten angepasst wird. In einem solchen Fall von *Overfitting* lernt das Modell das Rauschen eines speziellen Datensatzes mit, weshalb es für unbekannte Input-Daten nicht mehr gut generalisiert. Dies zeigt sich in einem sehr niedrigen *In-Sample* Fehler und einem hohen *Out-of-Sample* Fehler und wird meist durch eine *Cross-Validation* zu verhindern versucht.

Cross-Validation bedeutet jedoch in den meisten Fällen einen sehr hohen Rechenaufwand, da für eine *k-fold Cross-Validation* in der Regel k Modelle trainiert werden müssen. Für manche Methoden wie beispielsweise *Kernel Ridge Regression* ist ein geschlossener Ausdruck für den *Leave-One-Out Cross-Validation* Fehler möglich, der eine effizientere Implementierung von Cross-Validation ermöglicht. Dies ist bei Gauß-Prozessen ebenso möglich, jedoch wird bei einer Maximierung der *Marginal Likelihood* ein Overfitting der Daten automatisch vermieden. Um dies zu verstehen ist es sinnvoll zuerst die drei Summanden der Marginal Likelihood zu interpretieren:

- $-\frac{1}{2}\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y}$ bestraft eine schlechte Anpassung an die Daten.
- $-\frac{1}{2} \log |\mathbf{K}_y|$ bestraft hohe Komplexität der Kovarianzstruktur.
- $-\frac{n}{2} \log(2\pi)$ ist eine Normalisierungskonstante.

Hieran wird deutlich, dass eine sehr gute Anpassung an die Daten eine hohe Marginal Likelihood also nur dann erzeugt, wenn gleichzeitig die Komplexität der verwendeten Kovarianzfunktion nicht zu groß wird. Eine hohe Komplexität der Kovarianzstruktur bedeutet, dass beliebig komplexe Funktionen perfekt angepasst, also overfitted werden können. Somit wird ein Overfitting durch die Optimierung der Hyperparameter bei der Maximierung der Marginal Likelihood automatisch vermieden.

3 Vorhersage mit Gauß-Prozessen

3.1 Datenmaterial

Das Datenmaterial für diese Studienarbeit stammt vom Schweizer Forex-Broker Dukascopy², der Zugang zu einem hochliquiden ECN bietet und aufgrunddessen qualitativ hochwertige Daten des Devisenmarkts liefern kann.

Für die Vorhersage der täglichen Schlusskurse wurden dabei die TagesSchlusskurse vom 03.01.2000 bis zum 21.10.2011 genutzt, was einem Umfang von ungefähr 3000 Datenpunkten entspricht (an Wochenenden wird nicht gehandelt).

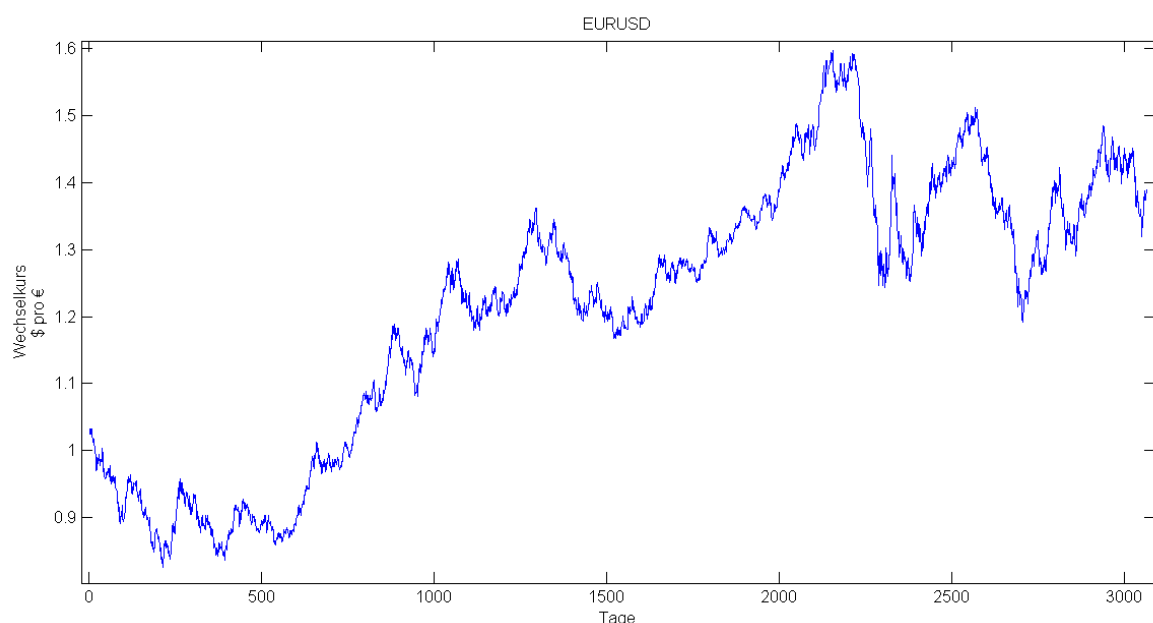


Abbildung 4: Tägliche Schlusskurse des Euro-US Dollar Wechselkurs vom 03.01.2010 bis zum 21.10.2011

Abbildung 4 zeigt die täglichen Schlusskurse Y . Augenscheinlich hat dieser Datensatz eine hohe Zufallskomponente, denn er ähnelt stark einem *random walk*. Jedoch ist auch die Ähnlichkeit mit einem Prior der Matérn-Kovarianzfunktion (**Abbildung 3**) sehr groß, was wiederum gegen einen *random walk* spräche.

Falls von einer hinreichend großen Nicht-Zufälligkeit ausgegangen wird, könnten abschnittsweise Trends in der Zeitreihe vorliegen. Von Tag 500 bis Tag 2100 herrscht ein starker Aufwärtstrend, anschließend ein leichter Abwärtstrend. Weiterhin fallen bei genauerem Hinsehen möglicherweise periodische Schwankungen auf, sowie die Konzentration von starken Kursschwankungen auf kurze Zeitabschnitte (siehe 1.3).

² www.dukascopy.com

3.2 Abbildung des Zeitpunkts

Zur Vorhersage der täglichen Schlusskurse $\mathbf{Y} = \{y_t\}_{t=1}^T$ kann durch den GP eine Abbildung

$$y_t = f(t)$$

vom Zeitpunkt t auf den Datenpunkt y_t gelernt werden. Durch das Lernen dieser einfachen Abbildung ist die Interpretation der Hyperparameter sowie das Erstellen einer passenden Kovarianzfunktion relativ einfach. Ein gutes Beispiel hierfür ist [13], Kapitel 5, wo beschrieben wird, wie die unterschiedlichen Komponenten einer Zeitreihe identifiziert und durch eigene Komponenten in der Kovarianzfunktion modelliert wurden.

Zur Vorhersage des Euro – US Dollar Wechselkurses stellt sich die Frage, welche Kovarianzfunktion zur Vorhersage am besten geeignet ist. Da die Daten in **Abbildung 4** eine große Ähnlichkeit mit einem GP Prior mit Matérn-Kovarianz besitzen (**Abbildung 3**), könnte diese auch am besten zur Modellierung in einem GP geeignet sein. Um diese Frage jedoch näher zu untersuchen, werden verschiedene Kovarianzfunktionen sowie Kombinationen³ davon miteinander verglichen.

Offen ist auch, welche Mittelwertfunktion verwendet werden sollte. In dieser Studienarbeit wird auch diese Fragestellung durch den Vergleich verschiedener Varianten untersucht. Die erste Variante ist hierbei, die Mittelwertfunktion gleich Null zu setzen. Obwohl diese Herangehensweise möglicherweise zu simpel scheint, kann sie dennoch gute Ergebnisse liefern, siehe 2.2.

Falls jedoch eine Mittelwertfunktion verwendet werden soll, kommt als zweite Variante eine *lineare Mittelwertfunktion* in Betracht, die den Trend aus den Trainingsdaten fortführen soll. Die Parameter der Mittelwertfunktion können ebenfalls durch die Maximierung der Marginal Likelihood gelernt werden.

Da die Marginal Likelihood durch die zusätzlichen Hyperparameter einer Mittelwertfunktion jedoch eine komplexere Oberfläche bekommt, die durch einen Gradientenabstieg häufig schwerer zu minimieren ist, wird auch ein dritter Ansatz getestet, bei dem die Mittelwertfunktion $m(x) = 0$ gesetzt wird und stattdessen die Trainingsdaten aus einem Trainingsfenster (zum Trainingsfenster, siehe weiter unten) mit der Breite n_{train} $\mathbf{y}_{t_{start}:t_{ende}} \in (n_{train} \times 1)$ vorverarbeitet werden indem ein *linearer Trend* abgezogen wird:

$$\mathbf{y}'_{t_{start}:t_{ende}} = \mathbf{y}_{t_{start}:t_{ende}} - \left(y_{t_{start}} + \frac{y_{t_{ende}} - y_{t_{start}}}{n_{train}} * \begin{bmatrix} 1 \\ \vdots \\ n_{train} \end{bmatrix} \right)$$

$$n_{train} = ende - start + 1$$

Ein solches mit dieser Methode normalisierte Trainingsfenster ist in **Abbildung 5** zusätzlich veranschaulicht. Der lineare Trend wird bei der Vorhersage wieder entsprechend addiert.

Siehe 2.2, Kovarianzfunktionen können durch verschiedene mathematische Operationen kombiniert werden, um neue Kovarianzfunktionen zu erzeugen³

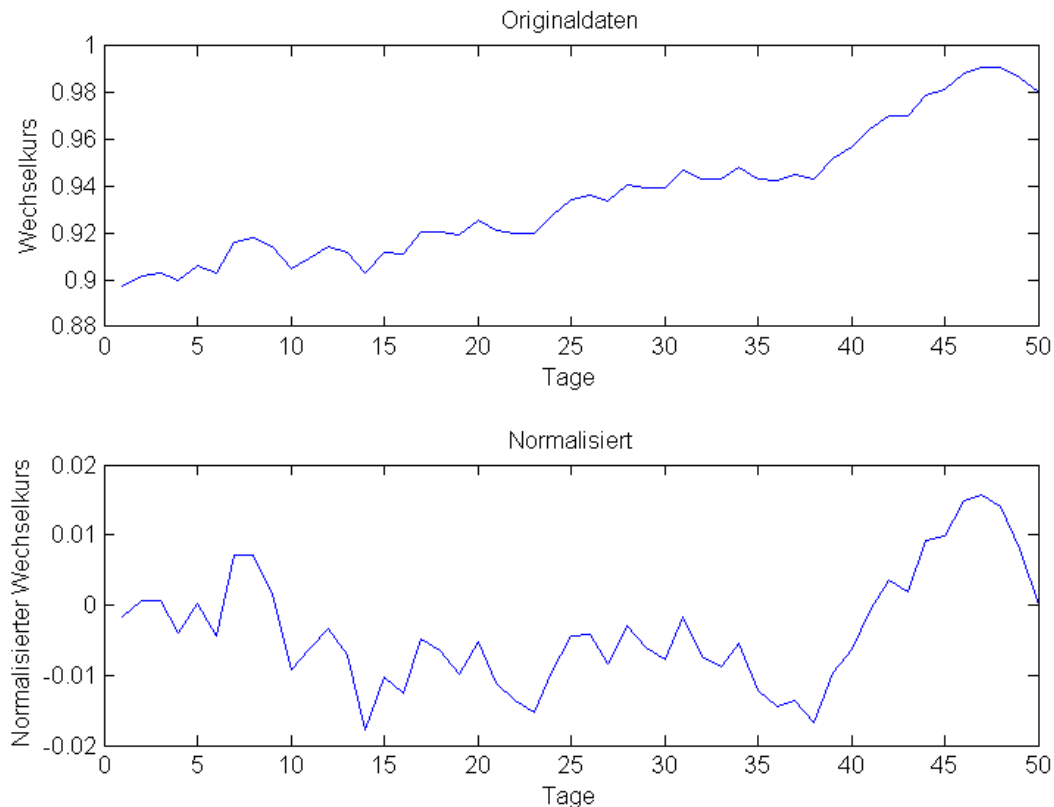


Abbildung 5: Veranschaulichung der Normalisierungsmethode

Die Vorhersagen der unterschiedlichen Ansätze werden anhand von 1, 10 und 40-Schritt Vorhersagen verglichen. Dies entspricht einer Vorhersage um einen Handelstag, zwei Handelswochen und ungefähr zwei Handelsmonaten.

Der GP wird nur auf Daten eines *Trainingsfensters* trainiert, da ältere Datenpunkte für die Vorhersage aufgrund veränderlicher Charakteristika des Wechselkurses wahrscheinlich nicht mehr relevant sind und so der Rechenaufwand reduziert werden kann. Die Breite dieses Fensters wird bei den Versuchen zwischen 50, 100, 150 und 200 Punkten variiert. Der vollständige Lern- und Vorhersagealgorithmus ist in einem Ablaufdiagramm (**Abbildung 6**) skizziert.

Als erster Benchmark für diese Vorhersagemethode wird ein *Exponential-Smoothing State-Space-Model* aus dem „forecast“ Package für *R* genutzt [15]. Dabei werden für jedes Trainingsfenster verschiedene Exponential Smoothing Modelle (gegebenenfalls werden Trend- und Saisonalitätsparameter genutzt) anhand des AIC verglichen und das beste Modell zur Vorhersage ausgewählt.

Als zweiter Benchmark werden klassische ARIMA Modelle genutzt, die ebenfalls im „forecast“ Package für *R* implementiert sind. Der Grad des genutzten ARIMA Modells wird durch Unit-Root-Tests für jedes Trainingsfenster neu bestimmt und die möglichen Modelle anhand eines spezialisierten Suchalgorithmus ausgewählt und miteinander verglichen. Auf diese Weise wird für jedes Trainingsfenster das beste Modell ausgewählt und damit vorhergesagt.

Schlussendlich wurde ein „naiver“ Benchmark zum Vergleich herangezogen, der für den nächsten Tag den Kurs des heutigen Tages vorhersagt. Angenommen der Wechselkurs sei ein *random walk*, so müsste diese Vorhersage die besten Ergebnisse liefern.

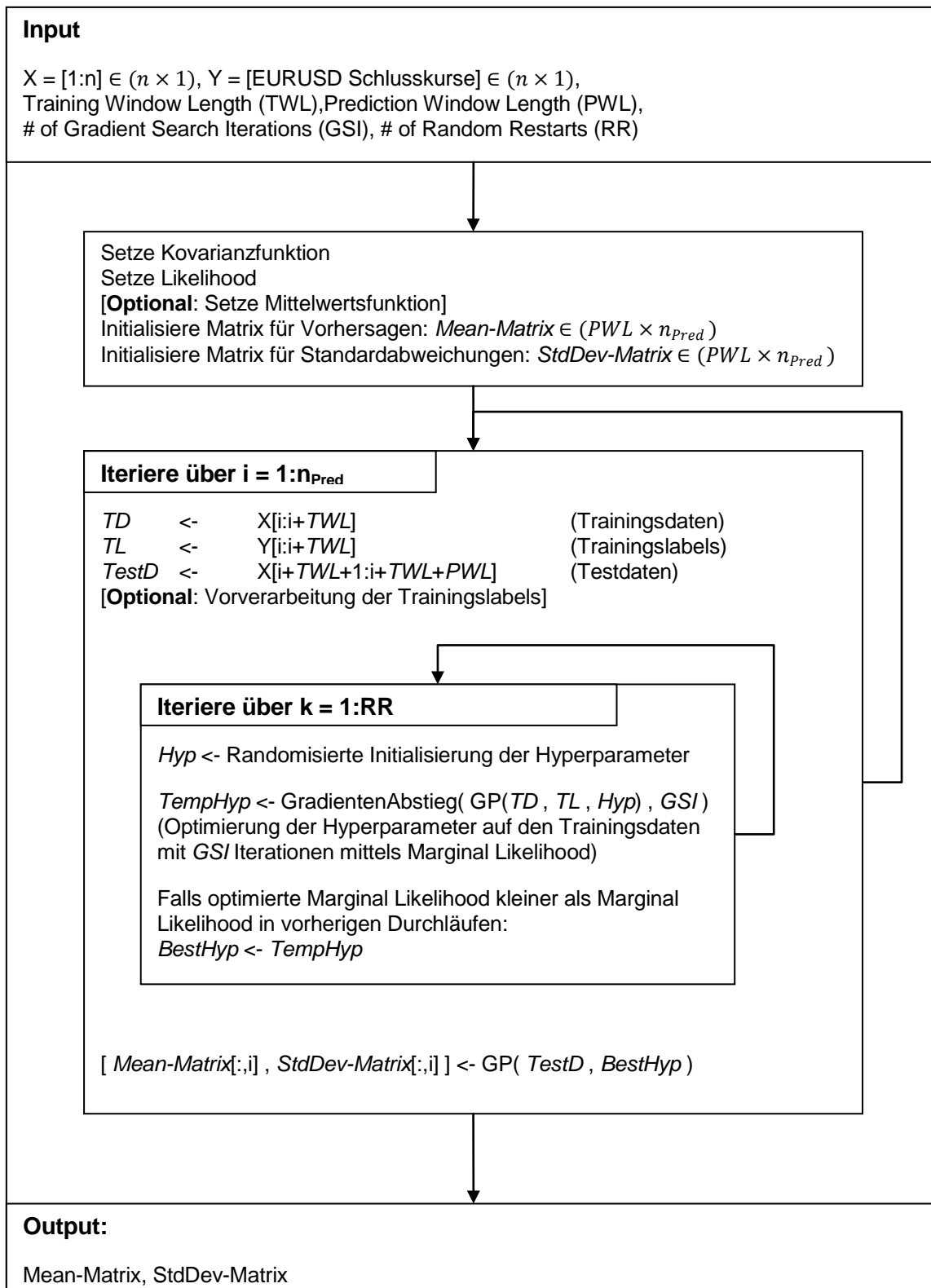


Abbildung 6: Ablaufdiagramm des Lern- und Vorhersagealgorithmus' zur Abbildung des Zeitpunkts

3.3 Abbildung des Zustandsvektors

Zur Vorhersage der täglichen Schlusskurse könnte mit einem Gauß-Prozess eine Abbildung

$$y_t = f(y_{t-1}, \dots, y_{t-d})$$

von d vergangenen Schlusskursen auf den zukünftigen Schlusskurs y_t gelernt werden. Die vergangenen Schlusskurse bilden hier einen *Zustandsvektor*, der den Zustand, in dem sich der zugrundeliegende Prozess befindet, möglichst gut beschreiben soll. Das Trainieren einer solchen Abbildung führt jedoch zu einem spärlich besetzten Input-Raum und es ist keine Generalisierungsfähigkeit für bisher nicht beobachtete Schlusskurse gegeben.

Aus diesem Grund wurde die zu lernende Abbildung verändert. Statt der vorangegangenen Schlusskurse wird eine Abbildung der d vorangegangenen täglichen Änderungen $r_t = y_t - y_{t-1}$ auf die heutige Änderung gelernt. Dies führt zu einem dichter besetzten Input-Raum und lässt den GP für bisher nicht beobachtete Kurse generalisieren.

$$r_t = f(r_{t-1}, \dots, r_{t-d})$$

Grundsätzlich können bei einer solchen Abbildung des Zustandsvektors komplexere Wechselwirkungen gelernt werden als bei einer Funktion der Zeitpunkte, jedoch entfällt die Möglichkeit, wie bei GPs mit Abbildung des Zeitpunkts die Charakteristika einer Zeitreihe intuitiv durch Komponenten in der Kovarianzfunktion zu modellieren.

Die erste Frage bei einem solchen Setup ist, wie groß die Dimension d des Zustandsvektors gewählt werden sollte. In den Experimenten für diese Studienarbeit wurde d sehr breit zwischen 10, 25, 50, 100, 150 und 200 Datenpunkten variiert.

Auch bei diesem Ansatz wird der GP nur auf einem Trainingsfenster trainiert. Da bei dieser Art von Abbildung aufgrund der obigen Überlegungen jedoch eine höhere Generalisierungsfähigkeit vermutet wird, wurden die Trainingsfenster etwas größer gewählt als in Abschnitt 3.2, es wurde variiert zwischen einer Breite von 150, 300, 450 und 600 Datenpunkten. Um zum vorherigen Ansatz vergleichbare Ergebnisse zu erhalten, wurden ebenfalls 1, 10 und 40-Schritt Vorhersagen des Wechselkurses durchgeführt. Der komplette Lern- und Vorhersagealgorithmus ist in **Abbildung 7** dargestellt.

Um diese Methode mit einem Benchmark zu vergleichen, der dieselbe Abbildung lernt, wird mit ihr auch ein *Neuronales Netz* trainiert. Es werden dabei dieselben Größen des Zustandsvektors und Trainingsfensters getestet. Das Neuronale Netz verfügt über eine innere Schicht, deren Größe 67% der Größe der Eingabeschicht (die wiederum abhängig von der Größe des Zustandsvektors ist) beträgt.

Die Gewichte des Neuronalen Netzes wurden für die kleineren Zustandsvektoren mit einem *Levenberg-Marquardt* Algorithmus trainiert, für die größeren Zustandsvektoren war dies aufgrund des hohen Speicherverbrauchs nicht möglich. Ab einer Größe von 100 Datenpunkten des Zustandsvektors wurde deshalb der *Scaled Conjugate Gradient* Algorithmus zur Optimierung der Gewichte verwendet. Das Trainingsfenster wurde randomisiert in 65% Trainingsdaten und 35% Validierungsdaten unterteilt, um ein Overfitting mit Hilfe der Validierungsdaten zu vermeiden. Nach 6 Versuchen, in denen der Trainingsfehler gesunken, aber der Validierungsfehler gestiegen ist, wurde dabei der Trainingsvorgang abgebrochen und die Gewichte mit dem geringsten Validierungsfehler genutzt.

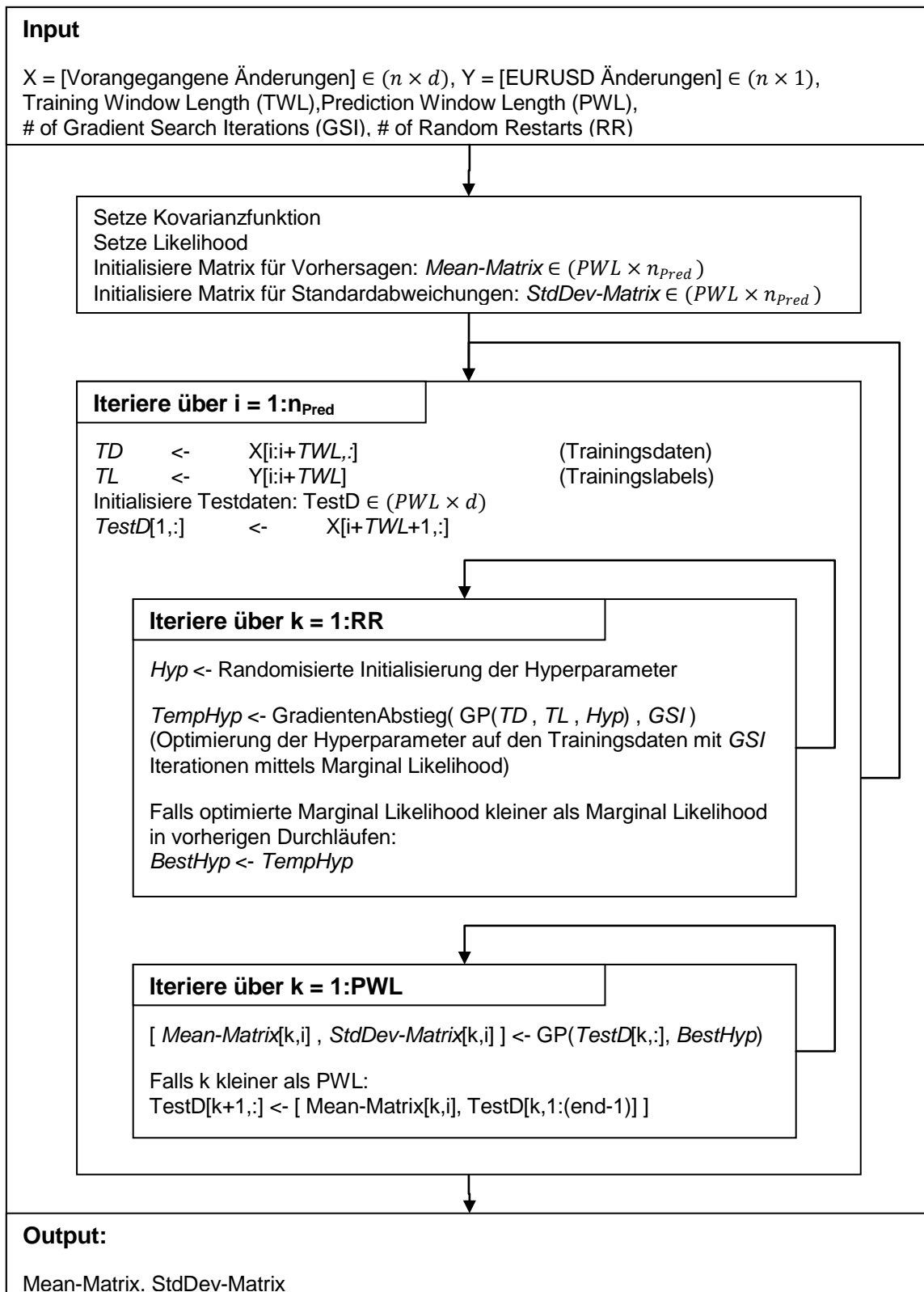


Abbildung 7: Ablaufdiagramm des Lern- und Vorhersagealgorithmus' zur Abbildung des Zustandsvektors

4 Resultate

4.1 Abbildung des Zeitpunkts

4.1.1 1-Schritt Vorhersage

In **Tabelle 1 - Tabelle 3** sind die *Mean Squared Errors (MSE)* der 1-Schritt Vorhersagen der unterschiedlichen Varianten dargestellt. Der MSE beschreibt den mittleren quadrierten Fehler zwischen der Vorhersage y_i^{Pred} und dem tatsächlichen Datenpunkt y_i für alle n_{Pred} vorherzusagenden Datenpunkte:

$$MSE = \frac{1}{n} \sum_{i=1}^{n_{Pred}} (y_i - y_i^{Pred})^2$$

Es wurden für jede Variante unterschiedliche Kovarianzfunktionen zur Vorhersage verwendet. Hier wurden zunächst die SE-, RQ- und Matérn-Kovarianzfunktion getestet. Da der Wechselkurs eine so hohe Ähnlichkeit mit einem Matérn-Prior hat (**Abbildung 3**), wurden darüber hinaus noch unterschiedliche Summen-Kombinationen der Matérn-Kovarianzfunktion getestet. Die Kovarianzfunktionen mit den geringsten MSE in jeder Variante sind in den Tabellen fett gedruckt.

	50	100	150	200
SE	1,5832e-02	9,3281e-03	7,2625e-03	6,9837e-03
RQ	1,1871e-04	1,0983e-04	1,0557e-04	9,8138e-05
Matérn	2,1207e-01	2,2849e-01	1,8221e-01	1,1704e-01
Matérn+Matérn	7,9835e-05	7,9671e-05	7,9703e-05	7,9808e-05
Matérn+SE	8,7651e-05	8,5451e-05	8,3590e-05	8,3234e-05
Matérn+RQ	8,8996e-05	8,4639e-05	8,3047e-05	8,3802e-05

Tabelle 1: MSE der 1-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster

	50	100	150	200
SE	3,7729e-04	4,0713e-04	4,9103e-04	4,4231e-04
RQ	1,1846e-04	1,2845e-04	1,4175e-04	1,4582e-04
Matérn	1,0375e-04	1,0163e-04	1,1038e-04	1,2568e-04
Matérn+Matérn	8,3764e-05	8,4775e-05	8,7628e-05	8,8304e-05
Matérn+SE	1,0055e-04	1,0056e-04	1,052e-04	1,1421e-04
Matérn+RQ	1,001e-04	9,6934e-05	1,0589e-04	1,0448e-04

Tabelle 2: MSE der 1-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Mean-Funktion)

	50	100	150	200
SE	9,9331e-05	1,0297e-04	1,0423e-04	1,0738e-04
RQ	8,6885e-05	8,6323e-05	8,6532e-05	8,7027e-05
Matérn	8,2171e-05	8,0545e-05	8,0399e-05	8,0390e-05
Matérn+Matérn	8,2428e-05	8,0603e-05	8,0526e-05	8,0602e-05
Matérn+SE	8,5532e-05	8,3402e-05	8,2039e-05	8,2059e-05
Matérn+RQ	8,5629e-05	8,3525e-05	8,2520e-05	8,2372e-05

Tabelle 3: MSE der 1-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Normalisierung)

	50	100	150	200
ES-SSM	8,269e-05	8,1766e-05	8,1201e-05	8,1162e-05
ARIMA	8,4589e-05	8,4021e-05	8,2209e-05	8,2535e-05
Naive	8,0479e-05	8,0479e-05	8,0479e-05	8,0479e-05

Tabelle 4: MSE der 1-Schritt Vorhersagen der Benchmarks mit unterschiedlichen Trainingsfenstern

Wie in **Tabelle 1** zu sehen ist, erzeugt die Vorhersage mittels SE- und Matérn-Kovarianzfunktion ohne Mean-Funktion und Normalisierung einen sehr großen mittleren Fehler. Die Vorhersage über die RQ-Kovarianzfunktion ist im Vergleich dazu um einige Größenordnungen besser. Als Kovarianzfunktionen mit den kleinsten MSE zeigen sich für diese Variante die zusammengesetzten Matérn-Kovarianzen.

Die Matérn- und SE-Kovarianzfunktionen erfassen beim Trainieren anscheinend nur die besonders starken kurzfristigen Autokorrelationen in den Daten, weshalb der length-scale Hyperparameter sehr klein wird. Außerhalb des Bereichs mit beobachteten Daten gehen diese Gauß-Prozesse daher sehr schnell zu ihrer Mean-Funktion zurück, die für diese Variante bei $m(x) = 0$ liegt. Die RQ-Kovarianzfunktion kann durch ihre weniger schnell abfallende Autokorrelationsfunktion jedoch längerfristige Abhängigkeiten erfassen, weshalb sie im Vergleich bessere Vorhersagen liefert. Der geringe Fehler der zusammengesetzten Matérn-Varianten bestätigt die Vermutung aus 3.2, dass diese Kovarianzfunktion zur Modellierung des Wechselkurses gut geeignet ist.

Um eine bessere Vorhersage auch mit einzelnen Matérn- und SE-Kovarianzfunktionen zu ermöglichen, wurde zum GP noch eine lineare Mean-Funktion hinzugefügt, deren Parameter ebenfalls über den Gradientenabstieg gelernt werden.

Wie in **Tabelle 2** zu erkennen ist, ermöglicht diese Variante eine relativ gute Vorhersage auch mit Matérn- und SE-Kovarianzfunktionen. Leider wird der Fehler bei den zusammengesetzten Matérn-Kovarianzen durch die Hinzunahme einer Mean-Funktion hingegen größer. Es hat den Anschein, dass die gelernte lineare Mean-Funktion entweder keine passende Annahme über den Wechselkurs darstellt oder beim Trainieren nicht gut auf die Daten angepasst wird.

Um dies zu prüfen wurde eine „manuelle“ Mean-Funktion angewendet, indem die Daten jedes Trainingsfensters durch eine lineare Normalisierung vorverarbeitet wurden (siehe 3.2). Aus **Tabelle 3** wird ersichtlich, dass damit im Vergleich zur Variante ohne Mean-Funktion und Normalisierung alle Vorhersagen mit Ausnahme der Matérn+Matérn Kombination besser werden. Es hat also den Anschein, dass die Vorverarbeitung besser funktioniert als die gelernte Mean-Funktion. Dabei darf jedoch nicht außer Acht gelassen werden, dass die Vorverarbeitung speziell darauf ausgelegt ist, dass der vorherzusagende Punkt sich nahe bei

Null befindet und dementsprechend einen Vorteil gegenüber der trainierten Mean-Funktion aus **Tabelle 2** hat.

Tabelle 4 zeigt die MSE der Benchmarks, die zu Vergleichszwecken getestet wurden. Auffallend ist zunächst, dass der naive Benchmark mit Abstand der Beste ist und auch viele Kovarianzfunktionen einen höheren MSE als die naive Vorhersage liefern. Unter Berücksichtigung der bereits vorhandenen Literatur ist dies jedoch nicht überraschend (siehe beispielsweise [16] und [17]) und zeigt zudem, warum die *random walk hypothesis* eine gute Näherung für den Wechselkurs darstellt.

Der Matérn+Matérn Summen-Kombination ohne Mean-Funktion (**Tabelle 1**) und der Matérn Kovarianzfunktion mit linearer Normalisierung (**Tabelle 3**) gelangen jedoch Vorhersagen mit geringerem MSE als dem naiven Benchmark. Dies zeigt, dass zwar eine starke Zufälligkeit im Wechselkurs vorhanden ist, da die naive Vorhersage im Vergleich so gut ist, jedoch noch eine gewisse Vorhersagbarkeit besteht, da die gut angepassten Modelle einen geringeren MSE als die naive Vorhersage liefern.

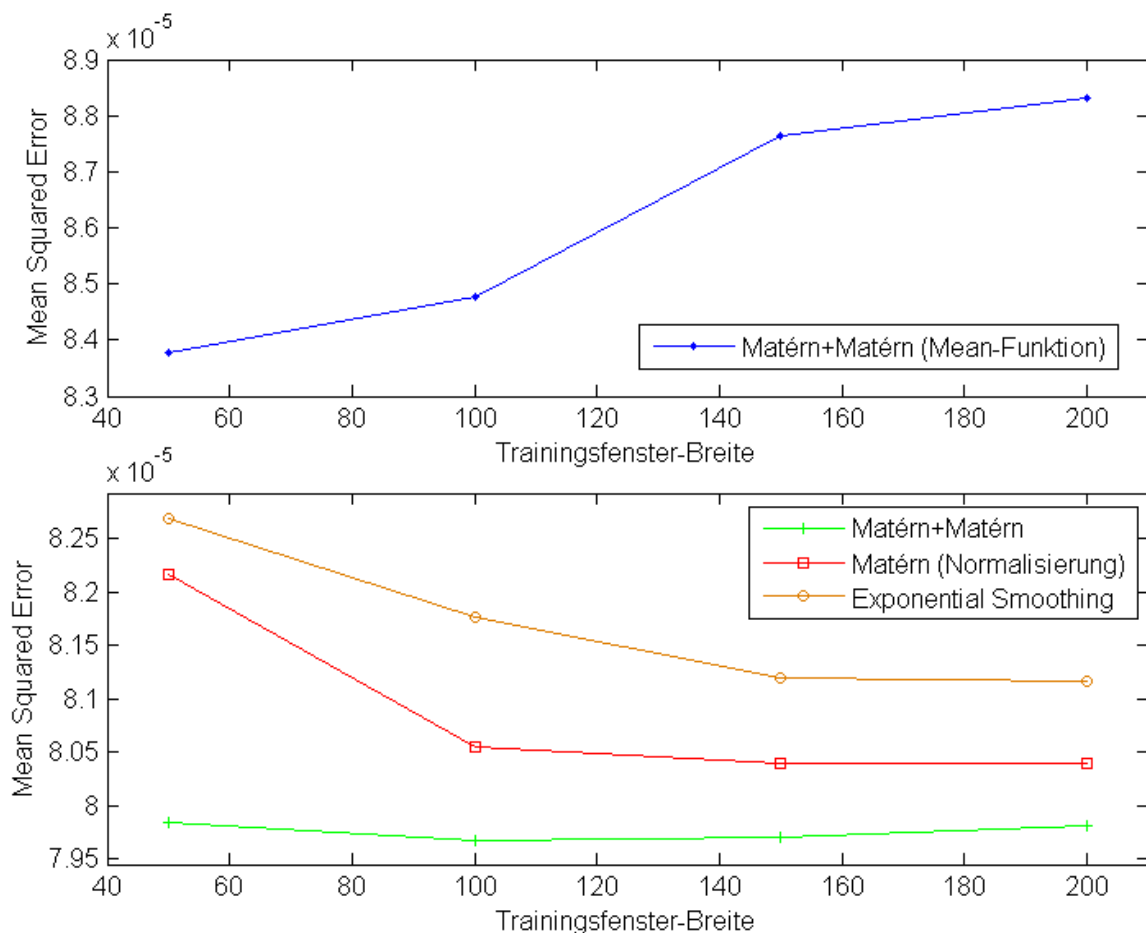


Abbildung 8: MSE in Abhängigkeit der Breite des Trainingsfensters für 1-Schritt Vorhersagen

In **Abbildung 8** ist dargestellt, wie der MSE der besten Kovarianzfunktionen jeder Variante von der Breite des Trainingsfensters abhängt. Für die „Matérn+Matérn (Mean-Funktion)“ Variante ergibt sich für größere Trainingsfenster ein steigender MSE, was einen weiteren Hinweis dafür liefert, dass die Spezifikation einer linearen Mean-Funktion für dieses Problem

nicht optimal ist. Die Vermutung aus Abschnitt 3.2 über den begrenzten Wert von alten Datenpunkten für die Vorhersage scheint sich zu bestätigen, denn die anderen Varianten haben ihr Minimum vor der maximalen Trainingsfenster-Breite von 200 Datenpunkten oder streben einem Grenzwert entgegen.

Aus den vorhandenen Vorhersagen wurde weiterhin ausgewertet, inwiefern die untersuchten Methoden in der Lage sind, einen Tag mit positiver oder negativer Änderung vorherzusagen. Die Vorhersage wird somit in ein Klassifikationsproblem umgewandelt, in **Abbildung 9** ist dazu die *Receiver Operating Characteristic* (ROC) dargestellt.

Die ROC trägt die *False Positive Rate* (der Anteil der Tage, die als positive Änderung vorhergesagt wurden, aber in Wahrheit eine negative Änderung brachten) gegen die *True Positive Rate* (der Anteil der Tage, die als positive Änderung vorhergesagt wurden und auch tatsächlich eine positive Änderung brachten) für variierende *Klassifikationsschwellen* auf.

Die Klassifikationsschwelle gibt an, welche Grenze die vorherzusagenden Tage in positive oder negative Änderungen einteilt. Da ein GP eine komplette Verteilung über die Vorhersage liefert, wurden als Klassifikationsschwellen die Quantile der Normalverteilung verwendet, wobei die Wahrscheinlichkeit in kleinen Schritten erhöht wurde. Für das Exponential Smoothing wurden zur Vorhersage positive oder negative Werte addiert, um die Klassifikationsschwelle zu verändern.

Für einen zufälligen Prädiktor wird bei einer Zunahme der True Positive Rate eine ebenso starke Zunahme der False Positive Rate erwartet, die zufällige Vorhersage wird also durch die Winkelhalbierende dargestellt. Ein guter Prädiktor sollte eine möglichst große Zunahme der True Positive Rate ohne Zunahme der False Positive Rate erzeugen, weshalb der ideale Prädiktor eine horizontale Linie bei *True Positive Rate* = 1 wäre. Dies ist für reale Daten jedoch kaum zu erreichen und für die Vorhersage von Wechselkursen noch weniger, ein kleiner Vorteil gegenüber der zufälligen Vorhersage bedeutet hier schon ein gutes Ergebnis.

Um die grafische Darstellung der ROC zu quantifizieren, kann zusätzlich die *Area Under Curve* berechnet werden. Die AUC ist die Fläche unterhalb der ROC-Kurve. Optimal ist dabei eine Fläche von 1, kein Vorteil gegenüber der zufälligen Vorhersage zeigt sich bei 0,5 und eine schlechtere Vorhersage als die zufällige ergibt sich aus AUC-Werten unterhalb von 0,5. Die AUC-Werte zu den berechneten ROC sind in **Tabelle 5** dargestellt.

Als Klassifikatoren wurden die besten Kovarianzfunktionen jeder Variante (ohne Mean-Funktion, mit Mean-Funktion, mit Normalisierung) gewählt, die in **Tabelle 1** - **Tabelle 3** fettgedruckt sind. Die naive Benchmark-Vorhersage kann keine Klassifizierung in positive und negative Änderungen liefern, da sie immer exakt den Schlusskurs des letzten Tages vorhersagt, deshalb wird hier die ROC des Exponential-Smoothing Modells als Benchmark genutzt.

Zu sehen ist, dass die „Matérn+Matérn (100)“ Variante ohne Mean-Funktion und Normalisierung nicht nur den besten MSE, sondern auch die beste Klassifikation liefert. Die ROC bewegt sich dabei konstant über der Winkelhalbierenden und verfügt auch über den höchsten AUC Wert.

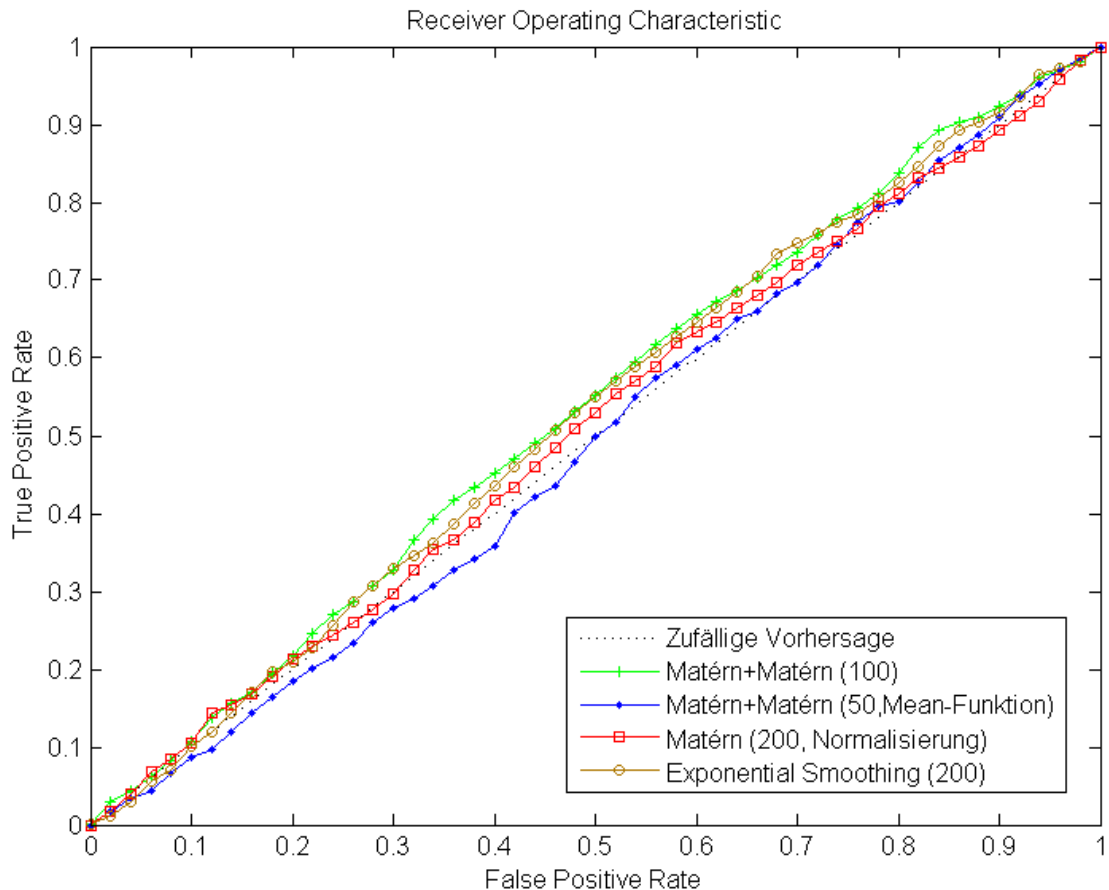


Abbildung 9: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 1-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.

Modell	Matérn+Matérn (100)	Matérn+Matérn (50, Mean-Funktion)	Matérn (200, Normalisierung)	Exponential Smoothing (200)
AUC	0,5337	0,4933	0,5111	0,5258

Tabelle 5: Area Under Curve Werte für die ROC aus **Abbildung 9**

Das Exponential-Smoothing Modell liefert auch eine relativ gute Klassifikation der täglichen Änderung, jedoch befindet sie sich größtenteils unterhalb der „Matérn+Matérn (100)“ ROC und hat dementsprechend einen niedrigeren AUC Wert, der jedoch immer noch höher ist als die AUC Werte der übrigen GP-Varianten.

Die Matérn Variante mit linearer Normalisierung scheint angesichts des Aussehens ihrer ROC und des AUC-Wertes knapp besser als die zufällige Vorhersage zu sein, wohingegen die Matérn+Matérn Variante mit Mean-Funktion nur sehr schlechte Klassifikationen liefert, die höchstwahrscheinlich nicht besser als eine zufällige Vorhersage sind, was einen weiteren Hinweis dafür liefert, dass das Lernen einer linearen Mean-Funktion für diese Daten nicht geeignet ist.

4.1.2 10-Schritt Vorhersage

In **Tabelle 6 - Tabelle 9** sind die MSE der gleichen Kovarianzfunktionen und Varianten für eine 10-Schritt Vorhersage dargestellt.

	50	100	150	200
SE	5,1046e-02	5,1424e-02	4,4504e-02	4,7097e-02
RQ	3,4006e-03	2,434e-03	1,8149e-03	1,5244e-03
Matérn	2,1464e-01	2,3066e-01	1,8648e-01	1,2098e-01
Matérn+Matérn	8,0927e-04	8,0924e-04	8,0983e-04	8,0987e-04
Matérn+SE	1,4849e-03	1,3697e-03	1,2965e-03	1,2112e-03
Matérn+RQ	1,5656e-03	1,331e-03	1,1993e-03	1,1543e-03

Tabelle 6: MSE der 10-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster

	50	100	150	200
SE	1,9653e-03	2,8089e-03	3,5907e-03	4,3374e-03
RQ	1,4921e-03	1,9213e-03	2,281e-03	2,3613e-03
Matérn	1,3311e-03	1,5435e-03	2,097e-03	2,9265e-03
Matérn+Matérn	9,976e-04	1,0835e-03	1,2377e-03	1,3143e-03
Matérn+SE	1,1147e-03	1,2425e-03	1,3433e-03	1,4482e-03
Matérn+RQ	1,081e-03	1,1840e-03	1,236e-03	1,2661e-03

Tabelle 7: MSE der 10-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Mean-Funktion)

	50	100	150	200
SE	1,0248e-03	9,5269e-04	1,0004e-03	1,0769e-03
RQ	1,0646e-03	9,675e-04	1,0169e-03	1,1021e-03
Matérn	9,9589e-04	8,8083e-04	8,7181e-04	8,7377e-04
Matérn+Matérn	9,9806e-04	8,7972e-04	8,7324e-04	8,7632e-04
Matérn+SE	1,031e-03	9,5685e-04	9,4116e-04	9,6831e-04
Matérn+RQ	1,0221e-03	9,2864e-04	9,3736e-04	9,476e-04

Tabelle 8: MSE der 10-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Normalisierung)

	50	100	150	200
ES-SSM	8,9998e-04	8,9242e-04	8,7379e-04	8,4072e-04
ARIMA	1,0486e-03	9,6677e-04	9,2363e-04	8,9989e-04
Naive	8,173e-04	8,173e-04	8,173e-04	8,173e-04

Tabelle 9: MSE der 10-Schritt Vorhersagen der Benchmarks mit unterschiedlichen Trainingsfenstern

Für die 10-Schritt Vorhersagen ist der MSE im Durchschnitt um etwa eine Größenordnung höher, auf diesem höheren Niveau verhalten sich die MSE relativ zueinander aber ähnlich zur

1-Schritt Vorhersage. Dies zeigt sich schon daran, dass die gleichen Kovarianzfunktionen wie auch bei der 1-Schritt Vorhersage den geringsten MSE für alle drei Varianten aufweisen. In **Tabelle 6** erkennt man wiederum, dass die „Matérn+Matérn (100)“ Variante den geringsten MSE aufweist und auch besser als die Benchmarks ist. Einzelne Kovarianzfunktionen mit Ausnahme der RQ-Kovarianzfunktion sind auch für die 10-Schritt Vorhersage schlecht geeignet und die Matérn Kombinationen liefern auch hier den kleinsten MSE. Auch die übrigen Tabellen zeigen ein ähnliches Verhalten des MSE.

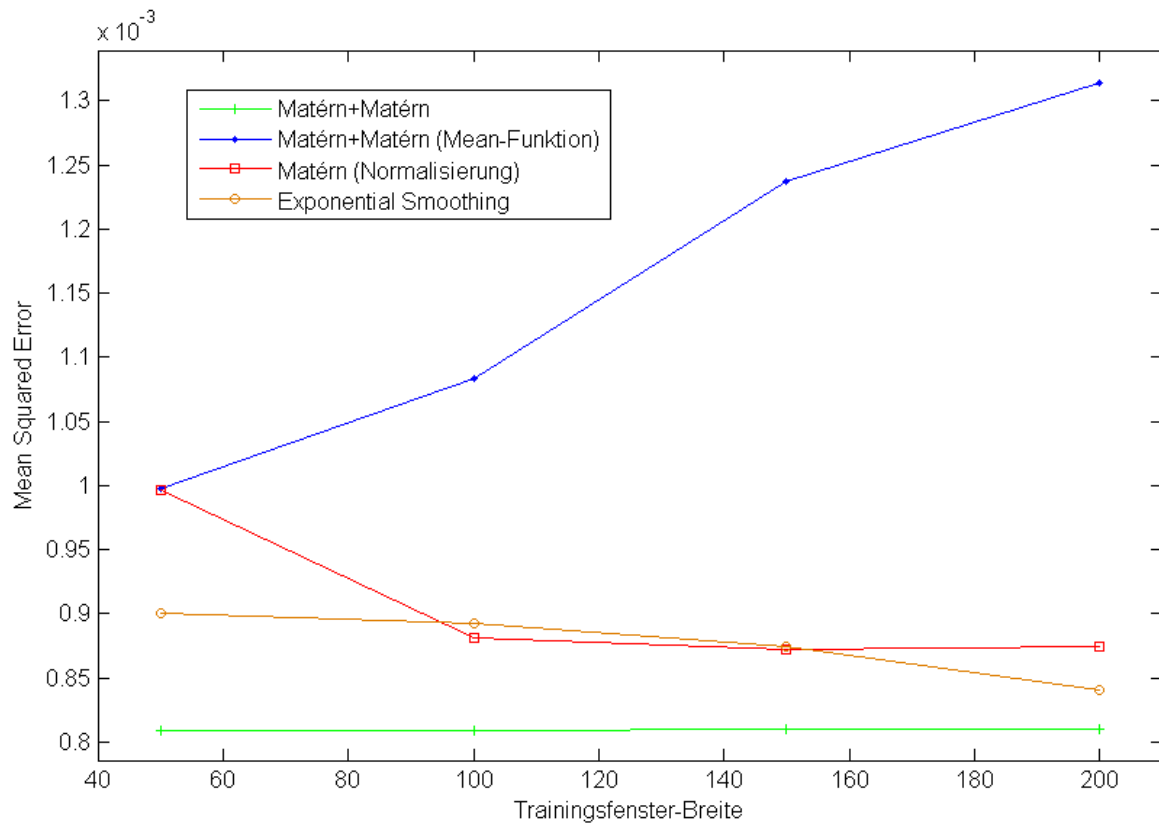


Abbildung 10: MSE in Abhängigkeit der Breite des Trainingsfensters für 10-Schritt Vorhersagen

Die Entwicklung des MSE für wechselnde Trainingsfenster bei 10-Schritt Vorhersagen ist in **Abbildung 10** dargestellt. Es zeigt sich wiederum das auf eine falsche Spezifikation der linearen Mean-Funktion hinweisende Verhalten der „Matérn+Matérn (Mean-Funktion)“ GP-Variante, sowie der limitierte Wert von älteren Datenpunkten zur Vorhersage.

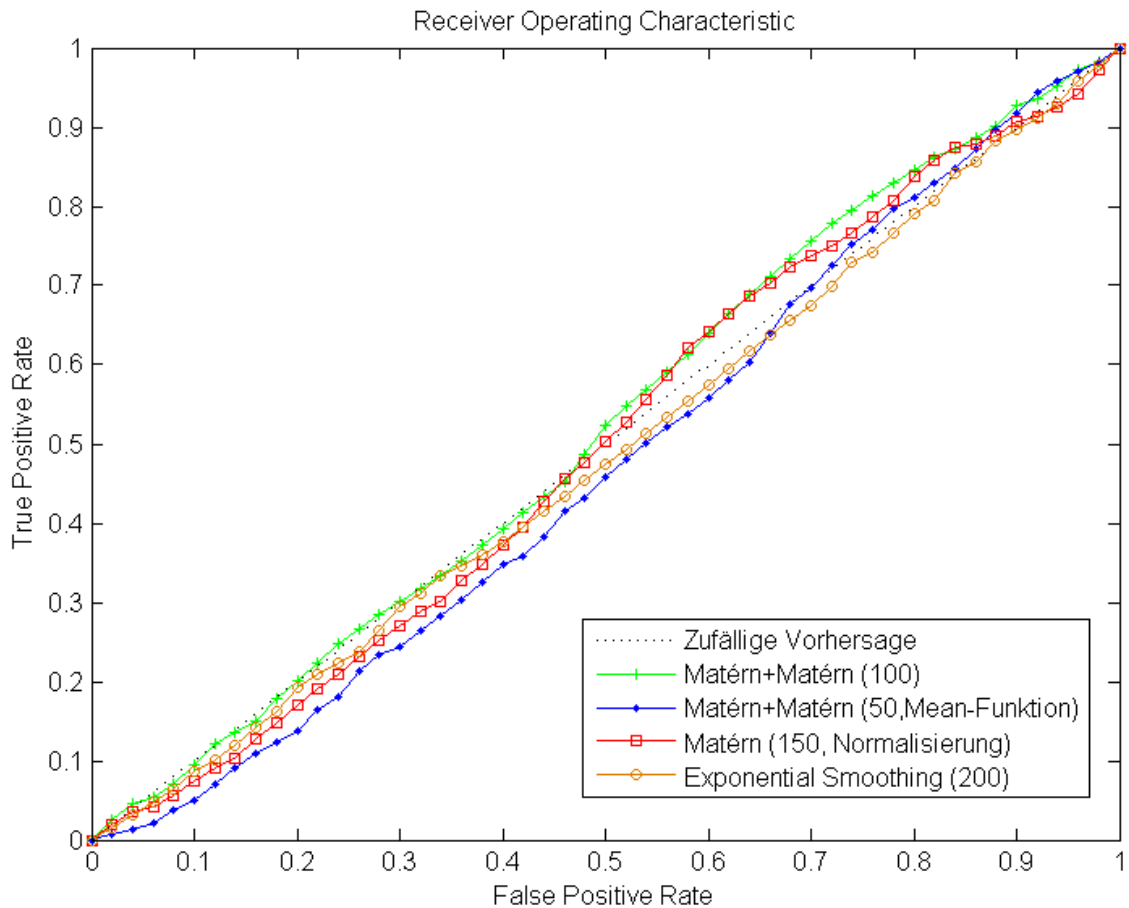


Abbildung 11: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 10-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.

Modell	Matérn+Matérn (100)	Matérn+Matérn (50, Mean-Funktion)	Matérn (150, Normalisierung)	Exponential Smoothing (200)
AUC	0,517	0,4726	0,4996	0,4848

Tabelle 10: Area-Under-Curve Werte für die ROC aus Abbildung 11

Die ROC der Klassifikatoren für 10-Schritt Änderungen in **Abbildung 11** zeigen im Vergleich zur ROC der 1-Schritt Änderungen jedoch ein deutlich verändertes Verhalten. Es zeigt sich, dass nur noch der Matérn+Matérn 100 Klassifikator besser als die zufällige Vorhersage ist, für die übrigen Klassifikatoren ist der AUC Wert kleiner als 0,5. Die Kurven des „Matérn+Matérn (100)“ und des „Matérn (150, Normalisierung)“ Klassifikators zeigt außerdem eine ausgeprägte Asymmetrie, nur für hohe positive Klassifikationsraten scheint die Klassifikation gute Ergebnisse zu liefern, außerhalb dieses Bereichs wirkt die Klassifikation zufällig. Dies ist womöglich auf eine Asymmetrie in den Verteilungen der Vorhersagen für beide Klassen zurückzuführen.

4.1.3 40-Schritt Vorhersage

In **Tabelle 11** - **Tabelle 14** sind die MSE der verschiedenen Varianten für eine 40-Schritt Vorhersage dargestellt.

	50	100	150	200
SE	9,7691e-01	1,9995e+00	1,4136e+00	4,6802e+00
RQ	2,6538e-02	1,6362e-03	8,8835e-03	7,1322e-03
Matérn	2,1718e-01	2,3805e-01	1,9212e-01	1,2822e-01
Matérn+Matérn	3,4772e-03	3,4818e-03	3,4801e-03	3,479e-03
Matérn+SE	3,2107e-02	2,6745e-02	2,5648e-02	2,0846e-02
Matérn+RQ	1,4216e-02	1,2217e-02	1,0947e-02	9,7619e-03

Tabelle 11: MSE der 40-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster

	50	100	150	200
SE	4,6333e-03	6,0338e-03	8,6182e-03	8,9769e-03
RQ	4,4265e-03	5,3033e-03	5,9001e-03	6,0353e-03
Matérn	4,5570e-03	5,4479e-03	8,1091e-03	1,2198e-02
Matérn+Matérn	4,2154e-03	4,7204e-03	5,3545e-03	5,6089e-03
Matérn+SE	4,3464e-03	5,3068e-03	6,4314e-03	6,6633e-03
Matérn+RQ	4,2519e-03	4,7542e-03	5,1197e-03	5,1847e-03

Tabelle 12: MSE der 40-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Mean-Funktion)

	50	100	150	200
SE	6,0347e-03	4,7206e-03	4,755e-03	4,553e-03
RQ	6,0155e-03	4,8179e-03	5,2401e-03	5,2456e-03
Matérn	6,0361e-03	4,7216e-03	4,7551e-03	4,5525e-03
Matérn+Matérn	6,0369e-03	4,7203e-03	4,7574e-03	4,5556e-03
Matérn+SE	6,0239e-03	4,8158e-03	4,8716e-03	4,9923e-03
Matérn+RQ	6,0096e-03	4,7408e-03	4,859e-03	4,8409e-03

Tabelle 13: MSE der 40-Schritt Vorhersagen für unterschiedliche Kovarianzfunktionen und Trainingsfenster (lineare Normalisierung)

	50	100	150	200
ES-SSM	4,3912e-03	4,5274e-03	4,3146e-03	3,9347e-03
ARIMA	6,8121e-03	5,3927e-03	4,6925e-03	5,241e-03
Naive	3,447e-03	3,447e-03	3,447e-03	3,447e-03

Tabelle 14: MSE der 40-Schritt Vorhersagen der Benchmarks mit unterschiedlichen Trainingsfenstern

In **Tabelle 11** sind die gleichen Effekte wie bei der 1- und 10-Schritt Vorhersage zu erkennen, für die SE-Kovarianzfunktion wird der Fehler bei dieser langfristigen Vorhersage sehr groß,

was sich aber wieder durch den kleinen length-scale Hyperparameter erklären lässt, der in einem groben Mißverhältnis zur langfristigen Vorhersage steht. In **Tabelle 12** und **Tabelle 13** fällt auf, dass die MSE für gleiche Trainingsfensterbreiten sehr ähnlich sind. Der Einfluss der Kovarianzfunktion geht bei einer solch langfristigen Vorhersage so stark zurück, dass letztendlich die Mean-Funktion für die Vorhersage maßgeblich ist. Da für alle Kovarianzfunktionen die gleiche Mean-Funktion spezifiziert wurde ähneln sich die Vorhersagen so stark. Die Vorhersagen mit Gauß-Prozessen zeigen für die 40-Schritt Vorhersage einen größeren MSE als der naive Benchmark. Dies deutet auf einen random walk für weit voneinander entfernte Punkte hin.

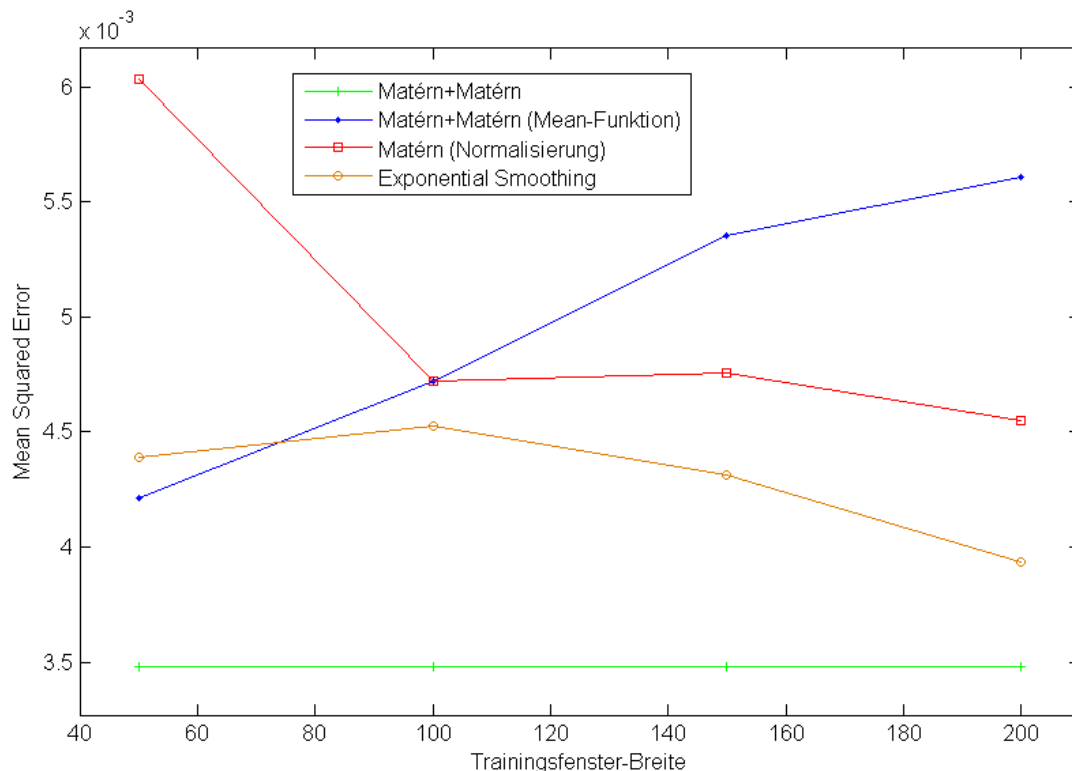


Abbildung 12: MSE in Abhängigkeit der Breite des Trainingsfensters für 40-Schritt Vorhersagen

In **Abbildung 12** zeigen sich wieder zur 1- und 10-Schritt Vorhersage ähnliche Ergebnisse, der MSE der GP-Variante mit trainierter Mean-Funktion steigt für größere Trainingsfenster, für die anderen Varianten und den Benchmark sinkt er jedoch.

Die ROC in **Abbildung 13** zeigen, dass es mit GPs nicht möglich ist, eine langfristige Vorhersage zu treffen. Keiner der Klassifikatoren liefert einen AUC Wert über 0,5, deshalb würde es sich lohnen, die Vorhersagen der Modelle umzukehren. Eine Umkehrung der Vorhersage würde bedeuten, dass die ROC an der Winkelhalbierenden gespiegelt würde. Besonders interessant wäre dies bei dem „Exponential Smoothing (200)“ Modell, da seine ROC stabil unterhalb der Winkelhalbierenden verläuft.

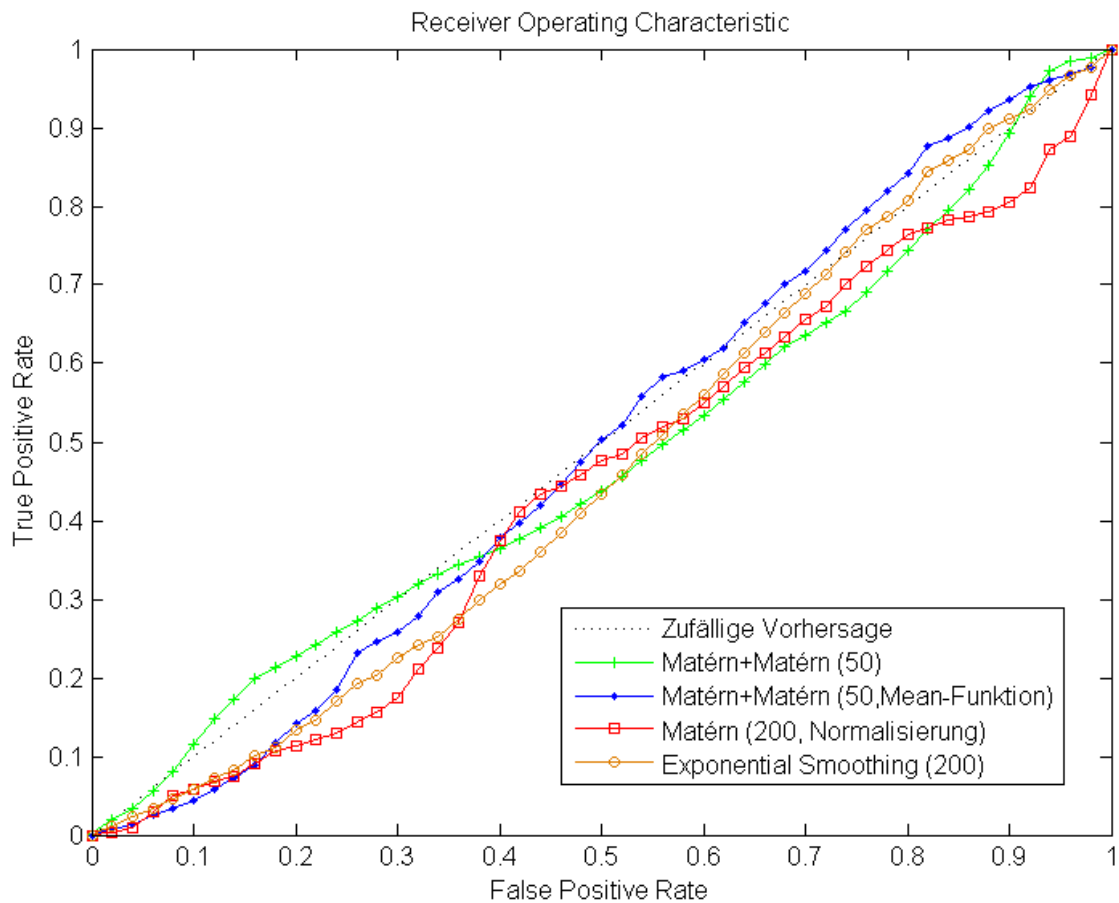


Abbildung 13: Receiver Operating Characteristic der GPs und des Benchmarks mit den geringsten MSEs für 40-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters angegeben.

Modell	Matérn+Matérn (100)	Matérn (50, Mean-Funktion)	Matérn (200, Normalisierung)	Exponential Smoothing (200)
AUC	0,4766	0,4928	0,4441	0,4633

Tabelle 15: Area-Under-Curve Werte für die ROC aus Abbildung 11

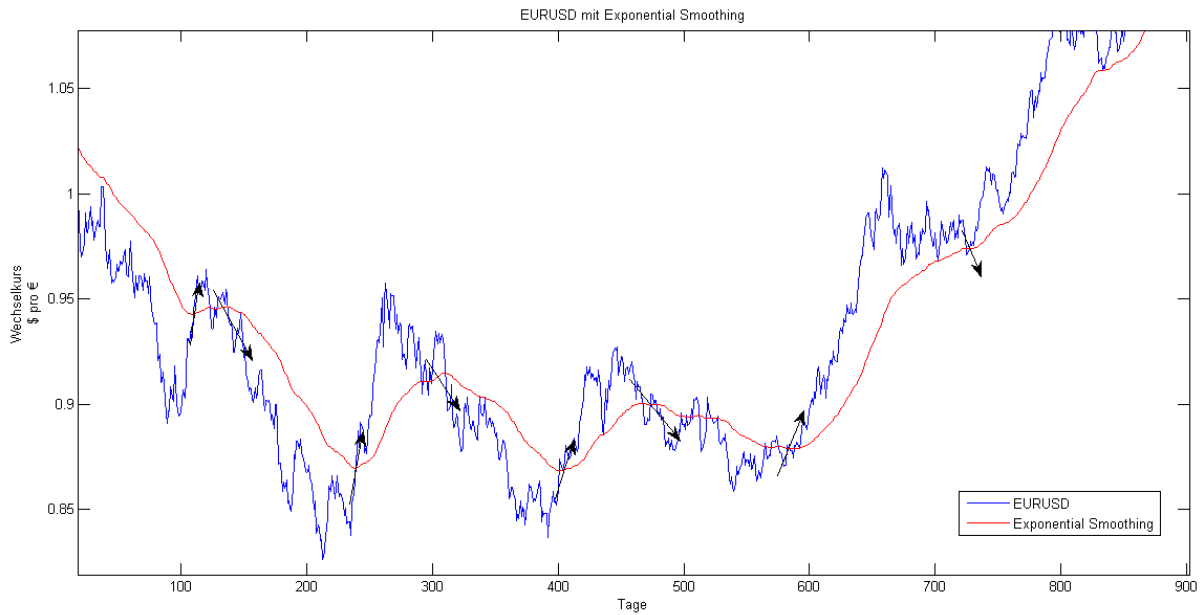


Abbildung 14: Darstellung der Kauf und Verkaufssignale eines Moving-Average Crossover Systems. Die schwarzen Pfeile geben die Richtung an, die durch ein Kreuzen des Exponential-Smoothings vorhergesagt wird. Dies entspricht der Umkehrung der untersuchten Exponential-Smoothing Vorhersage.

Ein interessantes Ergebnis ist dies trotzdem, da es die weite Verbreitung von *Moving-Average Crossover* Strategien erklärt. Unter Praktikern am Finanzmarkt sind verschiedene Strategien recht verbreitet, die eine gewisse Vorhersagbarkeit der Finanzmarktkurse ermöglichen sollen.

Die Moving-Average Crossover Strategie beruht auf der Annahme, dass bei einem Kreuzen des Kurses mit dem Moving Average (die Exponential Smoothing Methode ist ein Moving Average) von oben nach unten auf fallende Kurse und bei einem Kreuzen des Kurses von unten nach oben auf steigende Kurse gesetzt werden sollte. Die Kauf- und Verkaufssignale einer solchen Strategie sind beispielhaft in **Abbildung 14** dargestellt. Diese Signale wären exakt die Umkehrung der hier angewendeten Vorhersage, würde also zu einer ROC über der Winkelhalbierenden und einem AUC-Wert größer als 0,5 führen.

4.2 Abbildung des Zustandsvektors

4.2.1 1-Schritt Vorhersage

SE-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,0016e-05	7,9798e-05	7,9755e-05	7,9773e-05
	25	7,9998e-05	7,9784e-05	7,9752e-05	7,9774e-05
	50	8,001e-05	7,9807e-05	7,9742e-05	7,9756e-05
	100	8,0018e-05	7,9801e-05	7,9758e-05	7,9772e-05
	150	8,0017e-05	7,9807e-05	7,9752e-05	7,9779e-05
	200	8,0047e-05	7,9824e-05	7,9757e-05	7,9780e-05

Tabelle 16: MSE der 1-Schritt Vorhersagen für die SE-Kovarianz bei einer Abbildung des Zustandsvektors

RQ-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,0009e-05	7,9797e-05	7,9756e-05	7,9772e-05
	25	8,0028e-05	7,9790e-05	7,9745e-05	7,9774e-05
	50	8,0009e-05	7,9798e-05	7,9750 e-05	7,9767e-05
	100	8,0012e-05	7,9798e-05	7,9753e-05	7,9772e-05
	150	8,0072e-05	7,9790e-05	7,9753e-05	7,9771e-05
	200	8,0001e-05	7,9787e-05	7,9745e-05	7,9775e-05

Tabelle 17: MSE der 1-Schritt Vorhersagen für die RQ-Kovarianz bei einer Abbildung des Zustandsvektors

Matérn-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,01e-05	7,9888e-05	8,0007e-05	8,0427e-05
	25	8,021e-05	7,9853e-05	8,0436e-05	8,0706e-05
	50	8,0183e-05	8,0126e-05	8,0457e-05	8,1073e-05
	100	8,0231e-05	8,0124e-05	8,2131e-05	8,2121e-05
	150	8,0349e-05	8,0249e-05	8,1161e-05	8,1252e-05
	200	8,0642e-05	8,0946e-05	8,1508e-05	8,1419e-05

Tabelle 18: MSE der 1-Schritt Vorhersagen für die Matérn-Kovarianz bei einer Abbildung des Zustandsvektors

Neuronales Netz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	9,2467e-05	9,1331e-05	8,8935e-05	8,9352e-05
	25	1,3074e-04	1,2243e-04	1,2038e-04	1,1502e-04
	50	2,1752e-04	1,6938e-04	1,5869e-04	1,5920e-04
	100	5,5535e-04	3,1809e-04	2,3803e-04	1,9923e-04
	150	1,4842e-03	5,8343e-04	3,6863e-04	3,2797e-04
	200	2,2342e-03	9,9905e-04	6,3443e-04	4,6700e-04

Tabelle 19: MSE der 1-Schritt Vorhersagen für ein Neuronales Netz bei einer Abbildung des Zustandsvektors

In **Tabelle 16 - Tabelle 18** sind die MSE der GP-Modelle bei Abbildung des Zustandsvektors dargestellt. Über die Zeilen variiert die Dimension des Zustandsvektors und über die Spalten die Breite des Trainingsfensters. Hier ist der Unterschied zwischen den Kovarianzfunktionen bezüglich ihres MSE nicht so groß wie bei GPs mit Abbildung des Zeitpunkts. RQ- und SE-Kovarianz liefern sehr ähnliche MSE, der Fehler der Matérn-Kovarianz ist nur geringfügig größer.

Eine Änderung des MSE bei gleicher Kovarianzfunktion zeigt sich hauptsächlich für Änderungen der Trainingsfenster-Breite. Änderungen der Größe des Zustandsvektors verändern den MSE nur minimal. Das Minimum des MSE liegt für SE- und RQ-Kovarianz bei 450 Trainingspunkten, für die Matérn-Kovarianz bei 300 Trainingspunkten und überraschenderweise für alle drei Kovarianzfunktionen bei nur 25-50 Dimensionen des Zustandsvektors.

Obwohl bei der Abbildung des Zustandsvektors komplexere Wechselwirkungen in den Wechselkursdaten erfasst werden können, verbessert sich die Vorhersage nicht gegenüber dem besten Modell bei der Abbildung des Zeitpunkts. Womöglich liegen komplexe Wechselwirkungen in den Kursdaten nicht vor. Trotzdem besitzt das beste Modell (SE(50,450) aus **Tabelle 16**) einen besseren MSE als die naive Vorhersage.

Die Vorhersagen des Neuronalen Netzes können, gemessen am MSE, mit den Vorhersagen des GP nicht mithalten, der MSE der Vorhersagen des Neuronalen Netzes liegt für alle Zustandsvektoren und Trainingsfenster deutlich über den MSE der GP Modelle. Vor allem für eine Trainingsfenster-Breite von 150 und Dimensionen des Zustandsvektors von 150-200 zeigt sich zudem ein leichtes Overfitting beim Neuronalen Netz, da der MSE im Vergleich zu einer Trainingsfenster-Breite von 300 einen großen Sprung macht.

Auch diese Vorhersagen wurden genutzt, um eine Vorhersage von positiver oder negativer Änderung zu treffen. Die Klassifikationsschwelle für das GP-Modell wurde hierbei wieder mit Hilfe der Quantile der Normalverteilung verändert, zur Vorhersage des Neuronalen Netzes wurden positive oder negative Werte addiert, um die Klassifikationsschwelle zu verändern.

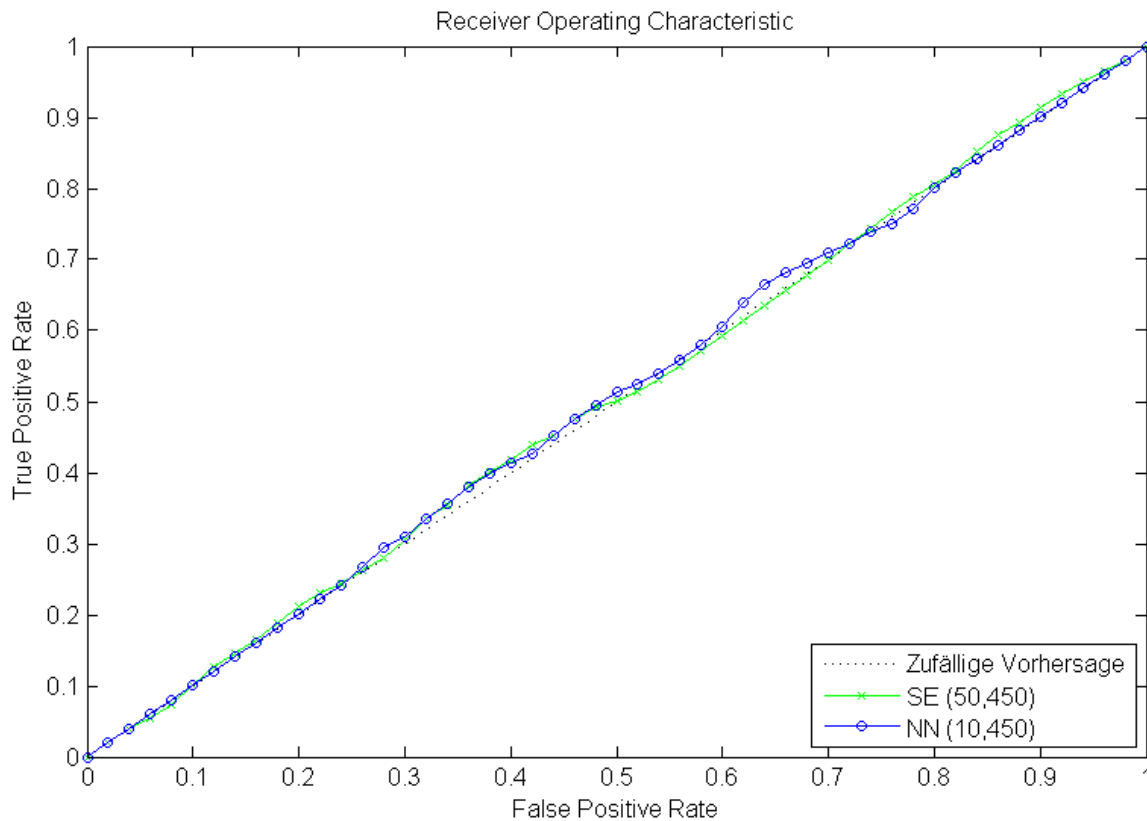


Abbildung 15: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSE für 1-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Dimension des Zustandsvektors und die Länge des Trainingsfensters angegeben.

Modell	SE (200,450)	NN (10,450)
AUC	0,505	0,5056

Tabelle 20: Area-Under-Curve Werte für die ROC aus Abbildung 15

Im Vergleich zum besten Modell bei Abbildung des Zeitpunkts bewegen sich die ROC in **Abbildung 15** deutlich näher an der zufälligen Vorhersage und bieten somit eine deutlich schlechtere Vorhersage, ob der künftige Tag eine positive oder negative Änderung bringen wird. Die AUC Werte bewegen sich dennoch leicht über der Marke von 0,5, wobei fraglich ist, ob dieses Ergebnis wirklich signifikant ist.

Das „SE (200,450)“ GP-Modell ist interessanterweise, entgegen den Ergebnissen bei den MSE, im Nachteil gegenüber dem Neuronalen Netz. Eine Erklärung dafür könnte sein, dass die Vorhersagen des Neuronalen Netzes vom Betrag her meist deutlich größer als die Vorhersagen des GP-Modells waren, und eine falsche Vorhersage des Neuronalen Netzes so einen größeren Beitrag zum MSE lieferte.

4.2.2 10-Schritt Vorhersage

SE-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,4014e-04	8,1384e-04	8,1080e-04	8,1277e-04
	25	8,4032e-04	8,1382e-04	8,1062e-04	8,1279e-04
	50	8,3964e-04	8,1410e-04	8,1040e-04	8,1269e-04
	100	8,3986e-04	8,1404e-04	8,1064e-04	8,1278e-04
	150	8,4016e-04	8,1444e-04	8,1053e-04	8,1314e-04
	200	8,4151e-04	8,1494e-04	8,1035e-04	8,1383e-04

Tabelle 21: MSE der 10-Schritt Vorhersagen für die SE-Kovarianz bei einer Abbildung des Zustandsvektors

RQ-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,4093e-04	8,1381e-04	8,1081e-04	8,1283e-04
	25	8,4052e-04	8,1363e-04	8,1065e-04	8,1268e-04
	50	8,4029e-04	8,14e-04	8,1052e-04	8,1267e-04
	100	8,4039e-04	8,1406e-04	8,1045e-04	8,1277e-04
	150	8,4098e-04	8,1436e-04	8,1045e-04	8,1280e-04
	200	8,4208e-04	8,1377e-04	8,1043e-04	8,1317e-04

Tabelle 22: MSE der 10-Schritt Vorhersagen für die SE-Kovarianz bei einer Abbildung des Zustandsvektors

Matérn-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	8,4401e-04	8,1568e-04	8,1239e-04	8,1372e-04
	25	8,3886e-04	8,17e-04	8,3441e-04	8,4439e-04
	50	8,4278e-04	8,1827e-04	8,5430e-04	8,6572e-04
	100	8,4433e-04	8,2407e-04	8,8839e-04	9,0285e-04
	150	8,5497e-04	8,4951e-04	8,7330e-04	8,8313e-04
	200	8,6408e-04	8,6357e-04	8,5841e-04	8,7428e-04

Tabelle 23: MSE der 10-Schritt Vorhersagen für die Matérn-Kovarianz bei einer Abbildung des Zustandsvektors

Neuronales Netz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	1,1530e-03	1,0475e-03	9,4852e-04	9,5433e-04
	25	1,8033e-03	1,6155e-03	1,5221e-03	1,4144e-03
	50	2,9383e-03	2,2742e-03	1,9913e-03	1,9897e-03
	100	9,4160e-03	4,3525e-03	2,9986e-03	2,5002e-03
	150	3,9634e-02	8,9586e-03	5,4583e-03	4,1673e-03
	200	6,7943e-02	1,9075e-02	8,5791e-03	6,2130e-03

Tabelle 24: MSE der 10-Schritt Vorhersagen für ein Neuronales Netz bei einer Abbildung des Zustandsvektors

Für die 10-Schritt Vorhersage ist der Unterschied in den MSE zwischen verschiedenen Kovarianzen zwar etwas größer als bei der 1-Schritt Vorhersage, ist im Vergleich zur 10-Schritt Vorhersage bei Abbildung des Zeitpunkts jedoch noch immer gering. Wieder ist zu beobachten, dass hauptsächlich Änderungen der Trainingsfenster-Breite zu Änderungen des MSE führen.

Die Minima bezüglich der Breite des Trainingsfensters liegen für die 10-Schritt Vorhersage bei allen Kovarianzfunktionen bei 450 Datenpunkten. Die Minima bezüglich der Größe des Zustandsvektors sind für die SE- und RQ-Kovarianz 200 Dimensionen, wobei dies auch ein zufälliger Effekt sein könnte, da weniger klare Tendenzen bezüglich der Dimension des Zustandsvektors zu erkennen sind. Bei der Matérn-Kovarianz zeigt sich hier jedoch die relativ klare Tendenz, dass kleinere Dimensionen des Zustandsvektors geringere MSE erzeugen.

Das Neuronale Netz liefert erneut deutlich schlechtere MSE als alle GP Modelle, der Unterschied ist jedoch nicht mehr so groß wie noch bei der 1-Schritt Vorhersage. Der minimale MSE liegt hier erneut bei 10 Dimensionen für den Zustandsvektor und 450 Trainingspunkten.

Das beste Modell (SE (200,450) aus **Tabelle 21**) ist auch hier erneut besser als die naive Vorhersage. Jedoch ist ebenso der GP mit Abbildung des Zeitpunkts bezüglich des MSE knapp besser. Es werden in **Abbildung 16** die ROC der 10-Schritt Vorhersagen dargestellt.

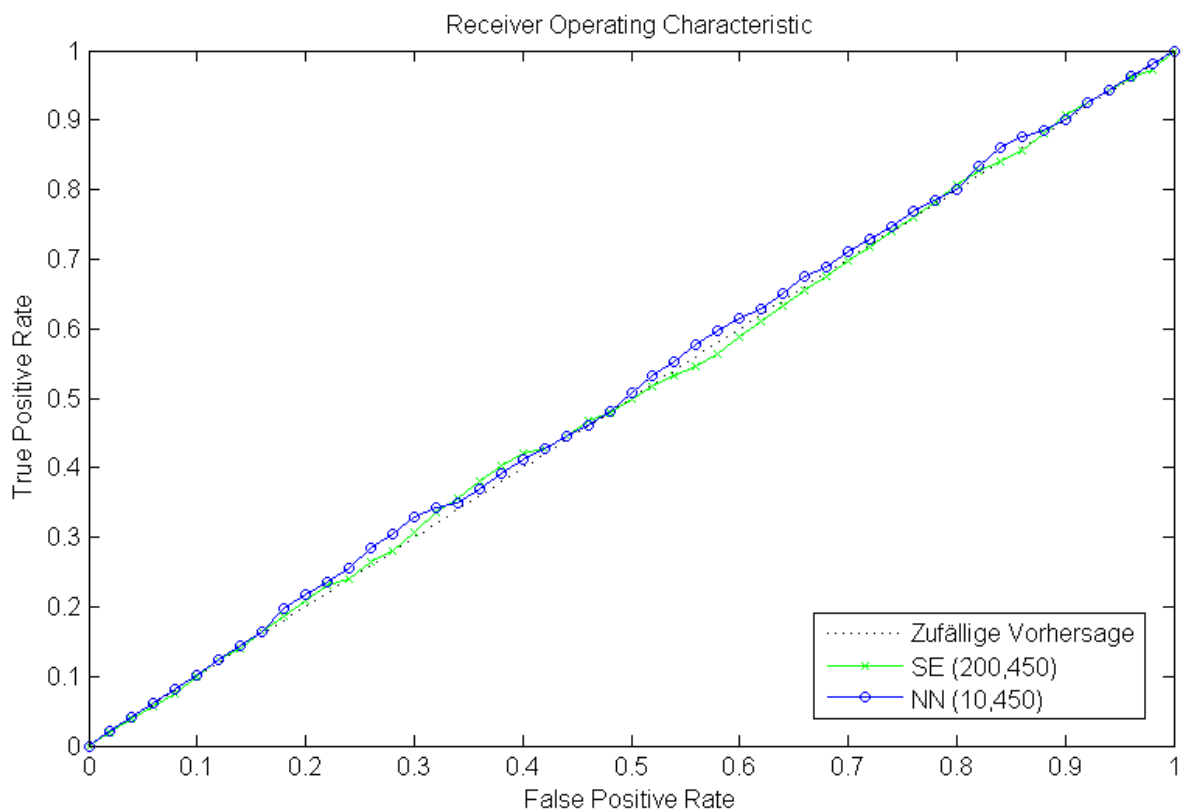


Abbildung 16: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSE für 10-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters und die Dimension des Zustandsvektors angegeben.

Modell	SE (200,450)	NN (10,450)
AUC	0,5014	0,50917

Tabelle 25: Area-Under-Curve Werte für die ROC aus Abbildung 16

Die Receiver Operating Characteristic zeigt wieder eher enttäuschende Ergebnisse. Die AUC bleiben zwar knapp über 0,5, jedoch ist erneut fraglich, ob dies statistisch signifikant ist. Das GP-Modell zeigt wie schon bei der 1-Schritt Vorhersage etwas schlechtere Klassifikationen als das Neuronale Netz. Interessanterweise ist der AUC Wert für die Klassifikationen des Neuronalen Netzes etwas höher als bei der 1-Schritt Vorhersage.

4.2.3 40-Schritt Vorhersage

SE-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	4,1077e-03	3,5806e-03	3,5124e-03	3,5565e-03
	25	4,1081e-03	3,5787e-03	3,5114e-03	3,5574e-03
	50	4,1039e-03	3,5794e-03	3,5091e-03	3,5583e-03
	100	4,1042e-03	3,5773e-03	3,5098e-03	3,5579e-03
	150	4,1055e-03	3,5806e-03	3,5093e-03	3,562e-03
	200	4,1168e-03	3,5901e-03	3,5096e-03	3,5603e-03

Tabelle 26: MSE der 40-Schritt Vorhersagen für die SE-Kovarianz bei einer Abbildung des Zustandsvektors

RQ-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	4,1085e-03	3,5787e-03	3,5126e-03	3,5561e-03
	25	4,1056e-03	3,5771e-03	3,5121e-03	3,5573e-03
	50	4,1057e-03	3,5784e-03	3,5093e-03	3,5579e-03
	100	4,1057e-03	3,578e-03	3,5088e-03	3,5579e-03
	150	4,1072e-03	3,586e-03	3,5075e-03	3,5578e-03
	200	4,1137e-03	3,5771e-03	3,5087e-03	3,5609e-03

Tabelle 27: MSE der 40-Schritt Vorhersagen für die SE-Kovarianz bei einer Abbildung des Zustandsvektors

Matérn-Kovarianz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	4,1048e-03	3,5857e-03	3,5755e-03	3,6522e-03
	25	4,0894e-03	3,6222e-03	3,6841e-03	3,8994e-03
	50	4,0756e-03	3,6663e-03	4,0299e-03	4,194e-03
	100	4,1448e-03	3,7778e-03	4,3131e-03	4,4812e-03
	150	4,2498e-03	4,007e-03	4,1838e-03	4,2687e-03
	200	4,3397e-03	4,157e-03	4,1924e-03	4,1974e-03

Tabelle 28: MSE der 40-Schritt Vorhersagen für die Matérn-Kovarianz bei einer Abbildung des Zustandsvektors

Neuronales Netz MSE		Trainingsfenster-Breite			
		150	300	450	600
Dimension des Zustands- vektors	10	9,4593e-03	7,8796e-03	5,7953e-03	5,5821e-03
	25	2,0607e-02	1,6172e-02	1,6591e-02	1,4427e-02
	50	3,8677e-02	2,3054e-02	1,8712e-02	2,2476e-02
	100	2,0835e-01	5,1376e-02	2,1551e-02	1,7967e-02
	150	1,3424e+00	1,4895e-01	5,7446e-02	4,0841e-02
	200	2,0514e+00	5,0281e-01	1,2210e-01	6,8685e-02

Tabelle 29: MSE der 40-Schritt Vorhersagen für ein Neuronales Netz bei einer Abbildung des Zustandsvektors

Für die 40-Schritt Vorhersagen bleibt das Minimum stabil bei 450 Trainingspunkten bei allen GP-Modellen. Die RQ und SE-Kovarianz zeigen erneut einen leicht besseren MSE als die Matérn-Kovarianz, wobei alle drei GP Modelle wie bei der 1- und 10-Schritt Vorhersage einen geringeren MSE als das Neuronale Netz liefern.

Die beste Vorhersage (RQ(150,450) aus **Tabelle 27**) kann im Gegensatz zur 1- und 10-Schritt Vorhersage die naive Vorhersage nicht verbessern. Dies deutet wie bei der 40-Schritt Vorhersage mittels Abbildung des Zeitpunkts auf einen random walk für langfristige Kursänderungen hin.

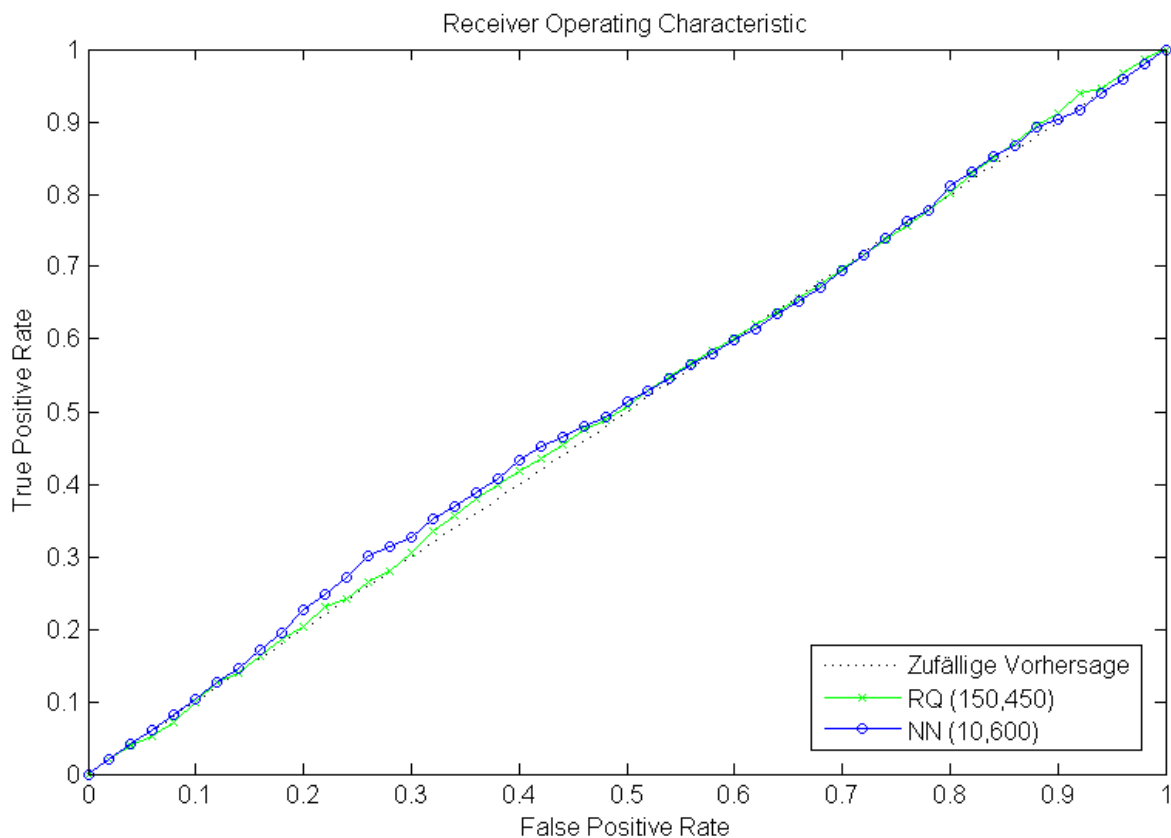


Abbildung 17: Receiver Operating Characteristic des GP mit Zustandsvektor und Benchmarks mit den geringsten MSEs für 40-Schritt Vorhersagen, zusätzlich zur verwendeten Kovarianzfunktion ist die Länge des Trainingsfensters und die Dimension des Zustandsvektors angegeben.

Modell	RQ (150,450)	NN (10,450)
AUC	0,5052	0,5096

Tabelle 30: Area-Under-Curve Werte für die ROC aus Abbildung 15

Bei den ROC zeigt sich wieder, dass das Neuronale Netz bessere Klassifikationen in positive und negative Änderungen liefert, als das beste GP Modell. Bemerkenswert ist jedoch, dass die AUC Werte zum Einen gegenüber der 10-Schritt Vorhersage ansteigen und im Gegensatz zu GPs mit Abbildung des Zeitpunkts über 0,5 liegen. Die Abbildung des Zustandsvektors ist also möglicherweise besser zur langfristigen Vorhersage geeignet. Dennoch muss angemerkt sein, dass die Ergebnisse, wenn überhaupt, nur sehr geringe Verbesserungen gegenüber einer zufälligen Vorhersage zeigen.

5 Fazit

In dieser Studienarbeit konnte gezeigt werden, dass eine gewisse Vorhersagbarkeit des Euro-Dollar Wechselkurses für 1- und 10- Schritt Vorhersagen mit Hilfe von Gauß-Prozessen, die eine Abbildung des Zeitpunkts lernen, besteht. Zum Einen konnte gezeigt werden, dass diese Vorhersage besser ist als eine zufällige, zum Anderen, dass die Vergleichsmethoden durchweg höhere MSE liefern.

Auch eine klassifizierende Vorhersage in positive und negative Änderungen am Folgetag ist mit Hilfe dieses GP-Modells möglich und liefert dabei auch nach dem Kriterium der ROC Ergebnisse, die einen Vorteil gegenüber der zufälligen Vorhersage zeigen. Die Klassifizierungen der Vergleichsmethode konnten durch einen Gauß-Prozess mit Abbildung des Zeitpunkts übertroffen werden.

Bei einem Gauß-Prozess mit Abbildung des Zustandsvektors waren die Ergebnisse gemessen am MSE ein wenig schlechter. Jedoch war die Wahl der Kovarianzfunktion bei Abbildung des Zustandsvektors für die Qualität der Vorhersage nicht so ausschlaggebend, was diese Methode robuster gegenüber einer fehlerhaften Spezifikation der Kovarianzfunktion macht. Der MSE der GP-Modelle war für jede Kovarianzfunktion geringer als der MSE der Vergleichsfunktion.

Der Gauß-Prozess mit Abbildung des Zustandsvektors lieferte jedoch weniger gute Klassifizierungen in Tage mit positiver oder negativer Änderung. Darüber hinaus mussten sich hier die GP-Modelle dem Vergleichsmodell geschlagen geben und bewegen sich, wie das Vergleichsmodell, sehr nah an der zufälligen Vorhersage.

Die Vorhersagbarkeit des Wechselkurses liefert starke Belege gegen die random walk hypothesis. Da der naive Benchmark der Beste sein müsste, wenn der Wechselkurs ein random walk wäre, kann gezeigt werden, dass dies nicht der Fall ist. Ob es jedoch profitabel ist, auf die Vorhersagen der hier untersuchten GP-Modelle zu spekulieren hängt von zahlreichen anderen Faktoren ab, hierbei sind sicherlich die Transaktionskosten entscheidend.

Da vor allem bei der Abbildung des Zeitpunkts gute Vorhersagen getroffen werden konnten, dürfte es interessant sein, diesen Ansatz weiter zu verfolgen. Erfolgsversprechend könnte dabei sein, Kovarianzfunktionen zu finden, die zur Vorhersage des Wechselkurses noch besser geeignet sind. Weiterhin könnten neben dem Zeitpunkt auch noch andere Informationen in die Vorhersage einfließen, wie Daten zur Wirtschaftsentwicklung oder Zinsunterschiede zwischen den Währungsräumen.

Quellen

- [1] Bank für internationalen Zahlungsausgleich, "Triennial Central Bank Survey - Report on global foreign exchange market activity in 2010," BIS, Basel, 2010.
- [2] Europäische Kommission. (13, July) Website der Europäischen Kommission. [Online]. <http://ec.europa.eu/trade/creating-opportunities/bilateral-relations/countries/united-states/>
- [3] Louis Bachelier, "Théorie de la spéculation," *Annales scientifiques de l'École Normale Supérieure*, 1900.
- [4] Fischer Black and Myron Scholes, "The pricing of options and corporate liabilities," *The Journal of Political Economy*, vol. 81, no. 3, pp. 637-654, May-Jun. 1973.
- [5] Eugene Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," vol. 25, no. 2, 1970.
- [6] Benoît B. Mandelbrot and Richard L. Hudson, *The (Mis) Behaviour of Markets: A Fractal View of Risk, Ruin and Reward.*, 2004.
- [7] Robert Sobel, *Panic on Wall Street: A Classic History of America's Financial Disasters- With a New Exploration of the Crash of 1987.*, 1988.
- [8] Grant McQueen and Steven Thorley, "Are Stock Returns Predictable? A Test Using Markov Chains," *Journal of Finance*, vol. XLVI, no. 1, March 1991.
- [9] Francis Tay and Lijuan Cao, "Application of support vector machines in financial time series forecasting," *The International Journal of Management Science*, pp. 309-317, March 2001.
- [10] R. Gencay and Min Qi, "Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 726-734, July 2001.
- [11] J. Galindo and P. Tamayo, "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications," *Computational Economics*, vol. 15, no. 1-2, pp. 107-143, 2000.
- [12] Alexander Gammerman and Vladimier Vovk, "Hedging Predictions in Machine Learning," *The computer Journal*, vol. 50, no. 2, pp. 151-163, February 2007.
- [13] Carl Edward Rasmussen and Christopher K.I. Williams, *Gaussian Processes for Machine Learning.*: MIT Press, 2006.
- [14] Bernhard Schölkopf and Alexander J. Smola, *Learning with Kernels.*: MIT Press, 2002.
- [15] Rob J. Hyndman and Yeasmin Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, vol. 27, no. 3, July 2008.
- [16] Lutz Kilian and Mark P. Taylor, "Why is it so difficult to beat the random walk forecast of exchange rates?," *Journal of International Economics*, vol. 60, no. 1, pp. 85-107, May 2003.

- [17] Barbara Rossi, "Are Exchange Rates Really Random Walks? Some Evidence Robust To Parameter Instability," *Macroeconomic Dynamics*, vol. 10, no. 1, pp. 20-38, February 2006.