

Vollautomatische Kalibrierungsmethode
hochauflösender Satellitenbildzeitreihen mittels
europaweiten MODIS Zeitreihen

Paul J. Arzberger

1. August 2021

Inhaltsverzeichnis

1	Motivation	3
2	State of the Art	4
2.1	Filterung verrauschter Satellitenbildzeitreihen	4
2.1.1	Methoden der Zeitreihenfilterung	4
2.1.1.1	Savitzky Golay	4
2.1.1.2	Asymmetrische Gauss'scher Fit	5
2.1.1.3	Doppel-Logistische Funktion	6
2.1.1.4	Fourier-Analyse	7
2.1.2	Vergleich unterschiedlicher Fittingmethoden anhand verrauschter Satellitendatenzeitreihen	7
2.2	Kalibrierung	11
2.2.1	Absolute Kalibrierung	11
2.2.1.1	Pseudo Invariante Feature Sites - PICS . . .	11
2.2.1.2	6S-Modell	16
	Absorptions- und Streuungseffekte	16
2.2.1.3	Bidirektionale Reflektionsverteilungsfunktion	19
2.2.1.4	MODIS (Collection V5) BRDF/albedo pro- duct	20
2.2.2	Relative Kalibrierung	22
2.2.2.1	Lineare Regression	22
2.2.2.2	PIF's	22
2.3	Bestehende Software	22
2.3.1	LEDAPS - Landsat Ecosystem Disturbance Adaptive Processing System	22
2.3.2	TIMESAT	25
3	Methodik	27
3.1	Datengrundlagen	27

3.1.1	MODIS - Allgemeine Informationen zum Sensor und Trägersystem	27
3.1.2	MODIS - Compositemethoden	28
3.1.2.1	Maximum Value Composite	28
3.1.3	Landsat - Allgemeine Informationen zum Sensor und Trägersystem	29
3.1.4	Sentinel	30
3.1.5	Sinusodial Referenzsystem	31
3.1.6	UTM-Referenzsystem	32
3.2	Datenakquisition	32
3.2.1	MODIS	34
3.2.2	Landsat	34
3.3	mathematische Grundlagen	35
3.3.1	statistische Kennzahlen	35
3.3.2	Lineare Regression	35
3.3.2.1	OLS - das klassische lineare Modell	35
3.3.3	Harmonische Schwingungen - Fourier Analyse	37
3.3.3.1	Harmonische Schwingungen , Diskrete Fourier Transformation und Analyse	37
3.3.3.2	Rekonstruktion mittels FFT	39
3.3.4	Anwendung Savitzky-Golay Filterung	41
3.3.4.1	Polynom 2. Grades	41
3.3.4.2	Gapfilling	43
3.3.4.3	Verspeicherung der Zeitreihe	44
3.3.5	Ausreisserdetektion	44
3.4	Verwendete Sprachen und Programme aus dem Bereich Data Science sowie Geoinformatik	45
3.4.1	Softwareentwicklung mit Python	45
3.4.1.1	Python 3.8	45
3.4.1.2	Python 3.8 - Multiprocessing	46
3.4.1.3	Python 2.7	46
3.4.1.4	Anaconda	46
3.4.2	ArcGIS, QGIS	47
3.4.3	Betriebssysteme, Speicherplatz und Rechenlogistik	47

Kapitel 1

Motivation

Zeitreihenanalysen in der Fernerkundung, mit Schwerpunkt zur Change Detection, setzen eine durchgehende, wenn möglich dicht besetzte Ansammlung aus Daten eines oder mehrere Sensoren voraus, um aus dem gesammelten Material einen Informationsgewinn ableiten zu können.

Die Sensoren um das Landsat System stellen einen derartigen Datenfundus zur Verfügung. Die historisch gesammelten Daten von Landsat 5 und 7 sowie von Landsat 8 stellen einen Datenpool von über 30 Jahren Fernerkundungsdaten dar welcher zur Zeitreihenanalyse herangezogen werden kann. Eine weitere Bestrebung kontinuierlich Information über den Zustand der Erde zu liefern stellt das MODIS (Moderate Resolution Imaging Spectrometer) System mit seinen beiden Satelliten Terra und Aqua dar. In Zukunft werden auch Daten von Sentinel 2 und 3 zur Verfügung stehen welche hochauflösende Daten in periodischen Abständen aufnehmen und im optischen wie im Frequenzband X Daten liefern werden. Anhand der dichten MODIS-Zeitreihe soll nun die Möglichkeit einer vollautomatischen relativen Kalibrierung von Satellitensensoren wie Landsat an die MODIS-Zeitreihe untersucht werden. Die Kalibrierung stellt eine Prototypenversion, dar welche portabel auf andere Systeme entwickelt werden soll, insbesondere mit Hinblick auf Sentinel 2 und 3. Die entwickelte Software soll rein auf Open Source Produkten /Modulen basieren. In die Kalibrierung fließen nicht nur die Rohdaten sondern auch von LEDAPS generierte Wolkenmasken, welche als Qualitätsparameter und Gewichtung herangezogen werden ein. Die zugrundeliegende MODIS-Zeitreihe wurde zudem noch zusätzlich gefiltert und stellt eine wolkenfreie, mit nahezu keinen NoData-Lücken, europaweite Zeitreihe dar welche zur Kalibrierung herangezogen werden kann. Die Güte der Kalibrierung selbst wird mit einem Qualitätsreport beschrieben.

Kapitel 2

State of the Art

2.1 Filterung verrauschter Satellitenbildzeitreihen

Satellitenbilddaten können durch atmosphärische Einflüsse wie Dunst, nicht identifizierte Wolkenpixel, Aerosole, gasförmigen Dämpfern etc zusätzlich zu den Effekten des Einfallswinkel und der Illumination verfälscht sein. In den meisten Studien über Phänologie sowie der Bestimmung von phenologischen Parametern basierend auf Satellitenbilddaten wird als erster Schritt die verrauschte Zeitreihe in eine gefittete Zeitreihe übergeführt.

2.1.1 Methoden der Zeitreihenfilterung

2.1.1.1 Savitzky Golay

Das dem Savitzky Golay Filter zugrunde liegende Polynom ist zweiter Ordnung und die Breite des Moving-Windows der zur Berechnung herangezogenen Daten wird vom User eingegeben.

$$f(t) = a_0 + a_1 \cdot t + a_2 \cdot t^2 \quad (2.1)$$

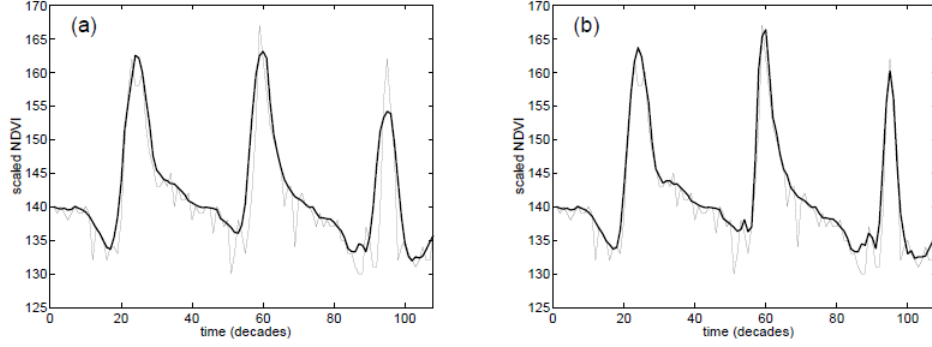


Abbildung 2.1: Gefittete Funktion nach Savitzky Golay. Die dünne Linie repräsentiert die Rohdaten, die dicke Linie repräsentiert die gefittete Funktion. In (a) ist die Moving-Window-Größe $n=5$, in (b) ist die Moving-Window-Größe $n=3$ (Eklundh L., Jönsson P., 2012)

2.1.1.2 Asymmetrische Gauss'scher Fit

$$f(t; x_1, \dots, x_5) = \begin{cases} \exp \left[- \left(\frac{t-x_1}{x_2} \right)^{x_3} \right] & \text{if } t > x_1 \\ \exp \left[- \left(\frac{x_1-t}{x_4} \right)^{x_5} \right] & \text{if } t < x_1 \end{cases} \quad (2.2)$$

$$t \quad \dots \quad \text{Zeit} \quad (2.3)$$

$$x_1 \quad \dots \quad \text{Position des Maximum oder Minimums} \quad (2.4)$$

$$x_2 \quad \dots \quad \text{Breite rechte Seite} \quad (2.5)$$

$$x_3 \quad \dots \quad \text{Wölbung rechte Seite} \quad (2.6)$$

$$x_4 \quad \dots \quad \text{Breite linke Seite} \quad (2.7)$$

$$x_5 \quad \dots \quad \text{Wölbung linke Seite} \quad (2.8)$$

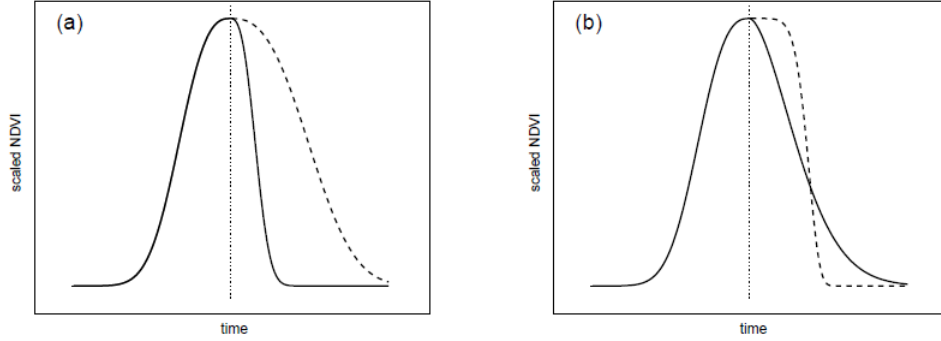


Abbildung 2.2: Auswirkungen auf die Breite und Wölbung wenn sich die lokalen Parameter ändern. In (a) wurde x_2 vermindert (durchgezogen) und erhöht (strichliert), dass sich in einer geänderten Breite und Wölbung auf der rechten Seite auswirkt. In (b) wurde x_3 vermindert (durchgezogen) und erhöht (strichliert). (Eklundh L., Jönsson P., 2012)

2.1.1.3 Doppel-Logistische Funktion

$$f(t; x_1, \dots, x_4) = \frac{1}{1 + \exp\left(\frac{x_1 - t}{x_2}\right)} - \frac{1}{1 + \exp\left(\frac{x_3 - t}{x_4}\right)} \quad (2.9)$$

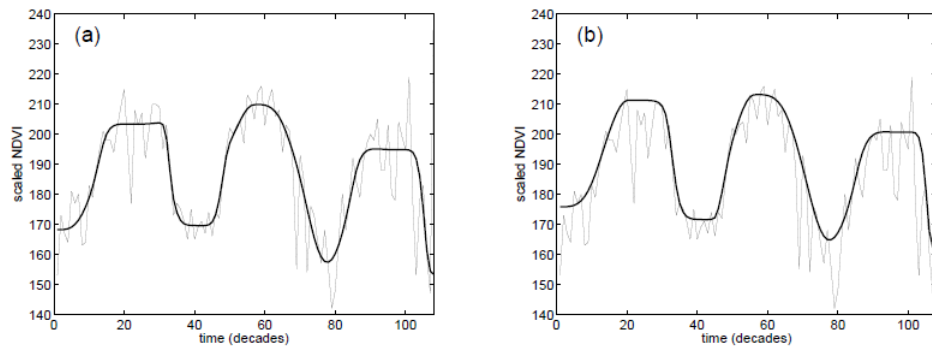


Abbildung 2.3: Gefittete Funktion nach mehreren Iterationen. Die dünne Linie repräsentiert die Rohdaten, die dicke Linie repräsentiert die gefittete Funktion (a). In (b) die erneute Berechnung des Modells mit einer verringerten Gewichtung der niedrigeren Datenwerte (Eklundh L., Jönsson P., 2012)

2.1.1.4 Fourier-Analyse

Die Diskrete Fourier Transformation ist gegeben durch

$$F(u) = \frac{1}{N} \sum_{t=0}^{N=1} VI(t) \cdot e^{-2\pi ut/T} \quad (2.10)$$

wobei $VI(t)$ in diesem Fall der einzusetzende Wert aus der Zeitreihe zum Zeitpunkt t , n die Anzahl der Fourier-Komponenten, t der Periodenindex, T die Gesamtperiodenanzahl wobei diese Zahl gleich der Gesamtanzahl an Epochen in der Zeitreihe. Die obere Gleichung besteht eigentlich aus zwei Zeilen: einem Kosinus-Teil (Realteil) sowie einem Sinus-Teil (Imaginärteil). Die Gleichung für den Kosinus lautet:

$$F_c(u) = \frac{1}{N} \sum_{t=0}^{N=1} \left(VI(t) \cdot \cos \left(2\pi \frac{ut}{T} \right) \right). \quad (2.11)$$

Die Gleichung für den Sinusteil setzt sich wie folgt zusammen:

$$F_s(u) = \frac{1}{N} \sum_{t=0}^{N=1} \left(VI(t) \cdot \sin \left(2\pi \frac{ut}{T} \right) \right). \quad (2.12)$$

Aus diesen beiden Gleichungen lassen sich die Phase F_p und die Magnitude F_m berechnen:

$$VI^*(t) = F_m(0) + \frac{1}{N} \sum_{t=0}^{N=1} \left(F_m(n) \cdot \cos \left(\frac{2\pi nt}{T} - F_p(n) \right) \right) \quad (2.13)$$

2.1.2 Vergleich unterschiedlicher Fittingmethoden anhand ver- rauschter Satellitendatenzeitreihen

Atkinson et al. (2012) wenden anhand des MERIS Terrestrial Chlorophyll Index (MTCI), ein Vegetationsindex der linear vom Chlorophyllgehalt in überschirmten Flächen abhängt und aus einer MERIS-Zeitreihe abgeleitet wurde, die Diskrete Fourier Transformation, eine Asymmetrische Gauss Funktion, eine Doppel-Logistische Funktion sowie einen Whittaker Smoother auf ihre MTCI-Zeitreihe an. Generell sei vorweg gesagt, dass jede Methode ihre Vor- und Nachteile mit sich bringt, wie sie auf Messfehler oder Ausreisser reagiert beziehungsweise zur Interpretation herangezogen werden kann (de Beurs K.M., Henebry G.M., 2010).

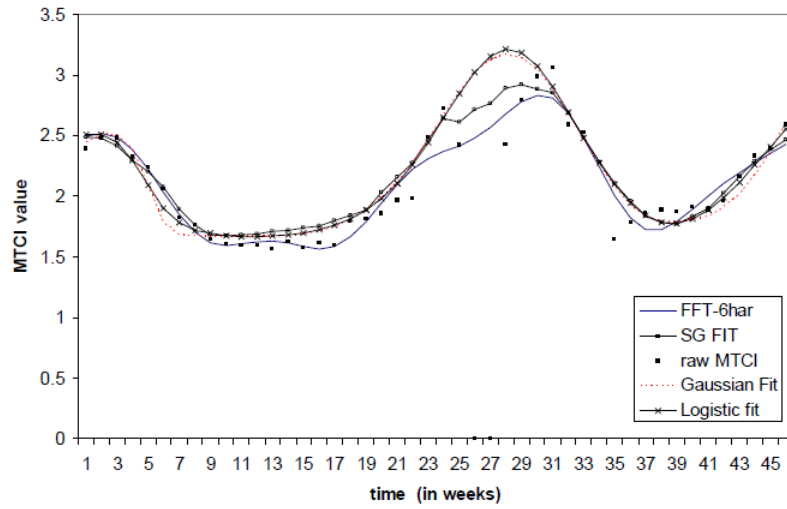


Abbildung 2.4: Vergleich der Fittingmethoden Savitzky-Golay (SG), Gauss'scher Fit, Logistischer Fit (hier mit TIMESAT) und Fourier (FFT) für eine landwirtschaftlich genutzte Fläche in Indien (Atkins et al., 2009)

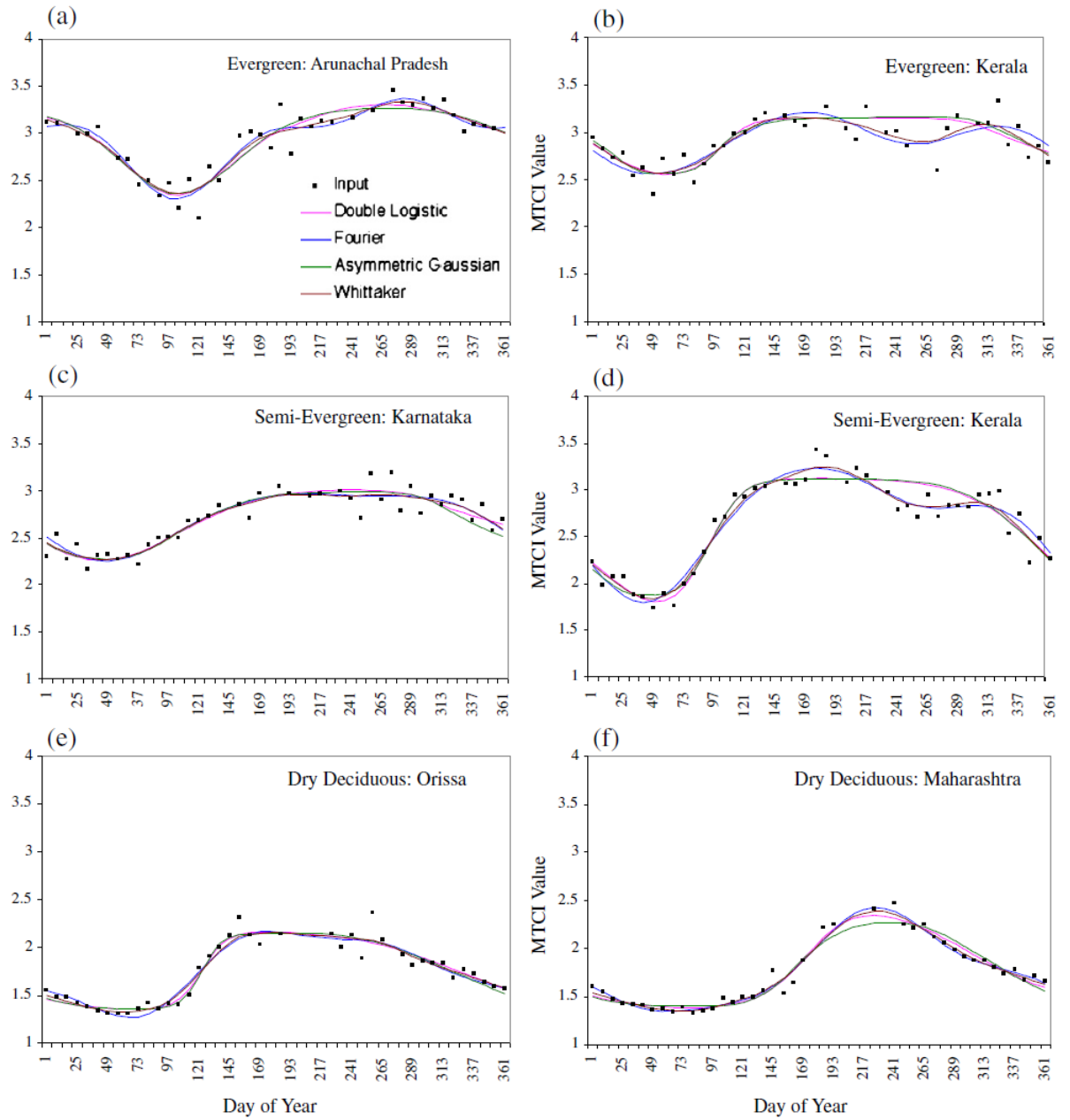


Abbildung 2.5: Vergleich der Fittingmethoden Savitzky-Golay (SG), Gauss'scher Fit, Logistischer Fit (hier mit TIMESAT) und Fourier (FFT) anhand von immergrünen Flächen (a,b), semi-immergrünen Flächen (c,d), trockene Laubwälder (e,f) (Atkins et al., 2012)

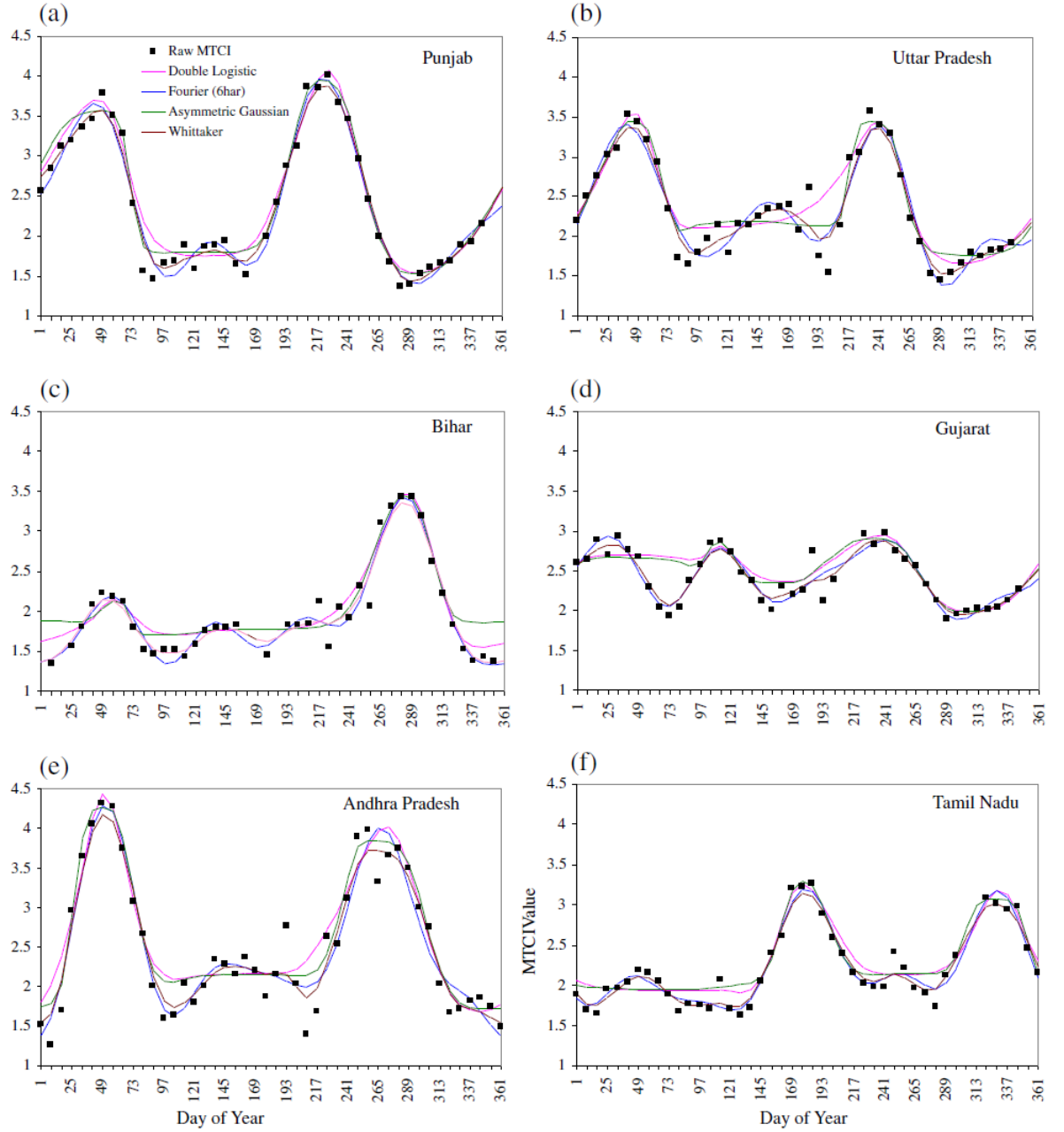


Abbildung 2.6: Vergleich der Fittingmethoden Savitzky-Golay (SG), Gauss'scher Fit, Logistischer Fit (hier mit TIMESAT) und Fourier (FFT) anhand von landwirtschaftlich genutzten Flächen - Punjab (a), Uttar Pradesh (b), Bihar (c), Gujarat (d), Andhra Pradesh (e), Tamil Nadu (f) (Atkins et al., 2012)

2.2 Kalibrierung

2.2.1 Absolute Kalibrierung

Unter der absoluten radiometrischen Kalibrierung versteht man die Umrechnung von Digital Numbers (DN) eines Satelliten in physikalische Einheiten in „am Sensor“ spektrale Strahlung $[Wm^{-2}sr^{-1}\mu m^{-1}]$. DN eines Sensors haben keine Gemeinsamkeiten mit einem anderen Aufnahmesystem. Die Konvertierung zu „am Sensor“ eintreffende Spektrale Information und Top-Of-Atmosphere Reflexion (TOA) sind fundamentale Vorgehensweisen um die Daten/Produkte zweier unterschiedlicher Sensoren zu vergleichen.

2.2.1.1 Pseudo Invariante Feature Sites - PICS

Es existiert eine Vielzahl an Möglichkeiten von pre-launch laborgetätigten Kalibrationen und post-launch on-orbit Kalibrierungen. Eine der meistverwendeten on-orbit Kalibrierungsmethoden ist die Verwendung von Pseudo Invarianten Calibration Sites (PICS). PICS sind radiometrisch stabile Orte und deren ursprünglicher Zweck der Überwachung der radiometrischen Stabilität eines Sensors mit einem hohen Genauigkeitsgrad. Jene Sites zeigen auch hohes Potential unter ähnlicher Genauigkeit zur absoluten Kalibrierung herangezogen zu werden. (5% Genauigkeit bei Kreuzkalibrierung MSS und Landsat 5 TM laut Helder et al., 2013). Die zeitliche Stabilität der Libya 4 - Kalibrierungssite wurde anhand des HYPERION EO-1, einem hyperspektralem Erdbeobachtungssensor in hochtransmittiven Spektren des SWIR (bei 1628 nm Wellenlänge ist die Atmosphäre nahezu transparent) untersucht, (siehe Helder et al, 2013)

1. Sonora Desert - Nord Amerika
2. Lybia 4
3. Sahara
4. Railroad Valley - U.S.A. Nevada
5. Ivanpah Playa - U.S.A. Kalifornien

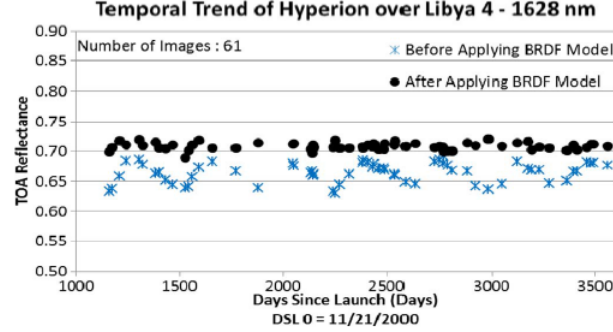


Fig. 3. TOA reflectance temporal trend of hyperion over Libya 4. BRDF was corrected to a nadir solar zenith angle for ease of viewing.

Abbildung 2.7: TOA Reflexion zeitlicher Trend von HYPERION über der Libya 4 Testsite und anhand einer BRDF korrigiert (Helder et al., 2013)

Helder et al wenden zuerst wird ein vereinfachtes lineares BRDF-Modell auf die Libya Testsite an. In diesem Modell wird lediglich der solare Zenithwinkel berücksichtigt. Sämtliche anderen Einflüsse werden vernachlässigt.

$$\rho_{Libya4}(\lambda, SZA) = k(\lambda)\rho_h(\lambda) [a(\lambda)(SZA) + 1] \quad (2.14)$$

TOA-Reflexion der Libya4 PICS wird von $\rho(\lambda)$ repräsentiert, $k(\lambda)$ ist der Skalierungsfaktor für das Hyperionspektrum, $\rho_h(\lambda)$ ist ein MODIS-Skalierungsfaktor und $a(\lambda)$ repräsentiert den linearen Koeffizienten welcher für die Korrektur des solaren Zenithwinkels benötigt wird. Die modellierten Werte und die gemessene Beobachtungen unterscheiden sich dabei um 1.3% (Helder et al., 2013) Um das Modell zu validieren wurde es auf Landsat 7 ETM+ Daten angewandt. Bei Landsat7 ETM+ wird die radiometrische Genauigkeit/Stabilität auf 5% angegeben. Da Landsat ein „am Sensor“ Messlösung ist muss die spektrale Welle die am Sensor gemessen wird, mittels eine Standartprozedur in die TOA übergeführt werden.

$$\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\Theta_s} \quad (2.15)$$

ρ_λ repräsentiert die Oberflächenreflexion [einheitslos], L_λ repräsentiert die spektrale Welle direkt am Sensor $[W/(m^2sr\mu m)]$, d Erd-Sonnen-Distanz [astronomischeEinheit], $ESUN_\lambda$ repräsentiert die mittlere atmopshärische Solarstrahlung $[W/(m^2sr\mu m)]$, Θ_s steht für den solaren Zenithwinkel. Die Kalibrierungen zeigen bei beiden Sensoren ein Ergebnis das innerhalb ihrer angegebenen Ungenauigkeit von 2% für MODIS und 5% für Landsat5 ETM+.

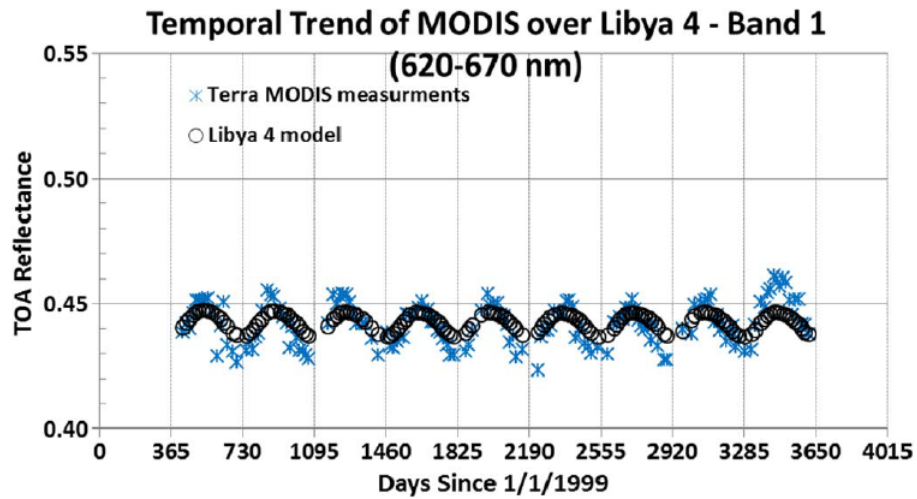


Abbildung 2.8: Gemessene MODIS und modellierte MODIS mittels linearer BRDF über Libya4 (Helder et al., 2013)

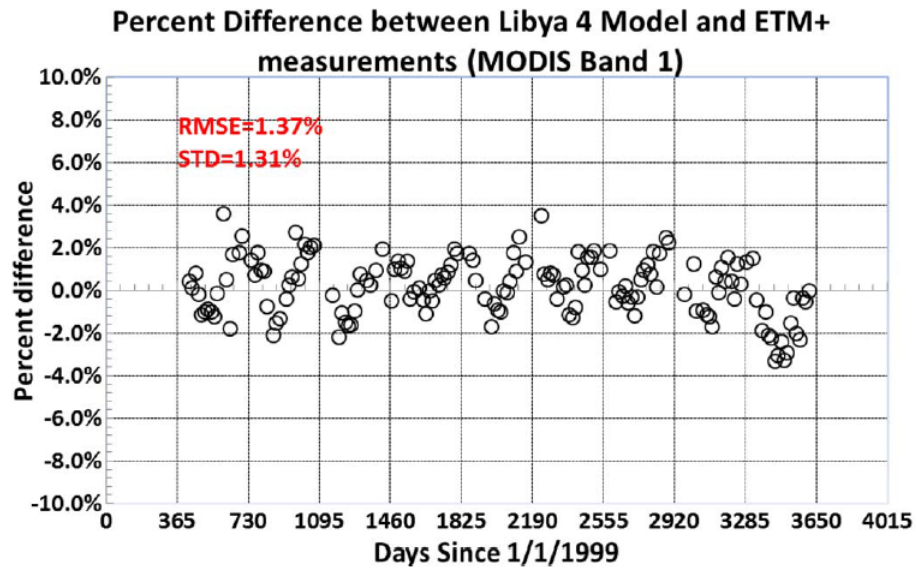


Abbildung 2.9: Prozent Differenz zwischen MODIS-Beobachtungen und Modellberechnungen (Helder et al., 2013)

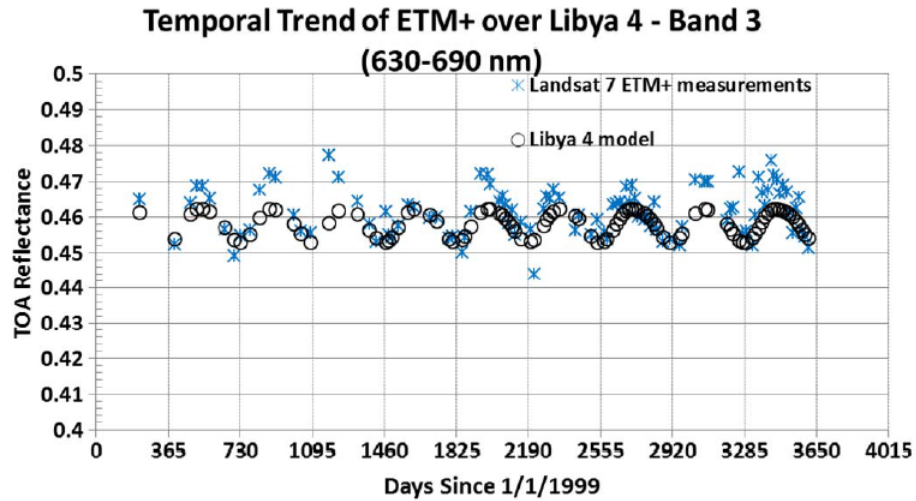


Abbildung 2.10: Gemessene ETM+ (asteriks) und modellierte Prediktionen (kreis) über Libya4 (Helder et al.,2013)

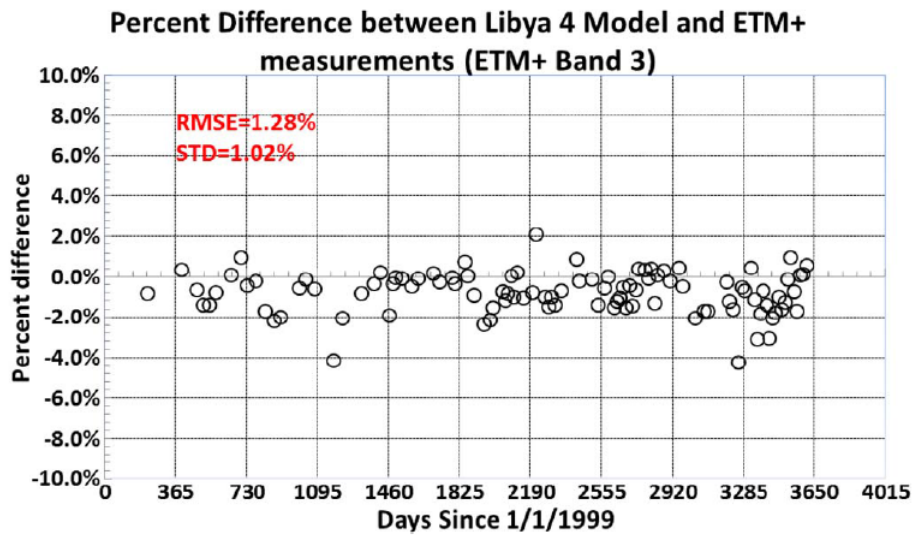


Abbildung 2.11: Prozent Differenz zwischen ETM+ -Beobachtungen und Modellberechnungen (Helder et al.,2013)

2.2.1.2 6S-Modell

In diesem Unterabschnitt wird grob die 6S Methode zur Berechnung eines empfangenen Signales beschrieben. Für eine genaue Diskussion wird auf die Literatur von Vermote et al., 2006, „Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV)“ hingewiesen.

Anhand der 6S (6S - Second Simulation of a Satellite Signal in the Solar Spectrum) radiometrischen Transfer Codes wird der Einfluss der Atmosphäre, speziell die Absorption und Streuung modelliert. Im Idealfall (mit keiner vorhandenen) Atmosphäre wird ein Teil der ausgesandten Photonen von der Erdoberfläche absorbiert und der Rest wieder in den Weltraum zurück reflektiert wird. Diese aufgefundenen Messungen selbst hängen direkt mit der Eigenschaft der Reflexionsoberfläche zusammen. bei vorhandener Erdatmosphäre Circa 80% der Photonen kommen bei einer Wellenlänge von $0.85\mu m$ und 50% bei $0.45\mu m$ an. Jene Lichtteilchen gehen bei zwei Prozessen , der Absorption (durch Gase, hauptsächlich O_2 , O_3 , H_2O , CO_2 , CH_4 sowie N_2O und der Streuung verloren. Die Absorption ist in der Regel klein gehalten und man vermeidet in der Fernerkundung Bänder in denen stark absorbiert wird. Die Streuung wird durch nicht absorbierende Moleküle verursacht. Das Photon wird dabei in eine andere Richtung reemittiert in die es eigentlich weiter gehen sollte. In der Regel durchläuft ein Photon mehrere solche Streuvorgänge, welche in der Verarbeitung mitberücksichtigt werden müssen.

Ein Teil der Strahlung wird direkt von der Atmosphäre zurück reflektiert ohne die Oberfläche zu erreichen somit tragen sie nur zum Strahlungshaushalt bei aber nicht zum Informationsgewinn über das Beobachtungsziel. Die überbleibenden Photonen beleuchten die Oberfläche abhängig von ihrem gestreuten Pfad diffus und werden auf ihrem Weg Richtung Weltraum bzw Sensor erneut gestreut, reflektiert und erneut gestreut (Trappingeffekt). Kurz es verlangt nach einer genauen Bestimmung der absorbierenden Gase der Atmosphäre sowie eine genaue Behandlung der Streuprozesse in Kombination beider Prozesse.

Absorptions- und Streuungseffekte

Die absorbierenden Gase Sauerstoff (O_2), Kohlendioxid (CO_2), Methan (CH_4) und Distickstoffmonoxid (N_2O) werden als Konstant in der Atmosphäre angenommen. Wasser (H_2O) und Ozon (O_3) werden als zeit- und orts-

abhängig betrachtet. Gase Absorbieren Strahlung indem sie ihre Rotations, Vibrations- sowie Elektronischen Zustände verändern. Veränderungen der Rotation werden im Mikrowellen oder fernen Infrarot sichtbar und sind typisch für schwache Frequenzen. Veränderungen in der Vibration sind im Nahen Infrarotbereich angesiedelt. Elektronische Veränderungen können im sichtbaren und ultravioletten Bändern sichtbar gemacht werden. Die Absorptionskoeffizienten variieren sehr stark mit der unterschiedlichen Wellenlänge und stellen ein sehr komplexes Konstrukt dar. H_2O beeinflusst hauptsächlich Wellenlängen die größer als $0.7\mu m$ sind. O_3 hat seine größte Absorptionsfähigkeit zwischen $0.55 - 0.65\mu m$ und limitiert die Erdbeobachtung auf Wellenlängen größer $0.35\mu m$. CO_2 beeinflusst Wellenlängen größer $1\mu m$ jedoch nicht so stark wie H_2O und wirkt störend auf den Wasserdampf. O_2 stärkster Einfluss ist um ca $0.7\mu m$, CH_4 zeigt besonders starken Einfluss bei 2.3 bzw bei 3.35, N_2O hat ebenfalls in zwei Bereichen einen starken Einfluss und zwar bei $2.9\mu m$ und $3.9\mu m$.

Bei der Annahme das die Oberfläche einheitliche lambertsche Reflexionseigenschaften aufweist und die Atmosphäre horizontal einheitlich geschichtet ist wird die Reflexion als

$$\rho^* = \frac{piL}{\mu_s E_s} \quad (2.16)$$

$$L \quad \dots \quad \text{gemessene Strahlung} \quad (2.17)$$

$$E_s \quad \dots \quad \text{solare Flux am Atmosphären-beginn} \quad (2.18)$$

$$(2.19)$$

Berücksichtigt man atmosphärischen Einfluss, Absorptionseffekte von H_2O , O_2 , CO_2 und O_3 , Streuung an Molekülen und Aerosolen und der Einfluss der Oberfläche, schreibt man nun (Vermote et al, 6S,2006)s

$$\rho'(\Theta_s, \Theta_v, \Phi_v) = t_g(\Theta_s, \Theta_v) \left\{ \rho_a(\Theta_s, \Theta_v, \Phi_v) + \frac{T(\Theta_s)}{1 - \langle \rho(M) \rangle S} \left[\rho_c(M) e^{-\tau/\mu_v} + \langle \rho(M) \rangle t_d(\Theta_v) \right] \right\} \quad (2.20)$$

Auf eine genaue Diskussion wird Um eine erfolgreiche Kalibrierung mittels der 6S Methode durchführen zu können benötigt man Informationen über

- geometrische Zustände der Strahlungsquelle und Sensor
- atmosphärische Modellierung der Gaskomponenten
- das Aerosolmodell
- spektralen Zustände
- die Bodenbeschaffenheit (Typ und Variation)

Für jeden Schritt in der Berechnung werden entweder standardisierte Zustände angenommen oder selbstdefinierte fließen ein. Die Satellitenposition wird generell als grob angenommen.

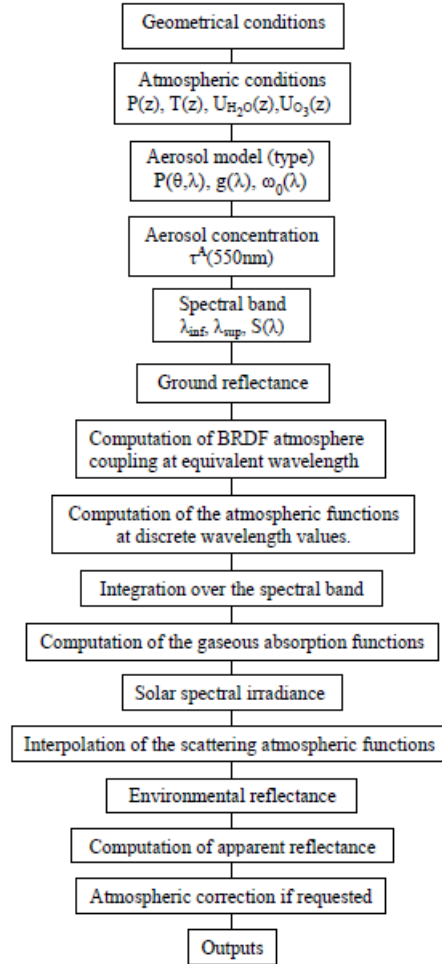


Abbildung 2.12: *Workflow der 6S Radiative Transformation Codes (Vermote et al., 2006)*

2.2.1.3 Bidirektionale Reflektionsverteilungsfunktion

Die Bidirektionale Reflektionsverteilungsfunktionen (BRDF) beschreiben die Reflektionseigenschaften von Materialien und Oberflächen und Berücksichtigt den Einstrahlwinkel des Emitters sowie den Blickwinkel des aufnehmenden Sensors. Verschieden Materialien reflektieren bei gleicher Beleuchtung unterschiedlich. Manche Oberflächen zerstreuen das einfallende Licht diffus während andere wiederum die Strahlung als Spiegel reflektieren. Generell ge-

ben sie an wie viel Licht von einer Oberfläche reflektiert wird abhängig von der pyhsikalischen Beschaffenheit (Wellenlänge) des ausgesendeten Lichtes sowie der Oberflächenbeschaffenheit und Zusammensetzung der reflektierenden Materie.

Die BRDF ist eine Funktion in vier Variablen

$$BRDF = f(\Theta_{in}, \Phi_{in}, \Theta_{ref}, \Phi_{ref}) = f(\mathbf{L}, \mathbf{V}) \quad (2.21)$$

wobei Θ_{in} , Φ_{in} , Θ_{ref} , und Φ_{ref} die Azimuth- und Deklinationswinkel der Einfalls- und Reflexionswinkel bezüglich der Strahlungsquelle sowie des Sensors sind. \mathbf{L} bezeichnet dabei die Richtung der Lichtquelle und \mathbf{V} die Richtung des Senors auf die Oberfläche. Ohne jetzt weitere Einflussparameter zu nennen oder in die Tiefe zu gehen siehe [N. Gebhardt] beschreibt einen BRDF für eine speziel definierte Wellenlänge das Verhältnis zwischen der differentiellen Leuchtdichte $RadiancedL_{ref}$ in Beobachtungsrichtung und der differentiellen Bestrahlungsstärke $IrradiancedE_{in}$ die aus der Beleuchtungsrichtung auf die beschienene Oberfläche wirkt, beschreibt.

$$BRDF(\Theta_{in}, \Phi_{in}, \Theta_{ref}, \Phi_{ref}) = \frac{dL_{ref}(\Theta_{ref}, \Phi_{ref})}{dE_{in}(\Theta_{in}, \Phi_{in})} \quad (2.22)$$

Allgemein ist die BRDF eine Funktion die nur in drei Variablen abhängig ist $f(\Theta_{in}, \Theta_{ref}, \Phi_{in} - \Phi_{ref})$. Als weiterführende Literatur wird Nikolaus Gebhardt und Shree Nayar empfohlen.

Die Oberflächenalbedo variiert sehr stark abhängig von sesonalen oder geographischen Parametern. Regenfall kann zum Beispiel die Oberflächenalbedo einer Baumkrone um bis zu 50% reduzieren (Román et al 2009). Zusätzlich dazu kommen noch Faktoren wie Schattenwurf durch Baumkronen oder spiegelnde Reflexionen abhängig vom Einfallwinkel des Sonnenlichtes oder atmosphärische Veränderungen und Schwankungen (Román et. al. 2009).

2.2.1.4 MODIS (Collection V5) BRDF/albedo product

Der Algorithmuis beruht auf multitemporalen, wolkenfreien, atmosphärisch korrigierten Oberflächenreflexionen der Terra und Aqua Satelliten um die

Anisotropie der Oberflächenreflexion über eine 16 Tage lange Periode zu modellieren. Verwendet wird dabei ein kernelbasiertes lineare Modell der BRDF auch bekannt als Ross-Lick/Li-Sparse-Reciprocal kernelartiges BRDF-Modell. Dieses Modell beruht auf den gewichteten Summen eines isotropischen Parameters sowie zwei Funktionen (oder Kernel) abhängig von Blickrichtung und Illumination. Θ steht für den solaren Zenithwinkel, ϑ Winkel unter dem der Pixel betrachtet wird, Φ ist der relative sonnengerichtet Azimuth in Blickrichtung des Senors und Λ ist das solarspektrale gewichtete Zentrum für ein gegebenes MODIS Band

$f_{iso}(\Lambda)$ ist die isotropische Streuungskomponente und gleich der bidirektionalen Reflexion bei einem Betrachtungswinkel von $\vartheta = 0$ und dem solaren Zenithwinkel $\Theta = 0$. $f_{geo}(\Lambda)$ ist der Koeffizient für des Li-Sparse-Reziproken geometrischen Streuungskerne K_{geo} welcher sich aus der Oberflächenstreuung und der geometrischen Schattenwurf Theorie herleitet. Der Parameter $f_{vol}(\Lambda)$ ist der Koeffizient des Ross-Thick Volumens-Streuungskernel K_{vol} , welcher sich aus radiativen Transfermodellen. Die best gefitteten Ross-Thick/Li-Sparse-Reziproken (RTLSR) Modellparameter werden an den ersten sieben MODIS Spektralbändern angebracht. Zusammengesetzt sieht das Modell wie folgt aus

$$R(\Theta, \vartheta, \Phi, \Lambda) = f_{iso}(\Lambda) + f_{vol}(\Lambda)K_{vol}(\Theta, \vartheta, \Phi) + f_{geo}(\Lambda)K_{geo}(\Theta, \vartheta, \Phi). \quad (2.23)$$

Anhand von Integration über alle Blickwinkel ergibt sich ein Modell für einen direkte hemisphärische Reflexion (DHR) oder Black-Sky-Albedo (BSA). Die Integration über alle Illuminationswinkel führt zur Bihemisphärischen Reflexion (BHR) unter isotropischer Illumination oder einer White-Sky-Albedo (WSA).

$$BSA(\Theta_s) = \sum_N f_k h_k(\Theta_s) \quad (2.24)$$

$$WSA = \sum_k f_k H_k \quad (2.25)$$

$h_k(\Theta)$ stellt das Integral des BRDF Kernels k über einen gegebenen Beobachtungszenith und (...weilerschreiben / vergleich der Integrationsparameter mit den Allgemeinen BRDF Parametern)

2.2.2 Relative Kalibrierung

2.2.2.1 Lineare Regression

(Tokol T. et al, 1999) gehen von zwei Grundlegenden Strategien aus, die bei der relativen Datenkalibrierung angewandt werden können. Manche Methoden basieren auf der Verteilung der Bildinformation während andere auf paarweise Pixelbeziehungen abziehen. Alle Regressionsmodelle wurden anhand von robusten Methoden berechnet, welche iterativ die Gewichte für die Regression neu berechnen. Um Band-zu-Band kalibrieren zu können wird folgendes Modell verwendet:

$$\hat{y}_i = \beta_0 + \beta_1 x_i \quad (2.26)$$

y_i ist hierbei ein Pixel eines Bandes aus einer nachfolgenden Epoche und x_i bezieht sich auf einen Pixel eines Bandes in einer früheren Epoche. (Vermote E.F., Saleous N.Z., 2006) führen eine Kreuzkalibrierung zwischen AVHRR und MODIS über einem radiometrischen stabilem Wüstengebiet im Niger anhand einer linearen Translation durch.

2.2.2.2 PIF's

In dieser Diplomarbeit werden als Ausgangsdaten der radiometrischen Kalibrierungsmethode mittels LEDAPS absolut kalibrierte Landsatdaten verwendet. Die Cross-calibration zweier Sensoren über einem radiometrisch stabilen Gebiet wird in der Literatur häufig angeführt. Bei diesen Pseudo-Invariant-Calibration Sites handelt es sich wiederholt um die selben Gebiete. Diese Sites wurden eigentlich dafür ausgesucht um die Langzeitstabilität eines Sensors zu überprüfen. Um die DN der Landsatdaten

2.3 Bestehende Software

2.3.1 LEDAPS - Landsat Ecosystem Disturbance Adaptive Processing System

Das Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) wurde ursprünglich im Jahre 2006 von der National Aeronautics and Space Administration - Goddard Space Flight Center und der Universität von Maryland entwickelt um Top-Of-Atmosphere (TOA) Reflexionsprodukte aus Daten der Landsat Thematic Mapper (Landsat5) und dem Enhanced Thematic Mapper Plus Level1 (Landsat7) Sensoren zu generieren. Der in dieser Diplomarbeit verwendete Algorithmus wurde vom USGS für den

Landsat Surface Climate Data Record angepasst. LEDAPS setzt sich aus sechs Modulen zusammen die drei Schlüsseloperationen beinhalten.

1. Konvertierung von Digital Numbers (DN) zu TOA Reflexionen
2. Wolkenpixeldetektion anhand der TOA
3. Transformation der TOA zu Oberflächenreflexion mittels zusätzlichen Daten

Die atmosphärische Korrektur wird anhand eines 6S-Korrekturmodells gerechnet welche Druck, Wasserdampf, Ozon sowie geometrischer Höheninformation.

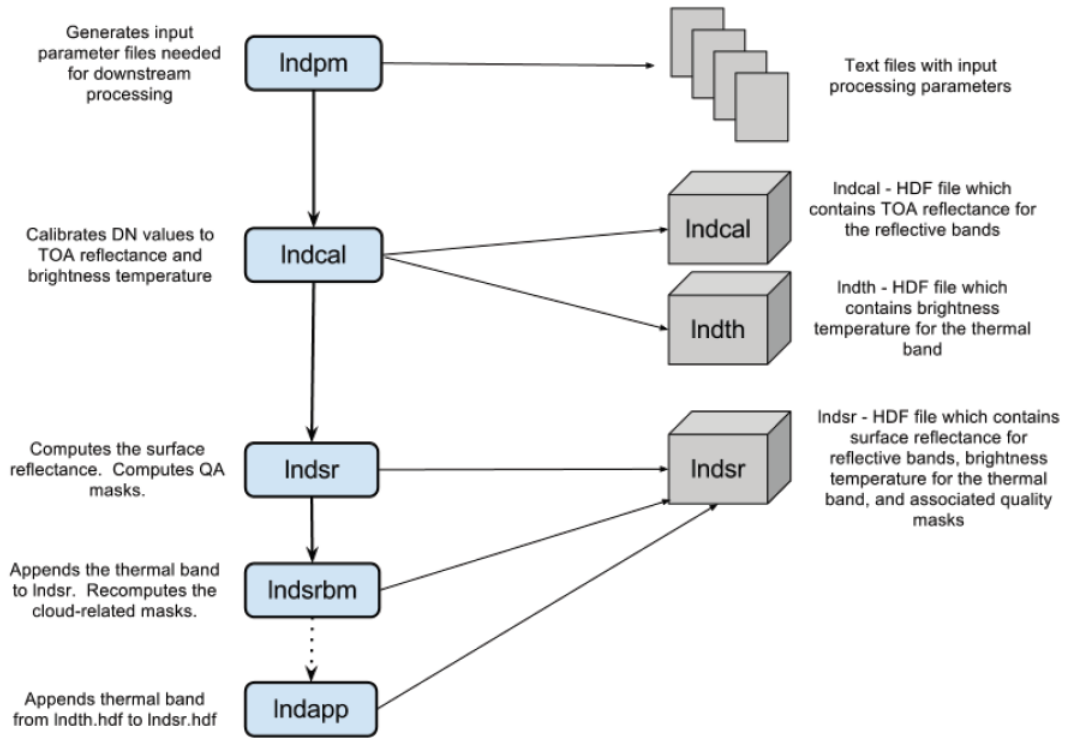


Abbildung 2.14: *Workflow des LEDAPS-Algorithmus, (U.S.G.S., 2013)*

Als Endergebnis liegen Landsat5 TM und Landsat7 ETM+ Daten als Oberflächenreflexion im hdf-Format vor.

2.3.2 TIMESAT

TIMESAT wurde primär zur Prozessierung von Vegetationsindexzeitreihen für die Ableitung von saisonalen Parametern wie Anfang, Ende, Länge etc der Vegetationsperiode entwickelt. Es erlaubt die Filterung von Satellitenbildzeitreihen sowie berücksichtigt es Zusatzinformation wie Qualitätsinformationen (z.B. Modis Qualitätsprodukt) für eine mögliche Gewichtung. Die verwendeten Algorithmen basierend auf der Methode der kleinsten Quadrate sind ein Savitzky Golay Filtern sowie zwei Methoden bei denen die Daten anhand von nonlinearer Funktionen unterschiedlicher Komplexität gefittet werden. TIMESAT arbeitet iterativ und mit einer dynamischen Gewichtung. Nach einer ersten Iteration und Schätzung der Parameter werden die Rohdaten-

werte unterhalb der gefitteten Kurve mit einem neuen geringeren Gewicht behaftet und das Modell wird neu berechnet. Dadurch resultiert eine an den oberen Verlauf der Daten adaptierte Modell. TIMESATs größter Nachteil ist die Prozessierungsgeschwindigkeit. In der Literatur wird TIMESAT meist nur für kurze Perioden oder einem beschränkten Gebiet eingesetzt siehe (Jönsson P., Eklundh L., 2004) sowie (Tan B. et al., 2011). Bei der Prozessierung einer Zeitreihe die Daten von mehr als einem Jahrzehnt umfasst steigt die Rechenzeit und das Programm wird nicht mehr brauchbar.

Kapitel 3

Methodik

3.1 Datengrundlagen

In diesem Kapitel werden die verwendeten Satellitendaten besprochen. Dabei handelt es sich um multispektrale Satellitendaten unterschiedlicher Auflösung und Datengehaltes sowie der Qualitätsinformation die bei der absoluten Kalibrierung der Landsatdaten mittels LEDAPS generiert wurde als fertiges Produkt für MODIS bereitgestellt wird.

3.1.1 MODIS - Allgemeine Informationen zum Sensor und Trägersystem

MODIS (Moderate Resolution Imaging Spectroradiometer) nennt sich ein optischer Sensor, welcher sich in einem Low Earth Orbit in 705 km Höhe, mit einer Blickwinkel von 2330 km auf den beiden Satelliten Terra (früher EOS AM-1) und Aqua (früher EOS PM-1) um die Erde befindet. Die Mission wurde am 18. Dezember 1999 mit dem Launch des Terra (EOS AM-1) Spacevehicles gestartet. Am 4. Mai 2002 wurde der Satellite Aqua (EOS PM-1) mit dem zweiten MODIS in seine Umlaufbahn geschossen. Die Lebenszeit von Terra und Aqua war mit jeweils 6 Jahren anberaumt und erfreut sich nun über ein mehr als 15 jähriges Bestehen. Die Orbits der Satelliten sind so synchronisiert, dass Terra den Äquator am Vormittag von Nord nach Süd passiert und Aqua den Äquator nachmittags von Süd nach Nord. Die räumliche Auflösung eines einzelnen Pixels reicht von 250 und 500 Metern im Nadir bis zu 1 km als Whisk-Broom-Scanner. Die Bänder 1-2 werden in 250m, die Bänder 3-7 in 500m und die Bänder 8-37 werden in 1000m Auflösung aufgenommen. Beide Satelliten vermögen in einer ein bis zwei Tage Periode die gesamte Erdoberfläche in 36 spektralen Bändern in einem Umfang von

0.4 μ m bis 14.4 μ m aufzunehmen. Die Aufnahmen finden dabei jeweils unter einem anderen Einstrahlwinkel, bidirektional, der Sonne statt. Nach der Anwendung von atmosphärischen Korrekturen und der Überführung der Strahlungswerten in Reflexionswerten der Erdoberfläche gibt das System die anisotropischen Reflexionseigenschaften mit dem Wandel der Landbedeckungstypen, Vegetation sowie Bodenkomponenten in jedem einzelnen Pixel wieder. Anhand von State-of-the-Art On-board-Kalibrierungseinheiten wird eine Genauigkeit von unter 2% angegeben (Vermote E.F., Saleous N.Z.,2006). MODIS Terra wird generell als einer der am besten kalibrierten Sensoren im Bereich der „reflektierenden“ Wellenlängen mit einer Unsicherheit von 2% (Helder et al., 2013).

BANDNAME	MODIS	nm	Landsat 7	nm	LANDSAT 5	nm	LANDSAT 8	nm
RED	Band 1	620-670	Band 3	630-690	Band 4	630-690	Band 5	630-680
NIR	Band 2	841-876	Band 4	750-900	Band 4	760-900	Band 5	845-885
BLUE	Band 3	459-479	Band 1	450-515	Band 1	450-520	Band 2	450-515
GREEN	Band 4	545-565	Band 2	525-606	Band 2	520-600	Band 3	525-600
MIR	Band 5	1230-1250	Band 6	1040-1250	Band 6	1040-1250	Band 9	1360-1390
MIR	Band 6	1628-1652	Band 5	1550 - 1750	Band 5	1550-1750	Band 6	1560-1660
MIR	Band 7	2105-2155	Band 7	2090-2350	Band 7	2080-2350	Band 7	2110-2300

Tabelle 3.1: Vergleich der aufzunehmenden spektralen Wellenlängen Bandweise MODIS zu Landsat Satelliten

3.1.2 MODIS - Compositemethoden

Der Grundgedanke des MODIS Compositionalgorithmus ist aus einem zeitlich vordefiniertem Pool an Epochen mittels einer reflexionsbasierten BRDF (Bi-directional Reflectance Distribution Function) mit einem eingeschränktem MVC (Maximum Value Composite) als Rückversicherung.

3.1.2.1 Maximum Value Composite

Das Maximum Value Composite (MVC) Verfahren zieht aus einem zeitlich vordefiniertem Pool für den NDVI den maximalen Wert jedes Pixels und trägt ihn in die abschliessende Epoche ein. Diese Vorgehensweise ist so geschaffen um störende atmosphärische Effekte zu minimieren , Wolken inkludiert. Der Kompositionsvorgang inkludiert in der Regel ein Wolken screening sowie einen Qualitätscheck der Daten. Die MVC funktioniert eher über Oberflächen mit Lambert’schen Eigenschaften, bei sich ändernden

atmosphärischen Eigenschaften und anisotropischem Oberflächenverhalten wird die Methode jedoch ungenau (Van Leeuwen et al., 1999). Nach atmosphärischer Korrektur nimmt die Ungenauigkeit der MVC noch zu da die Anisotropie der Oberflächen ebenfalls zunimmt. Generell arbeitet die MVC am besten bei atmosphärisch unkorregierten Daten und ist ungeeignet für die atmosphärisch korregierten Oberflächenreflexions Daten des MODIS-Sensors. Der NDVI tendiert zu einer Zunahme nachdem er aus atmosphärisch korregierten Daten berechnet wurde, jedoch bedeutet der höchste NDVI für ein Pixel noch nicht, dass es auch um den besten atmosphärisch korregierten Pixel handelt. Nicht entfernte Wolken und sowie unterschiedliche Sonnen- und Beobachtungswinkel verändern die Oberflächenreflexion und somit den ausgewählten maximalen Wert. Über das Jahr gesehen wäre somit der Pixel nicht konsistent (Van Leeuwen et al., 1999). Generell überschätzt das MVC-Verfahren die Rasterwerte und führt in weiterer Folge auch zu einer Überschätzung der abgeleiteten Parameters (Van Leeuwen et al., 1999). Wird der NDVI nun aus atmosphärisch korregierten Daten berechnet, so sind vor allem Pixel im Off-Nadir Bereich mehr fehlerbehaftet, da hier die Korrektur nicht so gut funktioniert. Einhergehend mit der immer weiter voranschreitenden Verbesserung und Entwicklung der atmosphärischen Korrekturalgorithmen wird bidirektionale Reflexion an Bedeutung gewinnen, da hier die unterschiedliche Struktur der Oberflächenbedeckung mit berücksichtigt werden kann.

3.1.3 Landsat - Allgemeine Informationen zum Sensor und Trägersystem

Landsat ist die weltweit längste zivile Erdbeobachtungsmission mittels Fernerkundungssensoren mittlerer Auflösung. Seit 1972 sind bis jetzt 8 Satelliten in Orbit gebracht worden wobei Landsat 6 beim Start verlorenging. Im näheren wird nur auf jene Sensoren eingegangen deren Daten in der Diplomarbeit verwendet wurden. Dies umfasst die Satelliten Landsat5, Landsat7 und Landsat8.

Landsat5 und Landsat4 (der Vollständigkeit halber erwähnt) wurden mit den verbesserten *Thematic Mapper* ausgestattet der im sichtbaren Bereich und nahen Infrarotbereich sowie im thermalen und panchromatischen Bereich Aufnahmen tätigt. Für das sichtbare und naheinfrarote Spektrum beträgt die Auflösung 30m, für das thermale Band 120m und im panchromatischen Kanal 15m.

Landsat7 führt den multispektralen *Enhanced Thematic Mapper Plus*

(ETM+) mit sich, welcher im sichtbaren Bereich und nahen Infrarotbereich sowie im thermalen und panchromatischen Bereich Aufnahmen tätigt. Für das sichtbare und naheinfrarote Spektrum beträgt die Auflösung 30m, für das thermale Band 60m und im panchromatischen Kanal 15m. Der Satellit führt drei on-board Kalibriereinrichtungen mit sich einen *Full Aperture Solar Calibrator*(FASC), einen *Partial Aperture Solar Calibrator* (PASC) und einen *Internal Calibrator* (IC)

Landsat8 wurde mit dem multispektralen Sensor *Operational Land Imager* (OLI) und dem *Thermal Infrared Sensor* (TIRS) ausgestattet. OLI liefert Aufnahmen in 8 spektralen Bändern in einer Auflösung von 30 m und einem panchromatischen Kanal in einer Auflösung von 15m. TIRS nimmt in zwei mittleren infraroten Bereichen in einer Auflösung von 100m auf. Da die TIRS Daten allerdings mit den OLI-Daten mitgeliefert werden, werden sie auf 30m neu aufgelöst.

Aktuell liefern Landsat7 und Landsat8 alle 16 Tage Aufnahmen des selben Erdoberflächenausschnittes. Die Orbits sind polare, sonnen-synchrone kreisförmige Bahnen in 705 km über Grund und ein Satellit umkreist die Erde in 99 Minuten. Die Bahnen der beiden Satelliten sind so synchronisiert das jeden achten Tag eine Aufnahmen des selben Beobachtungspunktes zur selben Uhrzeit vorliegt. Der beobachtete Groundtrack umfasst dabei einen 185 km breiten Streifen. Landsat5 hält bis jetzt den Rekord für den am längsten eingesetzten Erdbeobachtungssatelliten. Nach knapp 29 Jahren wurde seine Aktivität am 5. Juni 2013 eingestellt. Landsat7 weist seit dem 31 Mai 2003 Artefakte in seinen Satellitenbilddaten auf. Verursacht werden sie durch einen Defekt des *Scan Line Correctors* (SLC), der zur richtigen Abtastung des Beobachtungsgebietes bei Rotationsscannern eingesetzt wird. Seit diesem Ausfall finden sich zick-zack-artige Datenlücken in den Aufnahmen, die von einem zentralen nicht betroffenen Streifen horizontal nach außen laufen und von innen nach außen sich verbreitern.

3.1.4 Sentinel

Sentinel sind eine Reihe von Erdbeobachtungs- und Forschungssatelliten der ESA welche im Rahmen des Copernicus Forschungsprogramms gestartet und zum Einsatz kommen. Sie setzen sich aus 5 Satelliten zusammen die im optischen und Radarbereich eine dichte Bandbreite an Informationen über den Planeten sammeln. Sentinel 2 liefern Daten im optischen, nahen infraroten Bereich, hochauflösend von 10 mal 10 Meter, in einem Intervall von 5 Tagen. Dies gewährt Aussicht auf besonders dichte Zeitreihen die mit diesem System möglich sind. Sentinel 3 nimmt die Erdoberfläche mit einer höheren

Auflösung auf , soll allerdings zumindest auf dem Level der MODIS Daten kalibriert sein. Die entwickelte Methodik zielt auch auf eine portable Herangehensweise ab, dass man die besser kalibrierten Sentinel 3 Daten zum Kalibrieren von Satellitendaten anderer Systeme verwendet.

3.1.5 Sinusodial Referenzsystem

Die Daten der verwendeten MODIS-Zeitreihe wurden mittels dem MODIS-Reprojection-Tool in eine Sinusoidale Projection gebracht. Die Sinusoidal-Projektion ist eine flächentreue, pseudo-zylindrische Projektion, welche die gesamte Erdoberfläche darstellen kann. Der Zentralmeridian und der Äquator werden winkel- und längentreu wiedergegeben während Meridiane als Sinuskurven und Breitenkreise als Geraden dargestellt werden. Als Basis dient eine Kugel mit einem gemittelten Erdradius von 6371007,181 m (MODIS Land, web-quelle nr 11). Je weiter man sich vom Abbildungsmeridian entfernt umso verzerrter wird die Darstellung. Die meisten MODIS Land Produkte sind als Kacheln indiziert und nutzen das Sinusodial-Grid-Tilling-System. Am Äquator weist eine Kachel eine Größe von 10 Grad mal 10 Grad auf. Das Kachelkoordinatensystem startet bei (0,0) - (horizontale nummer, verticale nummer) in der linken oberen Ecke beginnt und bei (35,17) in der rechten unteren Ecke endet. Das System zählt dabei von rechts nach links (horizontal) und von oben nach unten (vertikal)

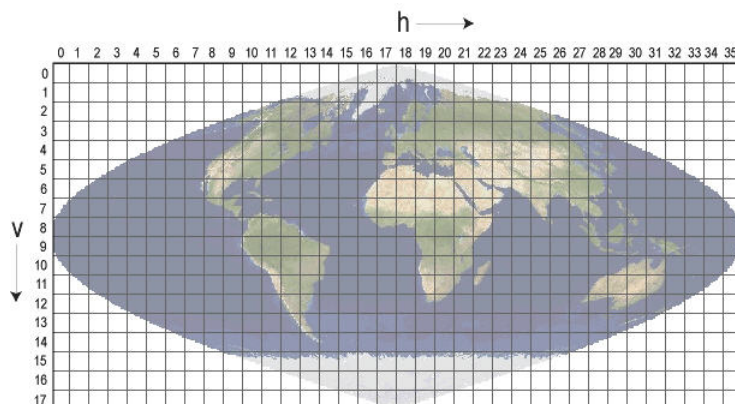


Abbildung 3.1: *MODIS Tiling System, (U.S.G.S., 2013)*

3.1.6 UTM-Referenzsystem

Sämtliche Daten der Landsatmissionen sind im *Universal Transverse Mercator*-Referenzsystem (UTM) projiziert. Es handelt sich hierbei um ein globales Koordinatensystem das die Erdoberfläche von 80 Grad Süd bis 84 Grad Nord streifenförmig in 6 Grad breite Zonen unterteilt. Als Basis dient das Bezugsellipsoid WGS84. UTM nutzt eine transversale Mercator-Projektion zum verebnen. Hierbei berührt der Projektionszylinder die Oberfläche nicht am Bezugsmeridian sondern schneidet sie. Dadurch entsteht ein Streifen

3.2 Datenakquisition

Die rohen MODIS-Zeitreihe liegt als verrauschte Zeitreihe von 2000-2012 im hdf-4 Format vor und wurden mittels der NASA-eigenen Software „MRT-TOOL“ in das GeoTIFF-Format entpackt und in das Sinusodial Reference System transformiert. Die Zeitreihe an sich selbst, setzt sich aus 16-Tägigen Kompositionen der besten spektralen Einzelaufnahme eines Pixels zusammen. Abhängig von der Auflösung des verwendeten Produktes zum Beispiel im 250 Meter aufgelöstem Produkt kann auch zwischen den Aufnahmen des Terra Satelliten mit dem Postfix ”MOD” und dem Aqua Satelliten mit dem Postfix ”MYD” unterschieden werden, welches zusätzlich in der Programmierung bedacht werden muss. Aufgrund der unterschiedlichen Launchzeitpunkte der Satelliten wurde die Zeitreihe zwischen 2000 und der ersten Aqua-Aufnahme künstlich verdichtet, indem die Terra-Aufnahmen einfach verdoppelt wurden und einen passenden Zeitstempel erhalten haben. Die komplettierte Zeitreihe an sich selbst weist alle 8 Tage eine Epoche eines MODIS-Sensors (Terra und Aqua-Epochen immer abwechselnd) auf. Für die Diplomarbeit selbst beschränkt sich der Author auf das „500m“ aufgelöste Produkt MCD43A4 welches alle 8 Tage Daten des Terra- oder des Aqua-Satelliten inklusive die passende Qualitätsinformation zur Verfügung stellt. Trotz des einwandfreien Rufes einer perfekten radiometrischen Korrektur der MODIS-Daten wurden Störungen in der visuellen Kontrolle ausfindig gemacht und die Zeitreihe anhand eines Glättungsalgorithmus nach Savitzky-Golay verbessert. Somit liegt eine europaweite wolkenfreie MODIS-Zeitreihe in 500 Meter Pixelauflösung als Basis für die relative Kalibrierung der Landsatdaten bereit.

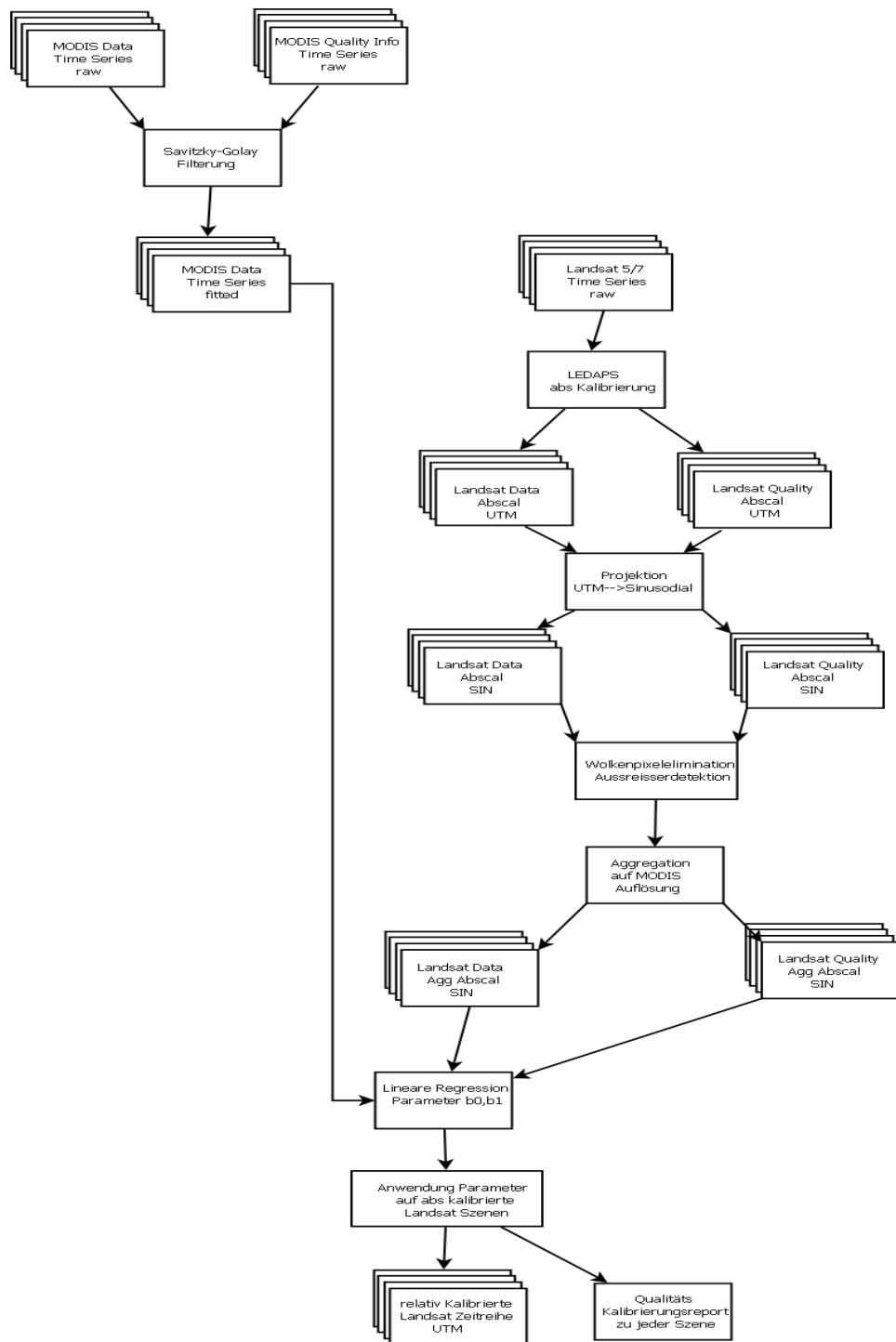


Abbildung 3.2: Workflow der relativen Kalibrierung von Landsatdaten mittels gefilterter MODIS Zeitreihe

3.2.1 MODIS

Die 500m aufgelösten MODIS-Daten sowie deren epochenzugehörige Qualitätsinformation im hdf4-Format wurden von einem mittlerweile geschlossenen FTP Server der NASA mittels einem Python FTP-Skript heruntergeladen und epochenweise lokal als hdf4-Stack gespeichert. Das hdf4-Format ist ein Containerformat welches vor allem in der wissenschaftlichen Community weit verbreitet ist. Das Datenprodukt trägt die Kennzeichnung MCD43A4 und das zugehörige Qualitätsprodukt die Kennzeichnung MCD43A2. Aus den hdf4-Stacks wurden mittels Python-Skript für die 500m Daten die Bänder

1. Rot
2. NIR
3. Blau
4. Grün
5. 3 mal MIR (siehe Tabelle)

mittels dem MRT-Tool sowie aus dem Qualitätsprodukt jener Layer mit der Qualitätsinformation der Albedo extrahiert und als GeoTIFF gespeichert. Die Rohdaten liegen nach dem Entpacken in der sinusoidalen Projektion vor.

3.2.2 Landsat

Die Daten des Landsatsystems wurden alle über die von der NASA und U.S.G.S. zur Verfügung gestellten Downloadsites *Earth Explorer* sowie *Glo-Vis* akquiriert. Eine Landsatszene definiert ihre Position über Path- und Row- Angaben welche für das ausgewählte Gebiet der Diplomarbeit im Schwarzwald liegen Pathindizes von 195 bis 196 sowie Rowindizes von 25 und 26 umfasst. Der maximale Bewölkungsgrad wurde mit 40% gewählt. Sämtliche Landsat5 und Landsat 7 Satellitendaten wurden entweder direkt mit der *Bulk Download Application* (BDA) der NASA downgeloadet oder über deren Bestellservice geordert und sobald fertig prozessiert ebenfalls mittels der BDA downgeloadet. Die Daten liegen als GeoTIFF für jedes einzelne Band pro Epoche auf. Sämtliche Landsatdaten wurden im Anschluss mittels LEDAPS absolut kalibriert. Die Kalibrierung liefert nicht nur die Landsatdaten in Oberflächenreflexion sondern vor allem die notwendige Wolkenmaske für jede zu kalibrierende LandsatSzene. Da LEDAPS seinerseits

alle seine Ergebnisse als hdf4 gespeichert müssen diese ebenfalls wieder in das GeoTIFF-Format übergeführt werden.

3.3 mathematische Grundlagen

Das Hauptziel der Diplomarbeit besteht darin, aus einer dichten Zeitreihe von atmosphärisch gut korrigierten MODIS-Satellitenbilddaten anhand von statistischen Methoden und Werkzeugen Landsat TM und Landsat ETM+ Satellitenbilddaten in Oberflächenreflexion überzuführen und radiometrisch zu verbessern. In diesem nachfolgenden Kapitel wird auf die mathematischen und statistischen Grundlagen eingegangen die für das Verständnis von Nöten sind. Die verwendeten Informationen stammen zumeist aus *Regression* (Härdle, 2007) sowie dem Vorlesungsskriptum aus *Parameterschätzung* von (Pail, 2010).

3.3.1 statistische Kennzahlen

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \dots \text{Mittelwert} \quad (3.1)$$

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \dots \text{Varianz} \quad (3.2)$$

$$\sigma = \sqrt{\sigma^2} \quad \dots \text{Standardabweichung} \quad (3.3)$$

3.3.2 Lineare Regression

3.3.2.1 OLS - das klassische lineare Modell

Generell gilt der Zusammenhang das y , die Variable die primär von Interesse ist, eine Funktion von x_1, \dots, x_k ist. Allgemein wird der Zusammenhang zwischen y und x_1, \dots, x_k als Funktion $f(x_1, \dots, x_k)$ beschrieben. Da diese Beziehung nicht exakt gemessen werden kann wird diese Beziehung immer durch zufällige Störungen verzerrt.

$$y = f(x_1, \dots, x_k) + \epsilon \quad (3.4)$$

Generelles Ziel ist die Schätzung der zugrundeliegenden unbekannten Funktion f . Über die unbekannte Funktion f und den Fehler ϵ werden folgende Annahmen getroffen:

1. Die systematische Komponente f ist eine Linearkombination der Kovariablen
Anhand der Kovariablen wird die Funktion als Linearkombination modelliert.

$$f(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (3.5)$$

Die Parameter $\beta_0, \beta_1, \dots, \beta_k$ sind allesamt unbekannt und werden geschätzt, wobei β_0 als Konstante verwendet wird (Intercept). Zusammengefasst in Vektoren mit der Dimension $p = (k + 1)$, $x = (1, x_1, \dots, x_k)'$ und $\beta = (\beta_0, \dots, \beta_k)$ ergibt sich die kompaktere Schreibweise

$$f(x) = x' \beta \quad (3.6)$$

2. Additivität der Störgröße Die Additivität der Störgröße wird als weitere Grundannahme des linearen Modells angenommen

$$f(x) = x' \beta + \epsilon \quad (3.7)$$

Für die zu schätzenden unbekannten Parameter β zu den erhobenen Daten y_i sowie $x = (1, x_1, \dots, x_k)'$, $i = 1, \dots, n$ existiert eine Beobachtungsgleichung für jede Beobachtung

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon_i = x_i' \beta + \epsilon_i \quad (3.8)$$

Anhand der Vektoren

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{sowie} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (3.9)$$

und der Designmatrix X

$$P_{ij} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \quad (3.10)$$

kann eine bequeme Matrixschreibweise angeführt werden:

$$y = \quad (3.11)$$

3.3.3 Harmonische Schwingungen - Fourier Analyse

3.3.3.1 Harmonische Schwingungen , Diskrete Fourier Transformation und Analyse

Da sich die Natur recht gut mit periodischen Funktionen beschreiben lässt, wurden auch hier einige Methoden aus der Signalanalyse und Verarbeitung gewählt um die Daten einer Zeitreihe der vorhandenen MODIS Satellitendaten zu analysieren, rekonstruieren und zu verbessern. Die Zerlegung nach Frequenzen entspricht dem, was ein Prisma mit dem einfallenden Licht macht. Der Lichtstrahl - zum Beispiel ein Sonnenstrahl - ist meist eine Überlagerung von Beiträgen verschiedenster Frequenzen. Da die Lichtbrechung beim Prisma frequenzabhängig ist, wird der Strahl zerlegt, der Ausfallwinkel hängt von der Frequenz des entsprechenden Anteils ab. Diese Zerlegung ist nichts anderes als die Projektion auf die Basisvektoren des Vektorraums periodischer Funktionen in L^2 . Die Methode heißt Fourierzerlegung oder **Fourieranalyse**, die Zusammensetzung der Funktion als Summe ihrer Komponenten ist die **Fouriersynthese** (Christian B. Lang, Norbert P. 2005).

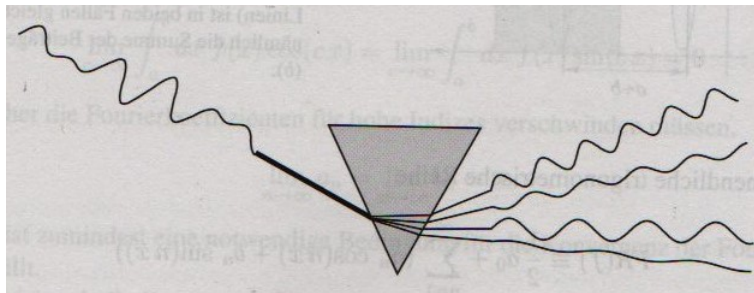


Abbildung 3.3: Aufteilung eines Lichtstrahls in seine Frequenzbereiche (Christian B. Lang, S401)

Die Daten der zu analysierenden Zeitreihe sind in der Regel keine reine sinusförmige Schwingung, besitzt jedoch ebenfalls eine Schwingungsdauer T und eine Kreisfrequenz $\omega_0 = 2\pi/T$. Diese zeitabhängige, periodische Funktion $y(t)$ lässt sich dann unter den bekannten Voraussetzungen in eine Fourier-Reihe vom Typ

$$\begin{aligned}
y(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cdot \cos(n\omega_0 t) + b_n \cdot \sin(n\omega_0 t)] = \\
&= \frac{a_0}{2} + a_1 \cdot \cos(\omega_0 t) + a_2 \cdot \cos(2\omega_0 t) + a_3 \cdot \cos(3\omega_0 t) + \dots \\
&\quad \dots + b_1 \cdot \sin(\omega_0 t) + b_2 \cdot \sin(2\omega_0 t) + b_3 \cdot \sin(3\omega_0 t) + \dots
\end{aligned} \tag{3.12}$$

entwickeln. Diese Entwicklung in unendlich viele Sinus- und Kosinusfunktionen bedeutet aus physikalischer Sicht eine Zerlegung der Schwingung $y(t)$ in ihre harmonischen Bestandteile, auch Schwingungskomponenten genannt. Sie bestehen aus der Grundschiwingung mit der Grundkreisfrequenz ω_0 und den harmonischen Oberschwingungen deren Kreisfrequenzen ganzzahlige Vielfache der Grundkreisfrequenzen sind: $2\omega_0, 3\omega_0, 4\omega_0, \dots$. Bringt man umgekehrt Grundschiwingungen und Oberschwingungen zur ungestörten Überlagerung, so erhält man als Resultierende genau die Schwingung $y = y(t)$ (Superpositionsprinzip der Physik ¹). Die Fourier-Koeffizienten $a_0, a_1, a_2, a_3, \dots, b_1, b_2, b_3, \dots$ bestimmen dabei die Amplituden der harmonischen Teilschwingung und somit letztendlich deren Anteile an der Gesamtschwingung (Lothar P., S182). Die als harmonische Analyse oder auch *Fourier-Analyse* bezeichnete Zerlegung einer nichtsinusförmigen Schwingung $y = y(t)$ in Grundschiwingung und harmonische Oberschwingungen läuft somit auf die Bestimmung der Fourier-Koeffizienten in der Entwicklung nach der vorangegangenen Formel (3.12) hinaus.

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cdot \cos(n\omega_0 t) + b_n \cdot \sin(n\omega_0 t)] \tag{3.13}$$

Sie können mit Hilfe der folgenden Integralformeln berechnet werden (Lothar P., S183)

$$\omega_0 : \quad \text{Kreisfrequenz der Grundschiwingung} \tag{3.14}$$

$$n\omega_0 : \quad \text{Kreisfrequenz harmonischer Oberschwingungen (n=1,2,3,4,5)} \tag{3.15}$$

$$T = \frac{2\pi}{\omega_0} : \quad \text{Schwingungsdauer (Periode)} \tag{3.16}$$

$$a_0 = \frac{2}{T} \cdot \int_T y(t) dt \tag{3.17}$$

¹ungestörte Überlagerung (Interferenz) mehrerer Wellen des gleichen Typs. Hier können sich die Amplituden der (z.B. elektromagnetischen) Wellen beim Überlagern gegenseitig verstärken oder abschwächen

$$a_1 = \frac{2}{T} \cdot \int_T y(t) \cdot \cos(n\omega_0 t) dt \dots n = 1, 2, 3, \dots \quad (3.18)$$

$$a_2 = \frac{2}{T} \cdot \int_T y(t) \cdot \sin(n\omega_0 t) dt \dots n = 1, 2, 3, \dots \quad (3.19)$$

Das Symbol T im Integral bedeutet, dass die Integration über ein beliebiges Periodenintervall von der Länge T zu erstrecken ist. Die Berechnung des Koeffizienten a_0 ist auch mit der Formel für $n = 0$ möglich, da $\cos(0\omega_0 t) = \cos 0 = 1$ ist (Lothar P., 2013, S183).

Die Fourier-Analyse einer nicht-sinusförmigen Schwingung $y = y(t)$ lässt sich auch in komplexer Form vornehmen:

$$y(t) = \sum_{n=-\infty}^{\infty} c_n \cdot e^{jn\omega_0 t} \quad (3.20)$$

Die Berechnung der komplexen Fourierkoeffizienten c_n erfolgt mit der Integralformel

$$c_n = \frac{1}{T} \cdot \int_0^T y(t) \cdot e^{-jn\omega_0 t} dt \quad (3.21)$$

Wobei ω_0 und T wieder wie folgt sich zusammen setzen

$$\omega_0 : \text{Kreisfrequenz der Schwingung} \quad (3.22)$$

$$T : \text{Schwingungs- oder Periodendauer} (T = 2\pi/\omega_0) \quad (3.23)$$

Die Umrechnung zwischen den reellen und der komplexen Darstellung erfolgt über folgende Gleichungen:

$$c_0 = \frac{1}{2}a_0, \quad c_n = \frac{1}{2}(a_n - jb_n), \quad c_{-n} = c_n^* = \frac{1}{2}(a_n + jb_n) \quad (3.24)$$

Die Umrechnung komplexer zur reellen Form passiert wie folgt:

$$a_0 = 2c_0 \quad a_n = c_n + c_{-n} = c_n + c_n^* \quad (3.25)$$

$$b_n = j(c_n - c_{-n}) = j(c_n - c_n^*) \quad (3.26)$$

3.3.3.2 Rekonstruktion mittels FFT

Die Fast Fourier Transformation (FFT) ist eine wichtige und bekannte Methodik aus der Signalverarbeitung, welche ein Signal in ihre einzelnen Spektralkomponenten zerlegt und Aufschluss über deren Zusammensetzung gibt,

Der Kern einer FFT ist eine Diskrete Fourier Transformation (DFT), jedoch ist der dahinterliegende Algorithmus optimiert, wovon auch der der Namen herrührt, sprich man verwendet die FFT um die Koeffizienten einer DFT zu ermitteln. Die FFT reduziert die Anzahl der Berechnungen für eine Anzahl N

$$\begin{aligned}
 \text{Diskrete Fourier Transformation} &= O(N^2) \\
 \text{Fast Fourier Transformation} &= O(N \cdot \log_2(N)) \\
 n &= 44000 \\
 DFT &= O(44000^2) = 1963000000 \\
 FFT &= O(4400 \cdot \log_2(44000)) = 204311
 \end{aligned} \tag{3.27}$$

Dies merkt man besonders wenn man große Mengen an Daten verarbeitet. Wie vorhin schon erwähnt ist die Fourier Transformation eine mathematische Operation, welche die Domäne der Zeit eines Signals in die Frequenzdomäne überführt, sprich in seine Frequenzanteile zerlegt und dadurch analysiert werden können. Bei der FFT wird ein zeitlich begrenztes Signal herangezogen und in seine Frequenzanteile zerlegt sowie deren Phasen und Amplituden bestimmt. In der Praxis wurde auf die bereits existierenden Routinen in **numpy 1.19.2** zurück gegriffen.

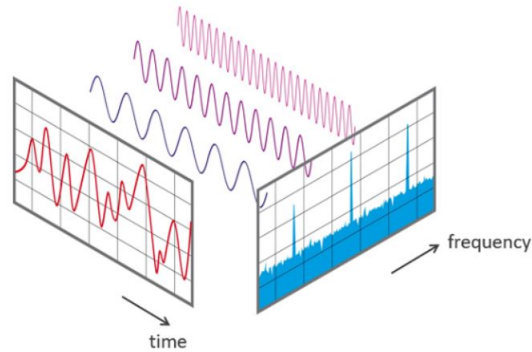


Abbildung 3.4: Aufteilung eines Signals in Zeit- und Frequenzbereich

$$\hat{x}_k = \sum_{i=1}^{n-1} x_i \quad \dots \text{FFT} \tag{3.28}$$

Bei genauerer Betrachtung des MODIS Datenmaterials stellt man fest, dass der Idealzustand bzw die optimalen Voraussetzungen bei der Berechnung der Reflexionswerte eher selten vorhanden sind. Dies äussert sich als schlechtere Pixelqualitätsinformation oder gar ein nichtvorhandener Wert der durch einen Füllwert repräsentiert wird. Die FFT ist jedoch ein sehr robuster Algorithmus der auch sehr gut bei verrauschten Signalen die wichtigsten Frequenzen heraus filtert. Voraussetzung für eine erfolgreiche Analyse ist jedoch ein lückenloses Signal. Wenn nun Datenlücken aufgetreten sind, wurden diese von den vorangegangenen und dem nachfolgenden Wert linear interpoliert um einen vollbesetzten Datenarray für die Ableitung der Parameter verwenden zu können.

3.3.4 Anwendung Savitzky-Golay Filterung

3.3.4.1 Polynom 2. Grades

Durch den Fakt, dass die MODIS Daten fehlerbehaftet sind, wird der Rohdatenwert durch einen gefilterten ersetzt. Um die verrauschte MODIS Zeitreihe für die später folgende relative radiometrische Kalibrierung zu filtern wurde die Methode nach Savitzky-Golay gewählt. Sie stammt aus der Signalverarbeitung und beschreibt einen Glättungsalgorithmus, der die Koeffizienten eines dem Datenmaterial zu Grunde liegend höheren Polynoms schätzt. Im Falle der Datenvorverarbeitung wurde ein Polynom 2. Grades gewählt,

$$f(t) = a_0 + a_1 \cdot t + a_2 \cdot t^2 \quad (3.29)$$

das sich auf Werte von 15 Epochen angleicht. Dieses Fenster bewegt sich nun über den ausgewählten Datenbereich und filtert die Daten anhand des ausgewählten Polynoms. Vorteile ergeben sich daraus, dass im Vergleich zu einem gleitenden Mittel die Spitzen nicht abgeschnitten werden sondern in die Charakteristik des Polynoms mit einfließen. Nachteil sind, dass nicht identifizierte Ausreißer ebenfalls so mit in die Berechnung einfließen und Einfluss nehmen. Der Wert in der Mitte des sich über die Zeit bewegendes Fensters ist im Interessenmittelpunkt und wird als später zu verwendender Wert in ein neues Rasterfile geschrieben.

Die Anzahl der Datenwerte auf der linken Seite werden als n_L , sowie die Datenwerte auf der rechten Seite mit n_R beschrieben, wobei gilt $n_L = n_R$. Die Parameter für das Polynom 2. Grades M von i , $a_0 + a_1 i + \dots + a_M i^M$ werden mittels der Methode Ausgleich nach Parametern für die Werte l_{-n_L}, \dots, l_{n_R} bestimmt. Die Designmatrix A setzte sich hierbei wie folgt

zusammen:

$$A_{ij} = i^j \quad i = -n_L, \dots, n_R, \quad j = 0, \dots, M \quad (3.30)$$

Der Normalgleichung ausgedrückt durch den Parametervektor $\hat{\mathbf{x}}$ und dem Beobachtungsvektor l_i in Matrizenschreibweise lautet

$$(\mathbf{A}^T \cdot \mathbf{A}) \cdot \hat{\mathbf{x}} = \mathbf{A}^T \cdot \mathbf{l} \quad or \quad \hat{\mathbf{x}} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{l}) \quad (3.31)$$

Anhand der Qualitätsinformation aus den MODIS-Epochen wird die Gewichtsmatrix \mathbf{P} aufgestellt:

$$P_{ij} = \begin{bmatrix} p_{11} & 0 & \dots \\ 0 & p_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (3.32)$$

Somit ergibt sich mit der Gewichtsmatrix folgende Form für die Normalgleichungen für den Parametervektor und den Beobachtungsvektor:

$$(\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A}) \cdot \hat{\mathbf{x}} = \mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{l} \quad or \quad \hat{\mathbf{x}} = (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{l}) \quad (3.33)$$

Die ausgeglichenen, gefitteten Beobachtungen erhält man anhand von

$$\hat{\mathbf{l}} = \mathbf{A} \cdot \hat{\mathbf{x}} \quad (3.34)$$

Da die Qualitätsinformation der MODIS-Epochen nicht immer ganz der Richtigkeit entsprechen (Restwolkenpixel, Dunst etc) wird eine zweite Iteration durchgeführt und die Daten werden anhand des Verbesserungsvektors neu gewichtet:

$$v = - \left[\mathbf{I} - \mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{P}) \right] \cdot \mathbf{l} \quad (3.35)$$

$$dv = \frac{1}{v} \quad (3.36)$$

$$P = \mathbf{I} \cdot dv \quad (3.37)$$

3.3.4.2 Gapfilling

Da es jahreszeitenabhängig zu länger anhaltender Wolkenbedeckung und somit zu Datenlücken für die Schätzung der Parameter des Polynoms kommen kann wird anstatt des Polynoms eine Gerade verwendet. Wenn die Anzahl der Elemente linksseitig und rechtsseitig eine gewisse Mindestanzahl unterschreiten, zum Beispiel $n_L \leq 3$ und $n_r \leq 3$ so wird anstatt eines Polynom 2. Grades eine Gerade durch die Daten gelegt und der zentrale Wert der neuen Berechnung als neuer Pixelwert verwendet. Die Identifikation der zu kleinen Anzahl erfolgt durch vordefinierte Pythonroutinen.

$$f(t) = a_0 + a_1 \cdot t \quad (3.38)$$

Die linear zu fittenden Daten werden mittels einer Pythonroutine ermittelt und zu einem Array zusammengefügt. n_C repräsentiert hierbei den eigentlich zu fittenden Wert um jenen herum normal das Polynom 2. Grades gelegt wird.

$$l_{ids} = l[n_L \leq 3; n_C = 0; n_R \leq 3] \quad (3.39)$$

Die Designmatrix bezieht sich hierbei nur auf die speziellen Situationen und variiert in ihrer Zeilenanzahl. Sie ist in ihrer Spaltenanzahl auf die Geradengleichung angepasst. Die Gewichtsmatrix verwendet jene Gewichte die bereits in der zweiten Iteration berechnet wurden. Der Beobachtungsvektor weist die gleiche Zeilenanzahl wie die Designmatrix auf

$$\hat{\mathbf{x}} = (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{P} \cdot l_{ids}) \quad (3.40)$$

Ist der interpolierte Wert nun jedoch kleiner als das Minimum oder größer als das Maximum im Beobachtungsvektor so wird der jeweilige Maximum oder Minimumwert gesetzt.

3.3.4.3 Verspeicherung der Zeitreihe

Der Filtervorgang findet Bandweise statt. Die gefilterten Daten werden zuerst als GeoTIFF verspeichert und in einem Folgeschritt zu europaweiten Mosaiken zusammengefügt sowie epochenweise in Stacks verspeichert. Als Ergebnis liegt nun eine geglättete, wolkenfreie Zeitreihe der MODIS-Sensoren für ganz Europa (ausgenommen Island) vor, welche zur radiometrischen Kalibrierung für Landsat5 und Landsat 7 Daten verwendet werden kann.

3.3.5 Ausreisserdetektion

Aussreissertests werden dazu benutzt um die Zuverlässigkeit von Daten zu überprüfen und vor allem extreme Erscheinungen in den Daten zu identifizieren und gegebenenfalls zu eliminieren. Ausreisser sind extrem hohe oder niedrige Werte in einer ansonst mäßig unterschiedlichen Messreihe. Im folgenden wird die Methode nach Grubbs beschrieben, welche in dieser Diplomarbeit Verwendung gefunden hat. Ziel der implementierten Ausreisserdetektion ist es vor der Aggregation der LandsatDaten auf die MODIS-Pixel-Auflösung nicht ausmaskierte Wolkenpixel aufgrund der nicht perfekten Wolkenmaske des LEDAPS-Programmes zu identifizieren und zu eliminieren. Dies impliziert das nach extrem hohen Werten in den Daten gesucht wird.

Das Verfahren selbst prüft eine berechnete Testgröße G anhand des Mittelwert \hat{x} und der Standardabweichung σ des Samples und prüft ob die Hypothese oder Gegenhypothese wahr ist, das es sich bei der Beobachtung um einen Ausreisser handelt.

$$G = \frac{\max |x_i - \hat{x}|}{\sigma} \quad (3.41)$$

$$\alpha \quad \dots \text{Signifikanzniveau} \quad (3.42)$$

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{(t_{\alpha/(2N), N-2})^2}{N-2 + (t_{\alpha/(2N), N-2})^2}} \quad (3.43)$$

Viele statistische Methoden werden durch das Vorhandensein von Ausreißern beeinflusst. Bei schon simplen Berechnungen wie dem Mittelwert oder

der Standardabweichung kann es zu einer Störung durch einen einzigen „verfälschten„ Wert kommen.

3.4 Verwendete Sprachen und Programme aus dem Bereich Data Science sowie Geoinformatik

3.4.1 Softwareentwicklung mit Python

Python ist wenn nicht die am meisten verwendete Sprache wenn es um Datenmanipulation geht. Sei es serverseitig im Backendbereich von Webservices oder Machine Learning Anwendungen, Python bietet so ziemlich zu jeder Problemstellung eine passende Herangehensweise oder Lösung. Die Community ist riesig und man findet sehr schnell Hilfe zu Problemstellungen in diversen Foren wie *www.stackoverflow.com*. Im Zuge dieser Diplomarbeit wurden zwei Versionen von Python verwendet (Python 3.8 und Python 2.7) um Datenmanipulationsschritte zu automatisieren. Der gesamte Prozess der Datenakquisition, Datenaufbereitung als auch jener der relativen Kalibrierung und Ergebnispräsentation ist entweder mit bereits vorhandenen Bibliotheken oder selbstgeschriebenen Modulen in Python umgesetzt worden und umfasst einige tausende Zeilen Code. Die von mir verwendete IDE zur Erstellung der Skripts und Software stammt von JetBrains und ist unter dem Namen PyCharm sehr bekannt. Hier sei darauf hingewiesen, dass das Thema dieser Diplomarbeit sich nicht im Detail mit Problemstellungen aus der Softwareentwicklung auseinandersetzt, sondern die Softwareentwicklung mit Python als Mittel zum Zweck verwendet. Nichts destotrotz sind ein paar verwendete Feinheiten beschrieben, die einen Geschwindigkeitsvorteil in der Berechnung oder bei der Manipulation der Daten gebracht haben.

3.4.1.1 Python 3.8

Python 3.8 war das Mittel zum Zweck für so ziemlich alle Arbeitsschritte um bis einschliesslich zur gefilterten MODIS-Zeitreihe zu gelangen. Die MODIS Daten wurden mittels einem Webscrabber-Script unter Verwendung von *bs4* und dem Pythoninternen *requests* Modul von den NASA Servern gedownloaded, mittels *GDAL* vom HDF5-Format in einzelne, bandweise GEOTIFFS konvertiert und anschliessend mittels *numpy* und den vorhin beschriebenen Algorithmen gefiltert oder rekonstruiert. Diese gefilterten Daten wurden weiters zu epochenweisen Stacks verarbeitet, da die relative Kalibration diese voraussetzt.

3.4.1.2 Python 3.8 - Multiprocessing

Da nur die gedownloadeten HDFS und rohen GEOTIFF Daten alleine fast einen ganzen Terabyte an Speicherplatz benötigen, war klar das es nicht sinnvoll sein kann, einem einzigen Pythonprozess die ganze Arbeit machen zu lassen. Multithreading ist am Papier mit Python möglich, aber auf Grund des Global Interpreter Locks immer an eine CPU gebunden. Paralleles Ausführen von Arbeitsschritten auf mehreren CPUs ist aber mittels dem in Python integrierten Multiprocessing Modul und seit der Python Version 3.8 auch mit einem Shared Memory oder gemeinsamen Arbeitsspeicherbereich möglich. Dies erlaubt einen Speicherbereich zwischen mehreren Python Prozessen zu teilen bzw zugreifbar zu machen, um große Datenmengen parallel zu verarbeiten. Jedoch sei hier darauf hingewiesen, dass man bei der Verwendung von Shared-Memory-Objekten sich nicht mehr auf die gewohnte Typesicherheit von Python verlassen darf und zusätzlich bedenken muss:

- welchen Typ hat das Shared-Memory-Objekt: -z.B. int16
- welcher Typ kommt bei der Berechnung heraus - z.B. float64
- in welchen Typ muss ich das Ergebnis wieder umwandeln um es am Shared-Memory-Object zu verspeichern

Wenn man dies nicht berücksichtigt kann dies es zu unerwartetem Verhalten führen und in elendslange Fehlersuche münden. Behirnt man dies jedoch korrekt, ist echtes paralleles Programmieren mit Python für CPU intensive Rechenoperationen vollumfänglich möglich.

3.4.1.3 Python 2.7

Die Rechenschritte der relativen Kalibration wurden alle in Python 2.7 getätigt. Grund waren die anfänglich verwendeten Module der arcpy API von ESRI, welche jedoch oft erhebliche Performanceprobleme hatten und im Endeffekt alle durch selbstgeschriebene Module oder durch Verwendung von bekannten Open Source Bibliotheken wie GDAL, numpy oder scipy ersetzt wurden.

3.4.1.4 Anaconda

Um zwischen den Interpreterversionen ohne Probleme hin und her wechseln zu können, wurde der Package und Virtual-Environment Manager Anaconda verwendet. PyCharm unterstützt auch unterschiedliche Interpreter die man sich als extra Runtime-Konfiguration in seinem Projekt erstellen kann.

3.4.2 ArcGIS, QGIS

Die abgeleiteten Daten benötigen natürlich eine menschliche, visuelle Kontrolle. Dies geschah in erster Linie mit ESRI's ArcMap 10.2 oder QGIS 3.16. Wie schon vorhin erwähnt, wurde anfangs auch Skripts oder Toolboxes von ArcMap verwendet, jedoch durch eigene Softwaremodule ersetzt. Besonders nützlich war in beiden Programmen der Information Cursor, welcher erlaubt einzelne Pixel-Zeitreihen auf schnelle Art und Weise tabellarisch zu visualisieren und zu vergleichen.

3.4.3 Betriebssysteme, Speicherplatz und Rechenlogistik

Hier möchte ich nur kurz Anreißern, dass die großen Datenmengen eine eigene Denkweise, Infrastruktur und Logistik erfordert haben. Die verwendeten Betriebssysteme unterscheiden sich erheblich in den Ausführzeiten der Python Programme welche für das Rekonstruieren der MODIS Zeitreihen verwendet wurden. Deshalb wurde beim Fitten und Filtern der Daten auf Linux (Debian 10, Ubuntu 18.04) vertraut. Es kamen bis zu vier Rechner mit mindestens 4 CPUs ab 3GHz Takt und mindestens 32GB RAM zum Einsatz um die Daten parallel zu verarbeiten. Pro Spektralband der MODIS Zeitreihe kann man für 20 Jahre und jeden 9ten Tag eine vorhandene Szene (oder 927 Datenpunkte) ca 17h Rechenzeit für Savitzky-Golay-Filter einplanen, die FFT brauchte hingegen nur an die drei Stunden. Der verbrauchte Speicherplatz für die MODIS Daten in der Version 6 sowie der Landsat Daten in roher als auch absolut Kalibrierter Form belaufen sich auf 2,42 TB (Terabyte). Um die Daten zu sichern wurde Speicherplatz im Gesamt Volumen von 11 TB angeschafft um sie auf drei unabhängigen Speichervolumen zu sichern.