

Sicherung der Qualität energiemeteorologischer Datenprodukte durch Metadaten im Projekt WISENT (AP 3.1)

Gerhard Gesell, Marion Schroedter-Homscheidt
Deutsches Zentrum für Luft- und Raumfahrt DLR, Deutsches Fernerkundungsdatenzentrum DFD
Münchner Straße 20
82234 Weßling

Telefon: +49 8153 28 1322
Telefax: +49 8153 28 1363
E-Mail: gerhard.gesell@dlr.de
Internet: www.caf.dlr.de

Inhaltverzeichnis

1.	Die allgemeine Bedeutung von Metadaten	3
2.	Qualitätssicherung (QS) mit Hilfe von Metadaten	4
2.1.	Ausgangssituation und Zielsetzung	4
2.2.	QS-Nutzeranforderungen	9
2.3.	QS-Metadaten für energiemeteorologische Produkte.....	13
2.3.1.	QS-Informationen der HELIOSAT-Prozesskette (DFD-Teil)	15
2.3.2.	QS-Metadaten-Definition für die HELIOSAT-Prozesskette (DFD-Teil)	21
3.	Transformation der QS-Metadaten in Nutzerinformation	26
4.	Risiken und Grenzen der QS mit Metadaten.....	28
5.	Zusammenfassung	29
Abbildung 2-1: Typischer Aufbau eines Datenproduktes im DIMS-Archiv des DFD		5
Abbildung 2-2: HELIOSAT-3 Prozesskette bei DFD und EdM		14
Abbildung 2-3: SCENES/APOLLO-Prozessierungskette für Wolkenparameter beim DFD		15
Abbildung 2-4: Beispiel für einen Auszug aus ^QP zur Vollständigkeit und Bestimmtheit		26
Tabelle 1-1: Aspektematrix Metadaten		4
Tabelle 2-1: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Gesamtprodukt)		7
Tabelle 2-2: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Primärkomponente)		7
Tabelle 2-3: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Quicklook-Komponente)		8
Tabelle 2-4: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Geografiekomponente)		9
Tabelle 2-5: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Prozessierungskomponente)		9
Tabelle 2-6: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Vollständigkeit“		11
Tabelle 2-7: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Konsistenz“		11
Tabelle 2-8: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Genauigkeit“		11
Tabelle 2-9: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Gültigkeit“		12
Tabelle 2-10: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Plausibilität“		12
Tabelle 2-11: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Überprüftheit“		12
Tabelle 2-12: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Bestimmtheit“		12
Tabelle 2-13: Verfügbare QS-Informationen der HELIOSAT-Prozesskette (DFD-Teil)		16
Tabelle 2-14: QS-Metadaten-Definition für energiemeteorologische Produkte (DFD-Teil)		22
Tabelle 3-1: Transformationstabelle QS-Metadaten zu Nutzerinformation		27

1. Die allgemeine Bedeutung von Metadaten

Datenprodukte können im Allgemeinen nur dann sinnvoll gespeichert und genutzt werden, wenn eine hinreichende Beschreibung ihres Inhaltes, ihrer Gültigkeit, ihrer räumlichen und zeitlichen Orientierung sowie gegebenenfalls ihrer Entstehung verfügbar ist. Eine solche Beschreibung wird mit Hilfe von Metadaten gegeben. Je nach dem, was mit den Datenprodukten geschehen soll bzw. wozu sie dienen sollen, werden die einen oder anderen Aspekte der Metadaten in den Vordergrund gestellt. Aus der Sicht des Archivators müssen die Metadaten es besonders leicht und eindeutig ermöglichen, ein Datenprodukt in eine Datenbank einzubinden und dort auch wieder zu finden. Aus der Sicht eines Nutzers ist es zunächst besonders wichtig, ein Produkt eindeutig über Metadaten identifizieren zu können oder beispielsweise auch gruppieren zu können. Bis hierhin geht es um so genannte äußere Metadaten, die separat bzw. in Zusatzdateien gespeichert über äußere Eigenschaften der Datenprodukte informieren. Darüber hinaus muss der Nutzer aber auch viele innere Eigenschaften der Datenprodukte kennen, um sie sinnvoll anwenden zu können. Solche inneren Eigenschaften sind zum Beispiel Angaben

- zu Struktur und Umfang von Datenobjekten und Variablen soweit sie nicht a priori festgelegt sind,
- zu physikalischen Größen und Einheiten,
- zu Prozessversionen sowie
- zur Vollständigkeit, Verlässlichkeit und Genauigkeit der Daten, also auch und vor allem zu deren Qualität.

Im Folgenden wird daher zwischen äußeren und inneren Metadaten unterschieden, wobei die inneren Metadaten sich stets auf die Eigenschaften von Einzelelementen oder Ebenen (Layer) eines Datenproduktes beziehen, die äußeren Metadaten dagegen auf das Datenprodukt als Ganzes. Weiter ist zu beachten, dass die äußeren Metadaten so weit möglich nach einer ISO-Norm standardisiert sein sollten, um eine Produkt-Interoperabilität zu gewährleisten. Die inneren Metadaten sollen so weit wie möglich ebenfalls einer ISO-Norm entsprechen, jedoch ist es bei der Beschreibung von Einzelobjekten auf sehr tiefer Ebene unter Umständen nicht möglich, eine passende Norm zu finden. Zur formatierten Übermittlung von standardisierten Metadaten kommen Metasprachen wie XML in Frage. Ein im Jahr 2003 definierter Standard für den Inhalt von Metadaten in Geoinformationssystemen wird durch die ISO-Norm 19115 beschrieben. Diese Norm beinhaltet Angaben zur eindeutigen Identifikation, zur Ausdehnung, zur Qualität, zum räumlichen und zeitlichen Schema, zum Referenzsystem und zur Nutzbarmachung. Dazu gibt ISO-19139 die Struktur vor, jedoch ist ISO 19139 noch kein fertiger Standard. Allgemein finden sich unter ISO-191xx Normen zur Geoinformation. Standards für interoperable Metadaten haben die Aufgabe, Metadaten aus unterschiedlichen Quellen nutzbar zu machen. Diese Standards umfassen drei Aspekte. Die Semantik beschreibt die Bedeutung, die meist durch eine Normung festgelegt wird bzw. festgelegt werden sollte. Das Datenmodell legt fest, welche Struktur die Metadaten besitzen können, wobei die Daten die Beschreibung und das Modell die Struktur betreffen. Die Syntax dient schließlich dazu, die durch die Metadaten gemachten Aussagen darzustellen.

In diesem Dokument wird ausschließlich auf die Semantik der für die Qualitätssicherung notwendigen Metadaten eingegangen, d.h. es handelt sich um eine Analyse, welche Informationen zur Qualitätssicherung der Datenprodukte übermittelt werden müssen. Gleichzeitig werden diese Informationen inhaltlich strukturiert und es werden Namen und Bedeutungen der Metadatenobjekte festgelegt bzw. vorgeschlagen. Hierbei wird von einer Aspektematrix (Tabelle 1-1) für die Bedeutung von Metadaten ausgegangen, die die äußeren bzw. inneren Eigenschaften der Datenprodukte mit den Präferenzen der Produktlieferanten bzw. –nutzer verknüpft. Die in dieser Matrix aufgeführten Kategorien von Metadaten sollen für die Datenprodukte umfassend sein, d.h. kein notwendiger Aspekt zur Erzeugung, Archivierung und Nutzung der Produkte darf unbeschrieben bleiben. Allerdings ist dabei stets darauf zu achten, dass Metadaten wirklich nur Aussagen über die Eigenschaften und die Entstehung der Datenprodukte machen. Daten, die zur Erzeugung von Produkten über die Eingangsdaten hinaus notwendig sind, etwa Kalibrationskoeffizienten oder Kartenparameter etc. werden „Zusatzdaten“ genannt und zählen selbst nicht zu den Metadaten. Die Eigenschaften von Zusatzdaten haben in der Regel aber erheblichen Einfluss vor allem auf die Produktqualität, weshalb etwa Metadaten zur Qualitätssicherung stets die Rolle der Zusatzdaten und damit auch deren Qualität berücksichtigen sollten.

Metadaten	äußere	innere
------------------	---------------	---------------

Sichtweise		
Provider	<ul style="list-style-type: none"> - Räumlich und zeitliche Orientierung - Orientierung in der Datenbank - Prozessor-ID(s) - Herkunft - etc. 	<ul style="list-style-type: none"> - Anzahl der Produktebenen - Datentyp und Feldgrößen - etc.
Nutzer	<ul style="list-style-type: none"> - Räumliche und zeitliche Orientierung - Versionskontrolle - Produktbeschreibung, soweit sie für das Entpacken notwendig ist - QS: Vollständigkeit des Gesamtproduktes - QS: Gesamtbewertung evtl. aufgetretener Fehler bei der Prozessierung - etc. 	<ul style="list-style-type: none"> - Anzahl der Produktebenen - Datentyp und Feldgrößen - vorkommende physikalische Größen und Dimensionen - Beschreibung der Produktebenen - QS: Vollständigkeit, Verlässlichkeit und Genauigkeit sowohl ganzer Produktebenen als auch von Einzelwerten - etc.

Tabelle 1-1: Aspektematrix Metadaten

Neben allen *notwendigen* Aspekten zur Erzeugung, Archivierung und Nutzung der Produkte kann es auch Metadateninhalte geben, die zwar hilfreich sind, nicht aber notwendig, um eine ausreichend umfassende Produktbeschreibung zu gewährleisten. Der Begriff „notwendig“ ist hier im Sinne von „zur Beschreibung zwingend erforderlich“ (englisch mandatory) zu verstehen, im Gegensatz etwa zu „optional“. Daher werden im Folgenden alle Metadatenobjekte als „notwendig“ oder „optional“ gekennzeichnet.

2. Qualitätssicherung (QS) mit Hilfe von Metadaten

Ausgehend von bestehenden Produkt- und Prozessierungsstrukturen und den sie beschreibenden Metadatenansätzen am DFD werden in diesem Abschnitt notwendige Qualitätssicherungs-Informationen identifiziert und die sie darstellenden Metadaten definiert.

2.1. Ausgangssituation und Zielsetzung

Als Ausgangssituation finden wir am DFD Produkt- und Metadatendefinitionen vor, die sich an den Anforderungen des Data- and Information Management System (DIMS) des DFD-Archivs für Fernerkundungs-Daten und -Produkte, dessen Nutzern und Betreibern orientiert. Die Abbildung 2-1 zeigt die typische Struktur der im DIMS gespeicherten Produkte. Ein solches Produkt besteht aus vier oder fünf Komponenten, wobei die Primärkomponente die eigentlichen (Mess-)Daten enthält, die übrigen Komponenten dagegen spezifische Daten zur Produktvisualisierung, -orientierung, -prozessierung und -qualität. Für das Gesamtprodukt sowie jede einzelne Komponente sollen sowohl äußere als auch innere QS-Metadaten angegeben werden. Die QS-Metadaten können wie alle Metadaten als Attribute abgespeichert werden, jedoch muss für energiemeteorologische Produkte auch die Möglichkeit zur Speicherung von inneren Metadaten zur QS der Einzelwerte als Layer in der Primärkomponente vorgesehen werden.

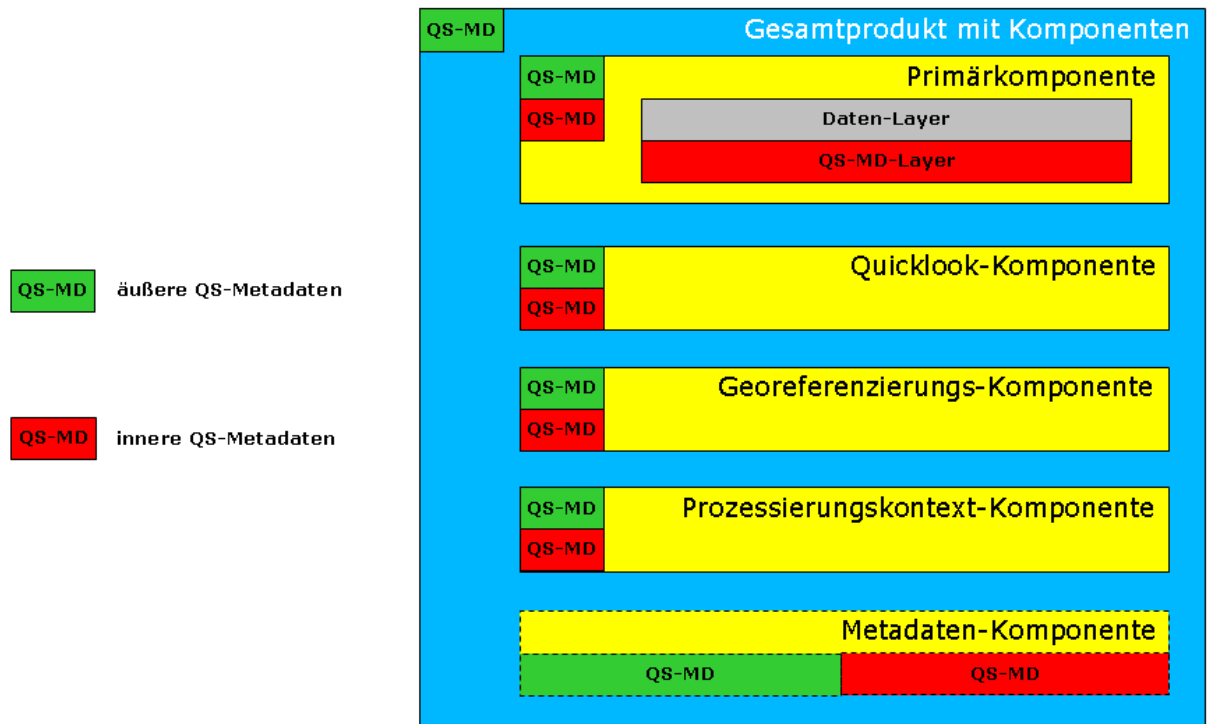


Abbildung 2-1: Typischer Aufbau eines Datenproduktes im DIMS-Archiv des DFD

Für die DIMS-Archivierung von Produkten des World Data Centers for Remote Sensing of the Atmosphere (WDC-RSAT) wurde in einem eigenen so genannten Product Definition Document (PPD) eine DIMS-konforme Produktstruktur und Metadatenammlung beschrieben. Allerdings existiert in dieser Beschreibung bisher nur eine Mindestanforderung an Metadaten zur Qualitätssicherung. Der Ausgangszustand dieser Metadatenbeschreibung für WDC-RSAT-Produkte ist in Tabelle 2-1 bis Tabelle 2-5 gegliedert nach den Produktkomponenten wiedergegeben. Gemäß den in Abschnitt 2.2 erhobenen Nutzeranforderungen werden die WDC-RSAT Metadatenpezifikationen als Ausgangspunkt genutzt, um QS-Metadatendefinitionen erweitert und mit einer eigenen Metadaten-Produktkomponente ergänzt. In Abschnitt 2.3 schließlich werden konkret für die HELIOSAT-3-Prozessierungskette diese QS-Metadaten hinsichtlich ihrer Bedeutung diskutiert und es werden Methoden und Algorithmen zu ihrer Gewinnung vorgeschlagen.

WDC-RSAT Gesamtprodukt

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
availability		C	+		QA
reliability		C	+		QA
quality		C	+		QA
revision		C	+	<revision>	
sphere		C	+	Earth	
spatialCoverage			+		
spatialCoverage		C		isGlobal	
or					
boundingBox					
westBounding Coordinate				-180.0:180.0	
eastBounding Coordinate				-180.0:180.0	
southBounding				-180.0:180.0	

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
Coordinate					
northBounding Coordinate				-180.0:180.0	
temporalCoverage			+		
startTime		Date	+	YYYY-MM-DD THH:mm:ss	
stopTime		Date	+	YYYY-MM-DD THH:mm:ss	
creation			+		
time		Date	+	YYYY-MM-DD THH:mm:ss	date/time of creation
site		C	+	DLR-DFD	
creator		C	+	<processor>	\$LID mission proc system name version (Proc system name from parameter file) e.g. EURAD AQ model
predecessor		list			\$RIDs of input products
auxiliary			+		
doi		C		<DOI>	Digital Object Identifier
mission		C	+		e.g. ENVISAT1
sensor		C	+		e.g. SCIAMACHY, TOMS
code		C	+		e.g. L3
datasets		list			List elements denote datasets in PRIM component. If PRIM contains a single dataset only, the list has length 1.
species		C		<species>	e.g. O3, NO2
units		C		<units>	e.g. Dobson
altitudeCoverage					
top		Float			
bottom		Float			
verticalResolution		C	+	valids <resolution>	e.g. VCD, TCD, SCD
temporalResolution		C		<nhours>	e.g. daily, monthly (in hours), applied only to lvl4 products
compositeType		C		valids<compositeType s>	<i>daily, weekly, monthly, yearly</i>
method		C	+	<method>	e.g. Kalman Filtering, Composite, ROSE
methodVersion		C	+	<methodVersion>	Method version
processorName		C	+	<processor>	Processor software name
processorVersion		C	+	<version>	Processor software version
algorithm					algorithm used to produce each of the output datasets
algorithmName		C	+	<algorithm>	Algorithm name
algorithmVersion		C	+	<algorithmVersion>	Algorithm version
externalProcessorInput		list			list of tuples identifying processor input

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
					originating from external sources.
name		C			product name; should also identify the source
date		Date		YYYY-MM-DD THH:mm:ss	
version		C			

Tabelle 2-1: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Gesamtprodukt)

WDC-RSAT Produkt - Primärkomponente

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
availability		C	+		QA
reliability		C	+		QA
quality		C	+		QA
revision		C	+	<revision>	
sphere		C	+	Earth	
spatialCoverage			+		
spatialCoverage		C		isGlobal	
or					
boundingBox					
westBounding Coordinate				-180.0:180.0	
eastBounding Coordinate				-180.0:180.0	
southBounding Coordinate				-180.0:180.0	
northBounding Coordinate				-180.0:180.0	
temporalCoverage			+		
startTime		Date	+	YYY-MM-DD THH:mm:ss	
stopTime		Date	+	YYY-MM-DD THH:mm:ss	
creation			+		
time		Date	+	YYY-MM-DD THH:mm:ss	date/time of creation
site		C	+	DLR-DFD	
creator		C	+	<processor>	\$LID mission proc system name version (Proc system name from parameter file) e.g. EURAD AQ model

Tabelle 2-2: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Primärkomponente)

WDC-RSAT Produkt - Quicklook-Komponente

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
availability		C	+		QA

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
reliability		C	+		QA
quality		C	+		QA
revision		C	+	<revision>	
sphere		C	+	Earth	
spatialCoverage			+		
spatialCoverage		C		isGlobal	
or					
boundingBox					
westBounding Coordinate				-180.0:180.0	
eastBounding Coordinate				-180.0:180.0	
southBounding Coordinate				-180.0:180.0	
northBounding Coordinate				-180.0:180.0	
temporalCoverage			+		
startTime		Date	+	YYY-MM-DD THH:mm:ss	
stopTime		Date	+	YYY-MM-DD THH:mm:ss	
creation			+		
time		Date	+	YYY-MM-DD THH:mm:ss	date/time of creation
site		C	+	DLR-DFD	
creator		C	+	<processor>	\$LID mission proc system name version (Proc system name from parameter file) e.g. EURAD AQ model
region		C	+	<region>	e.g. global, europe, south pole

Tabelle 2-3: Metadaten definition für WDC-RSAT-Produkte in DIMS (Quicklook-Komponente)

WDC-RSAT Produkt - Geografiekomponente

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
availability		C	+		QA
reliability		C	+		QA
quality		C	+		QA
revision		C	+	<revision>	
sphere		C	+	Earth	
spatialCoverage			+		
spatialCoverage		C		isGlobal	
or					
boundingBox					
westBounding Coordinate				-180.0:180.0	
+- eastBounding Coordinate				-180.0:180.0	

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
southBounding Coordinate				-180.0:180.0	
northBounding Coordinate				-180.0:180.0	
temporalCoverage			+		
startTime		Date	+	YYY-MM-DD THH:mm:ss	
stopTime		Date	+	YYY-MM-DD THH:mm:ss	
creation			+		
time		Date	+	YYY-MM-DD THH:mm:ss	date/time of creation
site		C	+	DLR-DFD	
creator		C	+	<processor>	\$LID mission proc system name version (Proc system name from parameter file) e.g. EURAD AQ model
epsgCode		C	+	<EPSG Code>	EPSG code specifying the coordinate reference system used in the GeoTIFF file

Tabelle 2-4: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Geografiekomponente)

WDC-RSAT Produkt – Prozessierungskontext

Parameter	Key Param.	Data Type	NOT NULL	Value Range	Description
availability		C	+		QA
reliability		C	+		QA
quality		C	+		QA
revision		C	+	<revision>	
creation			+		
time		Date	+	YYY-MM-DD THH:mm:ss	date/time of creation
site		C	+	DLR-DFD	

Tabelle 2-5: Metadatendefinition für WDC-RSAT-Produkte in DIMS (Prozessierungskomponente)

2.2. QS-Nutzeranforderungen

Die Erhebung der Anforderungen an QS-Metadaten gestaltete sich unerwartet schwierig aus folgenden Gründen:

- Provider und Nutzer haben zunächst ein völlig unterschiedliches Verständnis von Qualitätssicherung an sich sowie von Form und Inhalt der QS-Aussagen im Detail.
- Provider erzeugen QS-Metadaten in einem „Bottom-up-Prozess“. Den Nutzer dagegen interessieren ausschließlich integrale Aussagen und nur gegebenenfalls wird ins Detail gegangen, was einer Top-down-Sicht entspricht.
- QS-Metadaten können oft nur bei einer gewissen Grundkenntnis der physikalischen Grundlagen, der angewandten physikalischen, mathematischen und informatischen Methoden sowie der beteiligten Prozessoren und deren Implementation richtig interpretiert

werden. Hier hat der Provider gegenüber dem Nutzer einen größeren Wissensvorsprung, der durch zusätzliche Beschreibung der Nutzung der QS-Metadaten durch den Provider gegenüber dem Nutzer ausgeglichen werden muss.

Auf Grund des Wissensvorsprungs des Produkt-Providers muss dieser der Initiator und Koordinator bei der Erhebung der Anforderungen sein. Es gilt, die Sichtweisen von Nutzer und Provider gegenüber zu stellen und sinnvolle Zusammenfassungen und inhaltliche Transformationen zu beschreiben. Dabei werden, ausgehend von Provider-Sicht, zuerst die wichtigsten Klassen von QS-Metadaten bestimmt. Dies geschieht auf Grund einer groben Gliederung dessen, was der Nutzer wünscht und was der Provider bei Kenntnis der Methoden und Prozessoren anbieten kann. Dabei ergeben sich sieben QS-Metadaten-Klassen, die jeweils unterschiedliche Aspekte zum Inhalt haben, nämlich

- Vollständigkeit (auch Verfügbarkeit),
- Konsistenz,
- Genauigkeit,
- Gültigkeit,
- Plausibilität,
- Überprüftheit und
- Bestimmtheit.

Die ersten drei Klassen finden sich auch in der ISO-Norm 19115 wieder, d.h. die Metadaten in diesen Klassen lassen sich gegebenenfalls an diese Norm anlehnen bzw. können der ISO-Beschreibung entnommen werden. Die Klasse „Gültigkeit“ (Validity) tritt in der ISO-19115 als Unterklasse auf, ebenso wie eine Klasse „Correctness“, die auf die energiemeteorologischen Produkte wegen der großteils fehlenden physikalisch-analytischen Überprüfungsmöglichkeiten nicht ohne weiteres abgebildet werden kann. Stattdessen kann die Korrektheit eines Ergebnisses hier mit Hilfe der Klassen Plausibilität und Bestimmtheit abgeschätzt werden.

In Tabelle 2-6 bis Tabelle 2-12 werden bei Angabe der QS-Metadaten-Inhalte Provider- und Nutzersicht gegenüber gestellt. Dabei wird von Nutzerwünschen ausgegangen, die sich bei der Erhebung als ziemlich allgemein herausgestellt haben und sich auch auf Nachfrage bei der Nutzerseite nicht weiter detaillieren ließen. Danach werden von den Nutzern im Projekt WISENT (im Wesentlichen meteocontrol) folgende QS-Metadaten-Angaben gewünscht:

- Beschreibung der systematischen Fehler (BIAS)
- Beschreibung statistischer Fehler (MEAN und RMSE = root mean square error, Wurzel aus dem mittleren Fehlerquadrat)
- Angabe von Konfidenzintervallen und Gütekennzahlen (die erst zu definieren sind)
- Angaben zur Gültigkeit sowohl physikalisch, interpretativ als auch evtl. juristisch
- Beschreibung der Datenqualität, bezogen sowohl auf einzelne Werte als auch auf ganze Datensätze

Als beschreibende Klassen für die Datenqualität werden vom Nutzer überwiegend die Vollständigkeit, Genauigkeit und Gültigkeit gesehen.

Folgende Abkürzungen werden in Tabelle 2-6 bis Tabelle 2-12 benutzt:

- <QT = äußere QS-Metadaten für das Gesamtprodukt (Qualität total)
- <QP = äußere QS-Metadaten der Primärkomponente
- >QP = innere QS-Metadaten der Primärkomponente
- ^QP = QS-Metadaten-Layer der Primärkomponente (innere Metadaten)
- <QQ = äußere QS-Metadaten der Quicklook-Komponente
- >QQ = innere QS-Metadaten der Quicklook-Komponente
- <QG = äußere QS-Metadaten der Georeferenzierungskomponente
- >QG = innere QS-Metadaten der Georeferenzierungskomponente
- <QK = äußere QS-Metadaten der Prozessierungs-Kontext-Komponente
- >QK = innere QS-Metadaten der Prozessierungs-Kontext-Komponente
- <QM = äußerer Metadatenatz der Metadatenkomponente
- >QM = innerer Metadatenatz der Metadatenkomponente
- <QA = betrifft äußere QS-Metadaten aller Produktkomponenten
- >QA = betrifft innere QS-Metadaten aller Produktkomponenten

Vollständigkeit (auch Verfügbarkeit)	
Provider	Nutzer
<QT - Fehlende Produktkomponente oder Komponenten von Zeitreihen (Cubes)	Kombinierte Information zur Nutzbarkeit des Datenproduktes bei Ausfällen → <QT

Vollständigkeit (auch Verfügbarkeit)	
Provider	Nutzer
<QA - Systematisch fehlende Teile von Komponenten (a priori bekannte zeitliche und räumliche Ausfälle), z.B. Zeilen, Kanäle etc.	Die Kenntnis, ob fehlende Daten nachprozessierbar sind, ist hierbei wichtig.
>QA ^QP - Sporadisch fehlende Teile von Komponenten (statistische Ausfälle, Rauschen), z.B. Zeilen, Kanäle, Pixel etc., Einzelwerte und Einzelwertgruppen	
<QK - Vollständigkeit der Methodik, alle Algorithmen angewandt, z.B. Schnee/Wolken-Unterscheidung	Die Nutzer entscheiden selbst aktuell über die Produkt-Nutzbarkeit bei unvollständiger Methodik oder geben dem Provider einen Katalog mit Mindestanforderungen hierzu vor
<QM >QM - Fehlen von Metadaten (QS- und andere Metadaten)	Die Nutzer legen zusammen mit den Providern für jedes Produkt fest, welche QS-Metadaten essentiell und welche optional sind

Tabelle 2-6: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Vollständigkeit“

Konsistenz (auch Aktualität)	
Provider	Nutzer
<QG <QK >QP ^QP Übereinstimmung der Quellen von Ausgangs- und Zusatzdaten (z.B. aktuelle Daten oder Voreinstellungen oder Klimamittel etc.)	Erlaubt eine Einschätzung der Aktualität des Produktes bzgl. raumzeitlicher Start- und Rahmenbedingungen
<QP ^QP Ausreichende Grundgesamtheiten bei Statistiken (Mittelwerte, RMS etc.)	Erlaubt eine Einschätzung der Repräsentativität bzw. des Gewichtes von Statistiken
<QG >QG >QP ^QP Raumzeitliche Repräsentanz gleichmäßig (nicht 3 Pixel hier und 5000 dort)	Die räumliche und zeitliche Abdeckung mit Daten sollte möglichst gleichmäßig sein und Mindestpopulationen sollten erreicht werden. Die Anforderungen hierzu sind vorwiegend Nutzerangelegenheit.
<QM <QK >QM >QK Versionstreue entlang von Zeitreihen	Prozessierungsversionen sollten nicht gemischt werden und wenn doch, dann sollte das bekannt und einschätzbar sein.

Tabelle 2-7: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Konsistenz“

Genauigkeit	
Provider	Nutzer
<QP Genauigkeitsangaben auf Grund von Validierungsergebnissen	Kombinierte Information zur Genauigkeit des Datenproduktes → <QT
<QP <QG Genauigkeitsangaben auf Grund von Vergleichen	
<QP <QG Genauigkeitsangaben auf Grund von Abschätzungen	
<QP >QP ^QP Genauigkeitsangaben auf Grund von Mess- und Kalibrationsfehlern	Kann zur kombinierten Genauigkeitsinformation beitragen und/oder im Einzelfall zusätzliche Angaben erfordern, z.B. bei einer zeitlich befristeten Mess-Störung etc.

Tabelle 2-8: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Genauigkeit“

Gültigkeit	
Provider	Nutzer
<QQ <QG <QM >QQ >QG >QM Zeitliche und räumliche Einschränkungen	z.B. Tag/Nacht-Gültigkeit oder regionale Gültigkeit etc.

Gültigkeit	
Provider	Nutzer
<QM >QM Zweckbindung und sonstige Aussagen zu Bedingungen der Nutzbarkeit	Nach Vereinbarung zwischen Nutzer und Provider
<QK >QK Zugehörigkeit zu einem bestimmten Prozess mit daraus resultierenden Einschränkungen	
<QM Bedingungen der Interpretierbarkeit	Nach physikalisch-technischer Dokumentation und/oder Erfahrung von Nutzer und/oder Provider
<QM Juristische Einschränkungen	Nach Auflagen durch Provider, Nutzer oder Dritter, z.B. Datenbesitz- und Nutzungsrechte

Tabelle 2-9: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Gültigkeit“

Plausibilität	
Provider	Nutzer
>QP ^QP Ausreißermarkierung	Evtl. kombinierte Kennzeichnung nach Vereinbarung zwischen Provider und Nutzer
>QP ^QP Markierung typischer und untypischer Werte	
>QP ^QP Erkennung unsinniger Werte	
<QK <QM >QK >QM Kompatibilität	Betrifft Prozessorkonfiguration und – ablaufprotokoll sowie die Widerspruchsfreiheit verschiedener QS-Metadaten

Tabelle 2-10: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Plausibilität“

Überprüftheit	
Provider	Nutzer
^QP Überprüftheitsstatus von Einzelwerten in Bezug auf definierte Klassen, Kategorien, Typisierungen etc.	Einfache Aussagen, ob und auf welche Klassen, Kategorien, Typisierungen etc. überprüft bzw. nicht überprüft wurde
<QM >QM Überprüftheitsstatus in Bezug auf die Weiterverwendung (durch Tests)	z.B. Test auf Erfüllung von Bedingungen wie Definitionsbereiche etc. für Eingabedaten
<QM Überprüftheitsstatus von Zeitreihen (durch Tests)	Gemäß vereinbarter Anforderungen
<QK Status in Bezug auf die Prozessierung (operationell, experimentell, prototypisch etc.)	Betrifft Glaubwürdigkeit, Robustheit und wissenschaftliche Akzeptanz des Verfahrens
<QM Existenz von Validierungen oder anderen Referenzen	Betrifft Glaubwürdigkeit und wissenschaftliche Akzeptanz des Verfahrens

Tabelle 2-11: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Überprüftheit“

Bestimmtheit	
Provider	Nutzer
^QP (Grad der) Klassifikationssicherheit	Wahrscheinlichkeitsangaben für die Klasseneinordnung für den Nutzer
>QM Wahrscheinlichkeit alternativer Interpretationen	
^QP Statistische Signifikanzen	Verschiedene statistische Tests
^QP Unbestimmtheit im Sinne von Ergebnislosigkeit	Informationen, falls kein Ergebnis trotz Anwendung physikalischer Methoden erzielt wurde

Tabelle 2-12: Vergleich Provider-/Nutzersicht – Qualitätsklasse „Bestimmtheit“

2.3. QS-Metadaten für energiemeteorologische Produkte

In diesem Abschnitt werden die QS-Metadaten, die für energiemeteorologische Produkte benötigt werden, spezifisch beschrieben. Dies beinhaltet erstens eine Inhaltsangabe, zweitens die Zugehörigkeit zu den Qualitätsklassen, drittens eine Prioritätsangabe und außerdem einen Hinweis zur Standardisierbarkeit hinsichtlich der Norm ISO-19115. Darüber hinaus werden Verfahren angegeben oder empfohlen, wie die QS-Metadaten abgeleitet werden können. Als konkretes Anwendungsbeispiel für QS-Metadaten dient die HELIOSAT-Prozesskette (Abbildung 2-2) bis zu der Stufe, wo die Wolken- und Wasserdampfprodukte (MSG.SEVIRI.CLOUDS und MSG.SEVIRI.TWC) per FTP an T&M versandt werden. Die Kette beinhaltet bis dorthin die Prozessierung der SEVIRI-Daten von Level-0 zu Level-1.5, wofür der SCENES-Prozessor von EUMETSAT verwendet wird, die zusätzliche Einbeziehung von Ozon- und Aerosoldaten, letztere aus dem SYNAER-Prozessor, sowie den HELIOSAT-3-Prozessor für Atmosphärenparameter. Letzterer besteht aus dem APOLLO/SEVIRI-Prozessor, der die Wolkenerkennung und die benötigten Wolkenparameter liefert, sowie aus dem Prozessorteil, der die eigentlichen Atmosphärenprodukte fertig rechnet und zusammenstellt. Der APOLLO-Prozessor wird mit Eingabedaten aus SCENES gespeist und anschließend werden die APOLLO-Wolkenprodukte noch einer Nachprozessierung unterzogen und in die HDF-Produktdateien gepackt (Abbildung 2-3). Welche notwendigen und optionalen QS-Metadaten für die Produkte der HELIOSAT-Prozesskette erzeugt werden sollen, wird anhand der Betrachtung der Eingangsdaten, der prozeduralen Einzelschritte und Kontextgegebenheiten, der Qualität von Zwischenergebnissen und des a priori bekannten Wissens über die Verfahrensgüte erörtert und die beschreibenden QS-Metadaten werden definiert.

QS-Metadaten werden grundsätzlich auf drei Wegen, je nach Qualitätsklasse, erzeugt, nämlich

- nach physikalisch-analytischen Verfahren, z.B. durch Validierung und nach dem Fehlerfortpflanzungsgesetz,
- nach numerisch-statistischen Verfahren, z.B. durch Betrachtung von Abweichungen bei Vergleichen und Berechnung von Signifikanzen,
- sowie rein empirisch nach Erfahrungswerten, Vollständigkeitsanforderungen und Plausibilitätsbetrachtungen.

Bei der Erzeugung energiemeteorologischer Produkte ist der Provider tendenziell eher auf Statistik und Empirie angewiesen und wird seltener echte physikalische Qualitätsmaße oder Validierungen zur Verfügung haben. Die dabei erforderlichen Erfahrungswerte stammen naturgemäß meist vom Provider. Der Nutzer wird meist nicht von sich aus über einen solchen Erfahrungsschatz verfügen und muss sich daher auf die Providerangaben verlassen. Andererseits kann der Provider bei Qualitätsaussagen, die auf Erfahrung und Empirie beruhen, keine weiter reichenden Garantien gegenüber dem Nutzer übernehmen. Die QS-Metadaten geben also nur dann im Sinne einer Garantie belastbare Informationen, wenn sie ausschließlich mit physikalischen Verfahren erzeugt wurden oder auf Validierungen beruhen. In allen anderen Fällen geben QS-Metadaten nur Anhaltspunkte und (begründete) Einschätzungen der Güte energiemeteorologischer Produkte. Dieser Aspekt wird im Abschnitt 4 noch weiter diskutiert. Bei der Definition der QS-Metadaten für die HELIOSAT-Prozesskette wird im Folgenden integrativ vorgegangen. Zunächst werden modular für einzelne Prozessierungsschritte und Zwischenergebnisse die QS-Informationen identifiziert, die verfügbar gemacht werden können. Danach werden die Einzelinformationen sinnvoll anhand der Qualitätsklassen und den Nutzeranforderungen Rechnung tragend zu QS-Metadaten verdichtet. Dabei darf diese Vorgehensweise als allgemeine Empfehlung für die Definition von QS-Metadaten energiemeteorologischer Produkte verstanden werden.

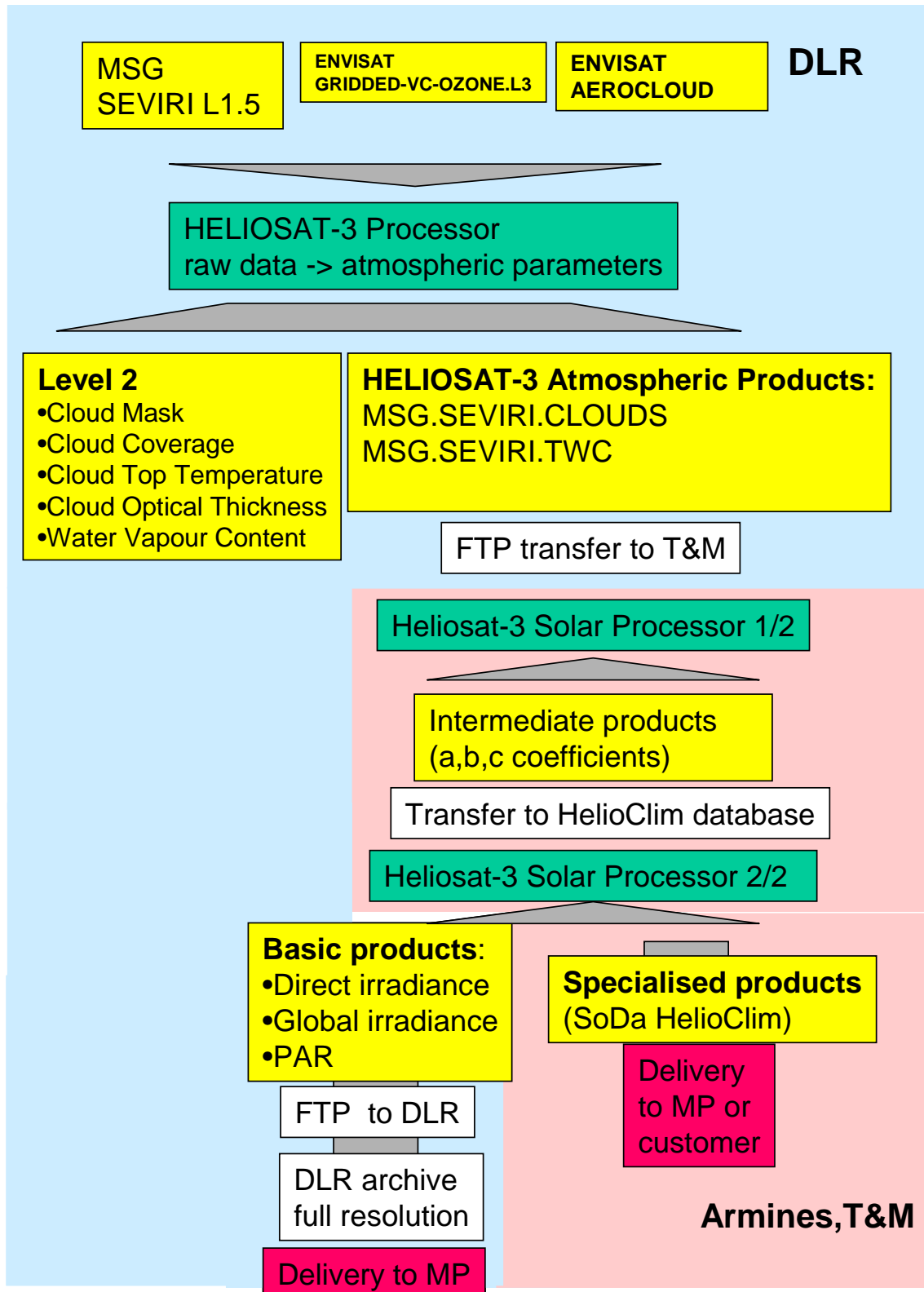


Abbildung 2-2: HELIOSAT-3 Prozesskette bei DFD und EdM

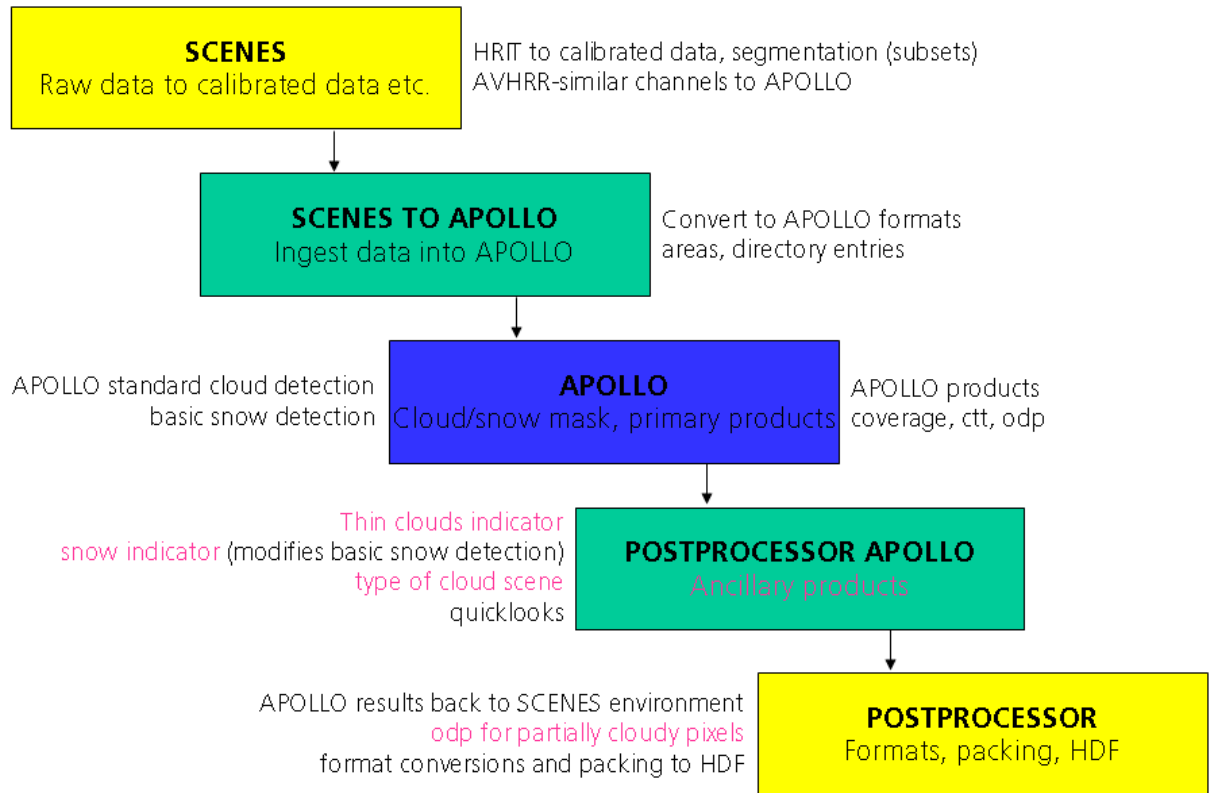


Abbildung 2-3: SCENES/APOLLO-Prozessierungskette für Wolkenparameter beim DFD

2.3.1. QS-Informationen der HELIOSAT-Prozesskette (DFD-Teil)

In Tabelle 2-13 werden alle im DFD-Teil der HELIOSAT-Prozesskette verfügbaren QS-Informationen aufgelistet und dabei nach Prozessierungsschritten und (Zwischen)Produkten gegliedert. Danach werden alle diese Informationen unter Angabe ihrer Referenznummer näher erläutert. QS-Informationen, die auf echten physikalischen Analysen oder Validierungen beruhen oder darauf prinzipiell begründet werden können, sind farblich hervorgehoben. Grün bedeutet, dass eine physikalische Herleitung oder Validierung der Genauigkeit tatsächlich existiert, blau, dass dies zwar möglich, jedoch (noch) nicht durchgeführt ist. Alle anderen Informationen beruhen auf rein statistischen Ableitungen oder empirischen Erfahrung. Der Begriff „Güte“ wird im Folgenden weitgehend gleichbedeutend mit dem Begriff „Genauigkeit“ verwendet, bezeichnet also eine absolute oder relative Fehlerangabe. Ein kleiner Unterschied besteht lediglich darin, dass „Güte“ sich mehr auf eine ganze Methodik, „Genauigkeit“ aber mehr auf einzelne Parameter bezieht.

Prozessierungsschritt	Zwischenprodukt	Verfügbare QS-Information	#
ENVISAT Ozondaten Input	Gridded vc ozone	Raumzeitliche Aktualität der vertical columns	1.1.1
		Klimawerte, Schätzwerte verwendet? (Vollständigkeit)	1.1.2
		Güte falls bekannt	1.1.3
SYNAER-Prozessor Aerosol Input	AOT	Raumzeitliche Aktualität (Verfügbarkeit)	2.1.1
		Klimawerte, Schätzwerte verwendet? (Vollständigkeit)	2.1.2
		Güte und Zuverlässigkeit falls bekannt	2.1.3
SCENES Level 1.5	Kalibrierte Kanäle – Reflexionsgrade und	Fehlzeilen und Badvalues (Vollständigkeit)	3.1.1

Prozessierungsschritt	Zwischenprodukt	Verfügbare QS-Information	#
	Temperaturen	Kalibrationsgüte	3.1.2
	BRDF-Winkel	Verfügbarkeit	3.2.1
	Land/Wasser-Maske	Verfügbarkeit	3.3.1
	Georeferenzierung	Verfügbarkeit	3.4.1
	Subset-Information	Verfügbarkeit	3.5.1
SCENES to APOLLO	Georeferenzierung	Plausibilität und Genauigkeit	4.1.1
	BRDF-Winkel	Plausibilität und Genauigkeit	4.2.1
	APOLLO Areas und Directories	Vollständigkeit	4.3.1
		Plausibilität der Wertebereiche	4.3.2
APOLLO	BRDF-Winkel	(nicht oder eingeschränkt) nutzbare Bereiche	5.1.1
	Land/Wasser/Sunglint-Maske	Einschränkungen Sunglint	5.2.1
		Einschränkungen Küste	5.2.2
	Wolken/Schnee-Maske, Erkennungsalgorithmen	Vollständigkeit Algorithmen	5.3.1
		Aktualität der dynamischen Schwellen	5.3.2
	Bedeckungsgrad	nicht zum Wolkenstockwerk zugeordnete Pixel	5.4.1
		Verfahrensgüte	5.4.2
	Optische Dicke Wolken	Plausibilität	5.5.1
		Zuverlässigkeit in Abhängigkeit von den BRDF-Winkeln	5.5.2
		interpolierte Werte bei Teilbewölkung	5.5.3
		Verfahrensgüte	5.5.4
	Temperatur bzw. Höhe der Wolkenobergrenze	Aktualität des Temperaturprofils bei Bestimmung der Wolkenhöhe	5.6.1
	Postprozessor APOLLO	Indikator dünne Wolken	Plausibilität bei Vergleich mit wolkenoptischer Dicke
Wolkentyp		unklassifizierbare Pixel	6.2.1
		Grad der Klassifikationssicherheit	6.2.2
Schnee-Indikator		Bestimmtheit der benutzten dynamischen Schneeschwelle	6.3.1
		Schnee-Indikator selbst als Schnee-Wahrscheinlichkeitsmaß	6.3.2
Quicklooks		Verfügbarkeit	6.4.1
Postprozessor Atmosphärenprodukte	Optische Dicke Wolken für teilbewölkte Pixel	Kennzeichnung der interpolierten Werte	7.1.1
	TWC	Plausibilität bei Vergleich mit Wolkentyp und evtl. mit LWP	7.2.1
	Metadatensammlung	Vollständigkeit der essentiellen (QS)-Metadaten	7.3.1
		Kombinierte QS-Metadaten-Informationen als Vorbereitung für Transformation in QS-Nutzerinformationen	7.3.2

Tabelle 2-13: Verfügbare QS-Informationen der HELIOSAT-Prozesskette (DFD-Teil)

Im nächsten Abschnitt wird im Detail auf die in Tabelle 2-13 aufgeführten verfügbaren QS-Informationen eingegangen:

1.1.1

Unter der Aktualität der verwendeten vertikalen Ozonsäulen ist die zeitliche und räumliche Nähe zu den übrigen für die Erstellung des Gesamtproduktes benötigten (Satelliten-)Daten zu verstehen. Die Qualität des Endproduktes wird davon beeinflusst, wie weit die Ozonsäule vom jeweiligen Messwert (Pixel) entfernt ist und wie viel Zeit zwischen der Erhebung der Ozoninformation und der anderen Daten liegt. Falls eine räumliche und/oder zeitliche Interpolation verwendet wurde, bedeutet ein Vermerk hierzu ebenfalls eine Qualitätsinformation.

1.1.2

Falls keine vertikalen Ozonsäulen verfügbar sind, können eventuell auch Klimamittelwerte oder Schätzwerte benutzt werden. Da die aktuellen Ozonsäulen von mittleren mehr oder weniger stark abweichen können, ist diese Information zur Beurteilung der Qualität des Endproduktes wichtig. Es können einzelne Werte fehlen oder ganze Ozon-Datensätze. Diese Angaben zur Vollständigkeit der Ozoninformationen hängen auch ab von deren Aktualität, d.h. ab einer bestimmten zeitlichen und räumlichen Distanz zwischen Produkt-Referenzzeit und -Ort und der nächsten Ozonsäule gilt die Ozoninformation als nicht aktuell und muss durch A-Priori-Wissen ersetzt werden. Welche Schwellen hierfür verwendet werden, kann allein im Ermessen des Providers liegen oder zwischen Provider und Nutzer vereinbart sein. 1.1.1 und 1.1.2 sollten stets gemeinsam und abgestimmt angegeben werden.

1.1.3

Ozonsäulen-Daten werden laufend mit Bodenmessungen des globalen Dobson und Brewer Netzwerkes verglichen. So können Effekte erkannt und eingeschätzt werden, die aus jahreszeitlichen Effekten, aus Sonnenstandeffekten oder aus der Sensordegradation resultieren. In der Regel sind die Vergleichsergebnisse geophysikalisch konsistent und es zeigen sich beim Vergleich mit den Bodenmessungen in Abhängigkeit vom Sonnenzenit Abweichungen zwischen 2% und 5%. Validierungs-Methoden und -Ergebnisse hierzu sind in Lambert et al, 2004, und von Bargaen et al, 2004, publiziert.

2.1.1

Für die Aktualität der Aerosolinformation gilt gleiches wie unter 1.1.1 für die vertikalen Ozonsäulen beschrieben. Allerdings können nicht verfügbare aktuelle Aerosolinformationen nicht einfach durch Schätzwerte oder Klimamittel ersetzt werden. Das liegt daran, dass Klimamittel für den Aerosolgehalt der Atmosphäre einerseits auf einer schmalen Datengrundlage beruhen, andererseits Aerosol sehr viel inhomogener verteilt ist und auch zeitlich stärker variiert. Bei der Aerosolinformation wird also die Aktualität und Vollständigkeit zu einer QS-Information zusammengefasst.

2.1.2

Analoges wie für Ozonsäulen unter 1.1.2 gilt auch für aerosol-optische Dicken.

2.1.3

Die Aerosol-Ergebnisse aus SYNAER ENVISAT wurden validiert gegenüber multi-spektralen bodengebundenen Sonnen-Photometer-Messungen aus dem AERONET-Netzwerk. Die Resultate sind unter <http://www.gse-promote.org> → "Air Quality" → "Satellite-based Particulate Matter Demonstration Service" → "European Multi-annual Records" dargestellt.

3.1.1

Im Level-1.5-Prozessor der SCENES-Software werden die MSG-SEVIRI-Rohdaten zu Strahldichten kalibriert und daraus für die solaren Kanäle Reflexionsgrade und für die thermalen Kanäle Temperaturen erzeugt. Auf der Ebene der Level-1.5-Daten genügt es, fehlende Pixel und Zeilen sowie Fehler bei der Kalibration und der Ableitung der Reflexionsgrade und Temperaturen (Badvalues) zu kennzeichnen. Die Plausibilität, das heißt der physikalische Sinn der abgeleiteten Werte wird auf der Schnittstelle zu APOLLO (Gruppe 4.x.x) überprüft.

3.1.2

Die Genauigkeit der verwendeten Kalibrationskoeffizienten wird von EUMETSAT angegeben. Eine Angabe zur Aktualität der Kalibrationskoeffizienten wird empfohlen, da Sensorsysteme altern, die Empfindlichkeit sich dabei kontinuierlich verändert, aber die Koeffizienten nur epochenweise aktualisiert werden.

3.2.1, 3.3.1, 3.4.1, 3.5.1

Neben den SEVIRI-Level-1.5-Daten liefert SCENES auch noch für die weitere Prozessierung notwendige Zusatzdaten.

Die BRDF-Winkel (3.2.1), das sind Sonnen- und Satellitenzenit sowie relativer Azimut zwischen Sonne und Satelliten werden in APOLLO für die Korrektur der Reflexions-Anisotropie benötigt. Die Land/Wasser-Maske (3.3.1) ist nötig, weil die Wolkenerkennungsalgorithmen über Wasser und über Land verschieden sind. Eine Zuordnung von Erdkoordinaten zu jedem Pixel wird trivialerweise auch gebraucht, d.h. SCENES liefert diese Georeferenzierung (3.4.1) in Form von Dateien mit geografischen Breiten- und Längenangaben für jedes Pixel. Zusätzlich wird auch noch eine Angabe zu Lage und Größe des zu bearbeitenden Ausschnittes (Subset, 3.5.1) von SCENES an APOLLO übergeben. Im Prozessierungsabschnitt SCENES werden diese Zusatzdaten noch nicht inhaltlich auf ihre Qualität geprüft, sondern es wird hier zunächst nur die Verfügbarkeit sicher gestellt. Dabei geht es

weniger darum, ob die Daten überhaupt vorliegen, denn liegen sie nicht vor, kann nicht mehr weiter prozessiert werden. Vielmehr ist bezüglich der Qualität des Endproduktes wichtig, welche Zusatzdaten-Versionen verfügbar sind.

4.1.1

In dem Prozessierungsschritt „SCENES to APOLLO“ (Gruppe 4.x.x), der die Schnittstelle zur Datenübergabe von SCENES nach APOLLO darstellt, werden sowohl Sensordaten als auch Zusatzdaten überprüft und es können QS-Informationen abgeleitet werden. Die Plausibilität der Georeferenzierung wird einfach dadurch getestet, dass Breiten- und Längengrade außerhalb der MSG-Scheibe nicht vorkommen dürfen. Eine automatische Überprüfung der Genauigkeit in aktuell laufender Prozessierung ist mit den implementierten algorithmischen Verfahren nicht möglich. Hier kann nur eine interaktive vorherige Kontrolle per Augenschein erfolgen, die dann lediglich als durchgeführt vermerkt werden kann.

4.2.1

Für die Genauigkeit der BRDF-Winkel gilt das Gleiche wie unter 4.1.1. ausgeführt für die Georeferenzierung. Die Plausibilität der BRDF-Winkel wird unter Berücksichtigung der Tageszeit und damit des Sonnenstandes der SEVIRI-Szene geprüft. Beispielsweise dürfen keine Sonnenzenitwinkel kleiner als 90° minus geografische Breite des Pixels plus Sonnendeklination vorkommen. Außerdem erwartet APOLLO bestimmte Wertebereiche, nämlich den Sonnenzenit zwischen 0° und 180° , den Satellitenzenit innerhalb der MSG-Scheibe zwischen 0° und 90° und den relativen Azimut zwischen 0° und 180° .

4.3.1

Die SEVIRI-Sensordaten der 5 AVHRR-ähnlichen Kanälen und die Land/Wasser-Maske, die als Binärdaten von SCENES an APOLLO übergeben werden, werden in so genannte APOLLO-Areas transformiert. Bei dieser Gelegenheit kann auch wieder die Vollständigkeit der Daten, wie sie für die Ableitung der Wolkenparameter benötigt wird, geprüft und angegeben werden. Da beim geostationären Satelliten die Land/Wasser-Maske zeitlich konstant ist, genügt für eine Überprüfung ihrer Vollständigkeit beispielsweise ein einfaches Zählen der Land- und Wasserpixel. Bei der Bestimmung der Vollständigkeit der SEVIRI-Sensordaten für die Prozessierung in APOLLO können die QS-Informationen aus 3.1.1 verwendet werden. Danach kann die unter 4.3.2 durchgeführte Plausibilitätsprüfung dazu führen, dass nicht plausible Pixelwerte ebenfalls als nicht verfügbar in der QS-Information angesehen werden.

4.3.2

Beim Anlegen der APOLLO-Areas mit den Sensordaten werden bestimmte Wertebereiche erwartet und Pixel, die aus diesem Bereich herausfallen, werden auf die untere bzw. obere Grenze des jeweiligen Bereiches begrenzt (Clip-Value). Für die Reflexionsgrade bei $0.6 \mu\text{m}$ und $0.9 \mu\text{m}$ wird auf das Intervall zwischen 0.005 und 2.000 begrenzt, bei $1.6 \mu\text{m}$ bzw. $3.7 \mu\text{m}$ auf das Intervall zwischen 0.005 und 1.000. Für die Temperaturen bei $11 \mu\text{m}$ und $12 \mu\text{m}$ wird das Intervall 170 K bis 330 K verwendet. Pixelwerte außerhalb dieser Intervall, die auf die jeweiligen Intervallgrenzen gesetzt werden, können als Ausreißer markiert werden.

5.1.1

Die BRDF-Winkel werden im APOLLO-Teil der HELIOSAT-Prozesskette gebraucht, um gegen die Anisotropie korrigierte Reflexionsgrade, vor allem von Wolken, zu berechnen. Dies ist eine notwendige Voraussetzung, um z. B. wolken-optische Dicken aus den Sensordaten ableiten zu können. Die Genauigkeit der abgeleiteten Wolkenparameter hängt dabei vor allem vom Sonnen- und vom Sensorzenitwinkel ab. Zu größeren Zenitwinkeln hin, also bei tief stehender Sonne und/oder tief stehendem Satelliten, nimmt die Genauigkeit nichtlinear und progressiv ab. Zwar gibt es keine quantitativen Untersuchungen zur Größe dieses Genauigkeitsverlustes mit zunehmenden Zenitwinkeln, jedoch lässt sich hierfür ein qualitatives relatives Qualitätsmaß im Sinne von Nutzungseinschränkungen bzw. Nutzungshinweisen generieren.

5.2.1, 5.2.2

Die Information über die Land- und Wasserverteilung in der SEVIRI-Szene wird in APOLLO zusammen mit einem Vermerk über die eventuelle Lage eines Pixels im Bereich des direkten Sonnenreflexes (Sunglint) abgespeichert. Diese Datei wird Land/Wasser/Sunglint-Maske genannt und enthält die genannte Information für jedes einzelne Pixel. Für alle Wasserpixel wird zunächst mit Hilfe der zugehörigen BRDF-Winkel geprüft, ob geometrisch ein Sonnenreflex auftreten kann. Die Pixel, die in diesem Bereich potentiellen Sonnenreflexes liegen, werden danach auf Grund ihrer spektralen

Signatur in den Sensorkanälen als reflexkontaminiert oder reflexfrei eingestuft. Zusätzlich wird in der Datei entlang der Übergangslinie Land-Wasser ein Küstenbereich von zwei Pixeln Breite definiert. Die Information, ob ein Pixel über Land, Küste oder Wasser liegt, sowie die Information, ob ein Wasserpixel im Sonnenreflex liegt oder nicht, entscheidet zum einen, welche Wolkenerkennungs-Algorithmen mit welchen Parametern laufen. Andererseits sind die bei der Anisotropiekorrektur verwendeten Koeffizienten für Land und Wasser unterschiedlich. Bei Pixeln, die als reflexkontaminiert gekennzeichnet sind, werden die Wolkenerkennungs-Algorithmen, die solare Kanäle verwenden, abgeschaltet. Einige Wolkenerkennungs-Algorithmen müssen wegen der Gefahr von Fehlklassifikationen im Küstenbereich abgeschaltet werden. Diese Einschränkungen im Bereich des Sonnenreflexes und über Küstenpixeln bedeuten eine reduzierte Effizienz bei der Wolkenerkennung und sollten in Form von QS-Metadaten vermerkt sein.

5.3.1, 5.3.2

Das Endergebnis nach dem Lauf aller Wolkenerkennungs-Algorithmen in APOLLO ist eine Datei, die die kombinierte Information über Wolken- und Schneekontamination der Pixel enthält. Diese Wolken/Schnee-Maske ist zusammen mit der Land/Wasser-Maske und den Sensordaten Ausgangsinformation für die nun folgende Berechnung der Wolkenparameter, z.B. der optischen Dicke. Wichtig für die Qualität des Endproduktes ist die vollständige Anwendung aller möglichen Wolkenerkennungs-Algorithmen sowie des Wolken/Schnee-Unterscheidungsalgorithmus. Normalerweise werden immer alle Algorithmen angewandt und im Fehlerfall die gesamte Prozessierung verworfen. Wie unter 5.2.1 und 5.2.2 erläutert, sind aber bei Küstenpixeln und im Sonnenreflex verfahrensbedingt Unvollständigkeiten gegeben. Die Gruppe 5.2.x und 5.3.1 sind also teilweise redundant, jedoch kann unter 5.2.x zwischen Küste und Sonnenreflex unterschieden werden. Die meisten der Wolkenerkennungs-Algorithmen sowie der Wolken/Schnee-Unterscheidungsalgorithmus bestimmen sich ihre Abfrageschwellen dynamisch aus den aktuellen Daten heraus. Falls das scheitert, d.h. falls gemäß bestimmter Kriterien solche dynamischen Schwellen nicht abgeleitet werden können, werden Voreinstellungen benutzt. Dies beeinflusst die Qualität der Wolken/Schnee-Maske und soll in Form von QS-Metadaten vermerkt werden.

5.4.1, 5.4.2

Für den APOLLO-Wolkenbedeckungsgrad aus AVHRR-Daten gibt es eine publizierte Validierung (Kriebel et al, 2003) gegen synoptische Beobachtungsdaten vom Deutschen Wetterdienst. Das Ergebnis zeigt eine Genauigkeit des totalen Bedeckungsgrades im Vergleich mit Synopdaten von weniger als $\pm 1/8$ Bedeckung für 70 % der Pixel, d.h. mehr als zwei Drittel der zugehörigen Bodenbeobachtungen für alle Pixel weichen vom mit APOLLO bestimmten Bedeckungsgrad um weniger als 12,5 % ab. Eine solche Validierung stellt eine echte physikalische Grundlage dar für Aussagen zur Genauigkeit (5.4.2). Die QS-Information 5.4.1 ist dagegen eine reine innere Vollständigkeitsinformation, die für so genannte teilbewölkte Pixel anfällt. Da nur die vollständig bewölkten Pixel über ihre Temperatur einem Wolkenstockwerk zugeordnet werden können, die teilbewölkten wegen ihrer Mischtemperatur dagegen nicht, muss eine statistische Methode angewandt werden, um auch teilbewölkten Pixeln ein Wolkenstockwerk zuordnen zu können. Dies geschieht in kleinen lokalen Ausschnitten nach dem Mehrheitsprinzip, d.h. die teilbewölkten Pixel werden dem Wolkenstockwerk zugeordnet, das in der lokalen Umgebung am häufigsten den vollständig bewölkten zugeordnet worden ist. Dies gelingt manchmal nicht, etwa wenn in der lokalen Umgebung keine vollständig bewölkten Pixel oder zwei oder mehr Klassen gleich häufig vorkommen. Diese keinem Wolkenstockwerk zugeordneten Pixel zu kennzeichnen ist beispielsweise wichtig für eine spätere Berechnung der Schattenwirkung von Wolken aus deren angenommenen Höhe.

5.5.1, 5.5.2, 5.5.3, 5.5.4

Zusammen mit dem Wolkenbedeckungsgrad ist die optische Dicke von Wolken eine zentrale Ausgangsgröße für die Abschätzung der solaren Einstrahlung am Boden. Für die optische Dicke lassen sich vier wesentliche QS-Informationen angeben. Die Plausibilität der abgeleiteten Werte muss zunächst rein numerisch in einem vorgegebenen Rahmen zwischen 0 und 500 liegen. Physikalisch sind aber auch schon Werte oberhalb von 200 meist unrealistisch. Außerdem kann die Plausibilität der wolken-optischen Dicke vor allem beim Vergleich mit dem Indikator für dünne Wolken (6.1.1) und dem Wolkentyp (6.2.x) des APOLLO-Postprozessors überprüft werden. Beispielsweise passen hohe optische Dicken nicht zu dünnen Wolken, wenn nicht gleichzeitig als Wolkentyp die Klasse „multi-layer“ angegeben ist. Wie bereits unter 5.1.1 erwähnt hängt die Zuverlässigkeit vor allem der optischen Dicke von Wolken von der BRDF-Geometrie des Pixels ab. Hierzu sind relative, qualitative QS-Angaben möglich. Die optische Dicke wird in APOLLO nur für vollständig bewölkte Pixel gerechnet, da wiederum im teilbewölkten Fall nicht zwischen Strahlung vom Boden und von der Wolke unterschieden werden kann. Dennoch möchte man auch für teilbewölkte Pixel wenigstens eine

Abschätzung der optischen Dicke der in diesem Pixel vorhandenen Wolken haben. Dies wird ebenfalls durch Bezugnahme zur lokalen Pixelumgebung durchgeführt. Zur Zeit wird eine solche Extrapolation der optischen Dicken auf die teilbewölkten Pixel im Postprozessor für Atmosphärenprodukte erstellt (7.1.1), könnte aber prinzipiell auch schon auf der APOLLO-Stufe der HELIOSAT-Kette implementiert werden. Die QS-Information 5.5.3 hat also derzeit nur eine Schalterfunktion, die besagt, dass in APOLLO noch keine teilbewölkten optischen Dicken erzeugt wurden. Die Kennzeichnung der extrapolierten Werte wird also unter 7.1.1 gemacht. Die Güte des in APOLLO verwendeten Verfahrens zur Ableitung wolken-optische Dicken ist nicht physikalisch validiert, jedoch existiert eine theoretische Abschätzung der Genauigkeit (Kriebel et al, 1989) auf der Grundlage vernünftiger physikalischer Annahmen zu den Genauigkeiten aller Ausgangsgrößen und Algorithmen.

5.6.1

Der für die Bestimmung der Höhe der Wolkenobergrenze notwendige Parameter ist die Wolkenoberflächen-Temperatur. Mit Hilfe eines Temperaturprofils kann einer bestimmten Temperatur eine Höhe zugeordnet werden. Die besten Ergebnisse erzielt man dabei natürlich, wenn das jeweils aktuelle Temperaturprofil bekannt ist. Häufig ist dies aber nicht möglich und man muss sich mit Annahmen behelfen, etwa mit Modellatmosphären oder mit in numerischen Modellen vorhergesagten Profilen. Eine Angabe zu dem aktuell verwendeten Temperaturprofil stellt daher eine bedeutende QS-Information dar, da beispielsweise eine Temperaturabweichung von fünf Grad schon einer Höhenabweichung von etwa 500 m entspricht.

6.1.1, 6.2.1, 6.2.2, 6.3.1, 6.3.2, 6.4.1

Aus den im APOLLO-Kern erzeugten Masken und Wolkenparameter werden im APOLLO-Postprozessor weitere für die Interpretation und Weiterverarbeitung hilfreiche Informationen zur Art der Bewölkung und zur Schneekontamination abgeleitet. Hierzu gehört auch die Erzeugung von Quicklooks zur Augenscheinprüfung der Datenqualität.

Ein Indikator-Layer gibt für jedes Pixel an, ob die Anwesenheit einer dünnen Wolkenschicht festgestellt wurde. Die Plausibilität dieser Information kann gegengeprüft werden, indem man mit der korrespondierenden optischen Dicke vergleicht. Ist die optische Dicke der Wolken unerwartet hoch, muss angenommen werden, dass unter der dünnen Wolkenschicht eine weitere, optisch dickere Schicht liegt. In diesem Fall muss bei der Wolkentypangabe der Vermerk „multi-layer“ gesetzt werden. Die Angabe des Wolkentyps ist ebenfalls für jedes Pixel vorgesehen, ist aber stellenweise wegen unzureichender Informationen nicht möglich. Solche unklassifizierbaren Pixel sollen gekennzeichnet werden. Zusätzlich kann zur Wolkentyp-Information der Grad der Klassifikationssicherheit angegeben werden. Sehr wichtig für die weitere Prozessierung ist die Kenntnis der Schneekontamination der Pixel. Hierzu wird ein Indikator berechnet, der im Bereich zwischen den Zahlenwerten 0 und 100 die Wahrscheinlichkeit für Schneekontamination des Pixels angibt. Die eindeutige Kennzeichnung der Pixel als schneekontaminiert oder schneefrei erfordert aber die Festlegung auf einen bestimmten Schwellenwert für den ermittelten Schneeindikator. Dieser Schwellenwert wird aus dem Erwartungswert des Histogramms ermittelt und dem Produkt als Metadateneintrag zusätzlich zum Indikator-Layer mitgegeben. Der Schwellenwert kann je nach Aussehen des Histogramms mehr oder weniger bestimmt sein, was als QS-Information angegeben werden soll. Der Schnee-Indikator selbst ist als innerer QS-Informationen-Layer anzusehen. Die Generierung von Quicklooks schließlich stellt eine visualisierende QS-Information dar, deren Verfügbarkeit ebenfalls in den Metadaten vermerkt sein sollte.

7.1.1

Beim Postprozessor für die Atmosphärenprodukte werden, wie bereits unter 5.5.3 erwähnt, wolken-optische Dicken für teilbewölkte Pixel extrapoliert. Diese extrapolierten Werte sollen für die Qualitätskontrolle bei der weiteren Prozessierung gekennzeichnet werden.

7.2.1

Die Qualität des Säulen-Wasserdampfgehalts (TWC) kann durch Vergleich mit dem in APOLLO gerechneten Flüssigwasserweg und mit dem Wolkentyp-Indikator auf Plausibilität geprüft werden.

7.3.1, 7.3.2

Die Aufgabe des Postprozessors für die Atmosphärenprodukte ist es, die Endprodukte der HELIOSAT-Prozesskette auf der DFD-Seite zusammenzustellen. Dazu gehört auch die Überprüfung der QS-Metadaten-Information. Zumindest die essentiellen QS-Metadaten sollten vollständig sein, aber auch die Vollständigkeit der optionalen sollte vermerkt werden. Für die Transformation in QS-Nutzerinformationen ist es notwendig, Zusammenfassungen aller QS-Metadaten-Informationen zu erstellen. Wie genau dies geschehen soll, wird in Abschnitt 3 beschrieben.

2.3.2. QS-Metadaten-Definition für die HELIOSAT-Prozesskette (DFD-Teil)

In Tabelle 2-14 werden nun auf der Basis der verfügbaren QS-Metadaten-Informationen die QS-Metadaten selbst definiert, d.h. sie werden in der Spalte „Parameter“ namentlich aufgeführt. Metadaten können gemäß der Norm ISO-19115 definiert werden, wenn in ISO19115 eine Rolle (Bedeutung) definiert ist, nach der aus der QS-Information die QS-Metadatenätze konstruiert werden können. In diesem Fall ist der Name der ISO-19115-Rolle oder -Klasse in grüner Farbe angegeben. Darüber hinaus werden in diesem Dokument keine weiteren Schritte zur Erstellung der Metadatenstruktur gemäß der Norm angegeben, da dies den Rahmen sprengen würde und Gegenstand jeweiliger prozessor-spezifischer Dokumentationen ist. In der Spalte „QS-Information“ wird als Referenz die Nummerierung der Tabelle 2-13 angegeben. In der Spalte „Typ“ wird gemäß der in Abschnitt 2.2 eingeführten abkürzenden Kennzeichnung angegeben, zu welcher Produktkomponente die Metadaten gehören und ob es sich um innere oder äußere Metadaten handelt. Unter der „Priorität“ wird verstanden, ob die Angabe der Metadaten notwendig, also unverzichtbar für die Prozessierung und die QS-Nutzerinformation, oder optional, also ergänzend ist. Falls Metadaten mit einer Menge von konkreten (alpha)numerischen Werten belegt werden, ist deren Wertebereich angegeben. In der Spalte „Klasse“ wird angegeben, zu welcher der sieben unter Abschnitt 2.2 eingeführten QS-Metadaten-Klassen der Metadaten-Parameter gehört. Im Anschluss an Tabelle 2-14 werden noch Details zur Zuordnung der Metadaten-Entitäten zu den Produktkomponenten erläutert.

Die Tabelle 2-14 gibt die QS-Metadaten-Sammlung für die HELIOSAT-Prozessierungskette wieder. Für die anderen Prozessorteile bis zum Endprodukt „solar irradiance“ müssen weitere solche Sammlungen definiert werden. Dabei kann und soll aber in analoger Weise vorgegangen werden. Die hier für energiemeteorologische Produkte vorgeschlagene Systematik der QS-Metadaten-Definition wird in Abschnitt 5 in ihren wesentlichen Aspekten zusammengefasst.

Parameter	QS-Inf.	Typ	Priorität	Wertebereich	Klasse
DQ_TemporalValidityOzone	1.1.1	^QP <QP	notwendig	Zeit: 0 - outdated	Gültigkeit
DQ_OzoneDefault	1.1.2	^QP <QP	optional	estimated, climatological	Konsistenz
DQ_ThematicAccuracyOzone	1.1.3	<QP	notwendig	Bias, RMSE	Genauigkeit
DQ_TemporalValidityAerosol	2.1.1	^QP <QP	notwendig	Zeit: 0 - outdated	Gültigkeit
DQ_AerosolDefault	2.1.2	^QP <QP	optional	estimated, climatological	Konsistenz
DQ_ThematicAccuracyAerosol	2.1.3	<QP	notwendig	Bias, RMSE	Genauigkeit
DQ_CompletenessScenesLv15	3.1.1	^QP	notwendig	any badvalues	Vollständigkeit
DQ_ThematicAccuracyScenesCal	3.1.2	<QP	notwendig	Bias, RMSE	Genauigkeit
DQ_CompletenessScenesBRDF	3.2.1	<QG	notwendig	(un)available	Vollständigkeit
DQ_CompletenessScenesLSM	3.3.1	<QG	notwendig	(un)available	Vollständigkeit
DQ_CompletenessScenesGeo	3.4.1	<QG	notwendig	(un)available	Vollständigkeit
DQ_CompletenessScenesSubset	3.5.1	<QG	notwendig	(un)available	Vollständigkeit
DQ_PlausibilityApolloGeo	4.1.1	>QG	notwendig	(un)reasonable	Plausibilität
DQ_ThematicAccuracyApolloGeo	4.1.1	<QG	optional	mean deviation	Genauigkeit
DQ_PlausibilityApolloBRDF	4.2.1	>QG	notwendig	(un)reasonable	Plausibilität
DQ_ThematicAccuracyApolloBRDF	4.2.1	<QG	optional	mean deviation	Genauigkeit
DQ_CompletenessApolloAreas	4.3.1	^QP >QP >QG	notwendig	any badvalues	Vollständigkeit
DQ_PlausibilityApolloAreas	4.3.2	^QP	notwendig	(un)reasonable	Plausibilität
DQ_PositionalAccuracyApolloBRDF	5.1.1	^QP	optional	0% - 100%	Genauigkeit

Parameter	QS-Inf.	Typ	Priorität	Wertebereich	Klasse
DQ_PositionalValidityApolloGlint	5.2.1	^QP	optional	(no) glint	Gültigkeit
DQ_PositionalValidityApolloCoast	5.2.2	^QP	optional	(no) coast	Gültigkeit
DQ_CompletenessApolloAlgorithms	5.3.1	>QK	notwendig	name missing Algorithm	Vollständigkeit
DQ_ApolloDefaultAlgorithms	5.3.2	^QP >QK	optional	none, which default	Konsistenz
DQ_CertaintyApolloCloudLevel	5.4.1	^QP	optional	(un)classified	Bestimmtheit
DQ_ThematicAccuracyApolloCov	5.4.2	<QP	notwendig	mean deviation	Genauigkeit
DQ_PlausibilityApolloOdp	5.5.1	^QP	notwendig	(un)reasonable	Plausibilität
DQ_PositionalAccuracyApolloOdp	5.5.2	^QP	optional	0% - 100%	Genauigkeit
DQ_CompletenessApolloOdp	5.5.3	<QK	optional	totals only, all	Vollständigkeit
DQ_ThematicAccuracyApolloOdp	5.5.4	<QP	notwendig	mean deviation	Genauigkeit
DQ_ApolloDefaultCth	5.6.1	^QP >QK	optional	actual, model, predicted	Konsistenz
DQ_PlausibilityApolloThinIndex	6.1.1	^QP	optional	(un)reasonable	Plausibilität
DQ_CertaintyApolloCloudType	6.2.1	^QP	notwendig	(un)classified	Bestimmtheit
DQ_CertaintyApolloCloudClass	6.2.2	^QP	optional	0% - 100%	Bestimmtheit
DQ_CertaintyApolloSnowIndexThr	6.3.1	>QK	notwendig	threshold, RMS	Bestimmtheit
DQ_CertaintyApolloSnowIndex	6.3.2	^QP	notwendig	0% - 100%	Bestimmtheit
DQ_CompletenessApolloQL	6.4.1	>QQ	optional	(un)available	Vollständigkeit
DQ_CompletenessAtmoProdOdp	7.1.1	<QK	optional	totals only, all	Vollständigkeit
DQ_PlausibilityAtmoProdTWC	7.2.1	^QP	optional	(un)reasonable	Plausibilität
DQ_CompletenessAtmoProdMDSet	7.3.1	>QM	notwendig	(un)complete	Vollständigkeit
DQ_CorrectnessAtmoProdMDSet	7.3.2	>QM	optional	which quality proofs checked	Überprüftheit

Tabelle 2-14: QS-Metadaten-Definition für energiemeteorologische Produkte (DFD-Teil)

Die Einordnung und Darstellung der Metadaten-Parameter aus Tabelle 2-14 wird im Folgenden kurz begründet und erläutert:

1.1.1

Hier wird der zeitliche Abstand zwischen den verwendeten vertikalen Ozonsäulen und der aktuellen Szene angegeben. Falls dieser zeitliche Abstand für alle Pixel, d.h. für alle Säulenwerte derselbe ist, kann ein äußerer MD-Parameter für die Produkt-Primärkomponente angegeben werden, falls nicht, wird für jedes Pixel ein innerer MD-Parameter im Metadaten-Layer der Primärkomponente angegeben. Wird eine bestimmt, evtl. zwischen Nutzer und Provider vereinbarte Zeitschwelle überschritten, wird ein Wert eingetragen, der die Bedeutung „veraltet (outdated)“ hat.

1.1.2

Falls 1.1.1 den Wert „veraltet“ hat, kann hier angegeben werden, ob als Ersatz geschätzte oder klimatologische Ozonsäulen verwendet wurden. Dies kann wiederum pauschal durch eine äußere Metadaten-Angabe oder durch pixelweise innere Metadaten-Angabe in der Primärkomponente erfolgen.

1.1.3

Die Genauigkeit der Ozonsäulen wird durch Angabe der RMSE aus Vergleichen angegeben. Da dies a priori bekannt ist, handelt es sich um einen äußeren MD-Parameter der Primärkomponente.

2.1.1

Angaben und Einordnung hierzu werden analog zu 1.1.1 verstanden.

2.1.2

Angaben und Einordnung hierzu werden analog zu 1.1.1 verstanden.

2.1.3

Angaben und Einordnung hierzu werden analog zu 1.1.1 verstanden.

3.1.1

Hier wird pixelweise das Fehlen von Pixeln oder ganzen Zeilen mit Hilfe von „Badvalues“ im inneren QS-Metadaten-Layer der Primärkomponente vermerkt.

3.1.2

Die Genauigkeit der in SCENES angewandten Kalibration sollte a priori bekannt sein und wird durch Bias und RMSE angegeben.

3.2.1

Im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente wird die Verfügbarkeit der BRDF-Winkel vermerkt. In der normalen Prozessorversion führt das Fehlen dieser Information zum Abbruch der Prozessierung. Es ist aber für spätere Versionen denkbar, die BRDF-Winkel aus Erdkoordinaten und Sensorgeometrie hinreichend genau zu berechnen. Voraussetzung hierfür ist aber, dass die Sensordaten jedenfalls in Erdkoordinaten orientierbar sind.

3.3.1

Im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente wird die Verfügbarkeit der Land/Wasser-Maske vermerkt. In der normalen Prozessorversion führt das Fehlen dieser Information zum Abbruch der Prozessierung. Die Land/Wasser-Maske könnte aber auch aus den Erdkoordinaten erzeugt werden.

3.4.1

Im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente wird die Verfügbarkeit der Georeferenzierung, also der Zuordnung von Erdkoordinaten zu den Pixeln, vermerkt. Ist diese Zuordnung nicht möglich, muss die Prozessierung entweder abbrechen oder das Endergebnis ist unsinnig.

3.5.1

Im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente wird die Verfügbarkeit der Lage und Größe des zu prozessierenden Ausschnittes (Subset) vermerkt. Ist eine Subset-Angabe nicht möglich, bricht die Prozessierung entweder ab oder es wird die komplette METEOSAT-Scheibe bzw. der gesamte Eingabebereich prozessiert.

4.1.1

Im inneren QS-Metadatenatz der Georeferenzierungs-Komponente wird das Ergebnis der Plausibilitätsprüfung der Erdkoordinatenzuordnung abgelegt. Falls die Präzision dieser Zuordnung auf der Ebene der Schnittstelle zwischen SCENES und APOLLO bekannt ist, wird eine mittlere Abweichung im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente vermerkt.

4.2.1

Im inneren QS-Metadatenatz der Georeferenzierungs-Komponente wird das Ergebnis der Plausibilitätsprüfung der BRDF-Winkel abgelegt. Falls deren Genauigkeit auf der Ebene der Schnittstelle zwischen SCENES und APOLLO bekannt ist, wird eine mittlere Abweichung im äußeren QS-Metadatenatz der Georeferenzierungs-Komponente vermerkt. Ungenauigkeiten in der Georeferenzierung selbst wirken sich nur in Horizontnähe (von Sonne und/oder Satellit) gravierend auf die BRDF-Winkel aus.

4.3.1

Hier wird die unter 3.1.1 gegebene Information auf die APOLLO-Areas abgebildet, d.h. die so genannten „Badvalues“ werden in die Areas selbst eingetragen. Zusätzlich wird im internen QS-Metadatenatz der Primärkomponente und der Quicklook-Komponente eine Zusammenfassung eingetragen. Der Eintrag in der Quicklook-Komponente ist sinnvoll, da sich das Fehlen von Pixeln in den Sensordaten auf die Quicklooks selbst auswirkt.

4.3.2

Pixelwerte, die aus dem als sinnvoll angesehenen Wertebereich herausfallen, werden im QS-Metadaten-Layer der Primärkomponente markiert.

5.1.1

Die BRDF-Winkel beeinflussen die Genauigkeit der abgeleiteten optischen Wolkenparameter in der Weise, dass etwa mit zunehmenden Zenitwinkeln deren Genauigkeit abnimmt. Dafür kann ein relatives Maß angegeben werden, ausgedrückt in Werten zwischen 0% und 100%. 100% bedeutet, dass an diesem Pixel die größtmögliche Genauigkeit erreicht werden kann, bei Werten darunter aber nur ein entsprechender Anteil.

5.2.1

Da im Bereich des direkten Sonnenreflexes die solaren Wolkenerkennungs-Algorithmen abgeschaltet werden, ist dort damit zu rechnen, dass die Wolkenerkennung weniger zuverlässig funktioniert. Pixel, an denen ein Sonnenreflex erkannt wurde, können daher gekennzeichnet werden, was im Prinzip der Übertragung eines Informationsauszuges aus der APOLLO-Land/Wasser-Maske in den QS-Metadaten-Layer der Primärkomponente entspricht.

5.2.2

Einige Wolkenerkennungs-Algorithmen funktionieren nicht im unmittelbaren Küstenbereich und müssen daher dort abgeschaltet werden. Daher können solche Küstenpixel analog zu den Sonnenreflex-Pixeln in 5.2.1 (die Küstenpixel sind ebenfalls in der Land/Wasser-Maske enthalten) gekennzeichnet werden.

5.3.1

Wurde ein als wesentlich angesehener Wolkenerkennungs-Algorithmus oder die Schnee/Wolken-Unterscheidung nicht angewandt, muss dies im internen QS-Metadatensatz der Prozessierungs-Kontext-Komponente vermerkt werden.

5.3.2

Falls Wolkenerkennungs-Algorithmen nicht in der Lage waren, dynamische Schwellenwerte zu ermitteln, müssen Voreinstellungen (Defaults) verwendet werden. Diese Defaults können im QS-Metadatenlayer der Primärkomponente angegeben werden. Im QS-Metadatensatz der Prozessierungs-Kontext-Komponente kann dazu eine Zusammenfassung gegeben werden.

5.4.1

Können teilbewölkte Pixel keinem Wolkenstockwerk zugeordnet werden, kann dies pixelweise vermerkt werden.

5.4.2

Die aus einer Validation mit Hilfe von Synop-Daten gewonnenen Genauigkeitsaussagen für den APOLLO-Wolkenbedeckungsgrad (Kriebel et al, 2003) werden im äußeren QS-Metadatensatz der Primärkomponente vermerkt.

5.5.1

Die Plausibilität der Werte für die optische Dicke von Wolken sollte pixelweise festgestellt werden. Werte, die aus dem als sinnvoll angesehenen Wertebereich herausfallen oder mit der Typisierung von Wolken konfliktieren, werden im QS-Metadaten-Layer der Primärkomponente markiert.

5.5.2

Das unter 5.1.1 beschriebene relative Genauigkeitsmaß für die unter Nutzung der BRDF-Winkel abgeleiteten optischen Wolkenparameter wird hier analog angewandt.

5.5.3

Im äußeren QS-Metadatensatz der Prozessierungs-Kontext-Komponente kann vermerkt werden, ob während der APOLLO-Prozessierung eine Extrapolation der wolken-optischen Dicke für teilbewölkte Pixel durchgeführt wurde.

5.5.4

Für die mit dem in APOLLO implementierten Verfahren abgeleiteten wolken-optischen Dicken gibt es eine Genauigkeitsabschätzung auf der Basis von plausiblen physikalischen Annahmen zu den

möglichen Fehlerquellen (Kriebel et al, 1989, enthält diese Angaben für den Wolkenparameter „Flüssigwasserweg“, der eine aus der wolkenoptischen Dicke abgeleitete Größe ist). Diese Genauigkeitsangaben werden im äußeren QS-Metadatenatz der Primärkomponente vermerkt.

5.6.1

Für die Umrechnung von Wolkenoberflächen-Temperaturen in eine Höhe der Wolkenobergrenze werden Temperaturprofile benötigt. Diese Profile können aktuell oder vorhergesagt sein, oder aber einer Modell-Atmosphäre entsprechen. Was genutzt wurde, kann im QS-Metadatenlayer der Primärkomponente angegeben werden. Im QS-Metadatenatz der Prozessierungs-Kontext-Komponente kann dazu eine Zusammenfassung gegeben werden.

6.1.1

Die Plausibilität der Indikatoren für dünne Wolken kann in Verbindung mit der wolken-optischen Dicke und der Wolkentypisierung geprüft werden. Der Vermerk dazu wird pixelweise im QS-Metadatenlayer der Primärkomponente abgelegt.

6.2.1

Pixel, die in Bezug auf den Wolkentyp nicht klassifizierbar sind, sollen im QS-Metadatenlayer der Primärkomponente gekennzeichnet werden.

6.2.2

Zusätzlich zu 6.2.1 kann für die hinsichtlich des Wolkentyps klassifizierbaren Pixel die Klassifikationssicherheit angegeben werden.

6.3.1

Im inneren QS-Metadatenatz der Prozessierungs-Kontext-Komponente wird der Schwellenwert vermerkt, der für die Separation schneekontaminierter und schneefreier Pixel mit Hilfe des Schnee-Indikators verwendet wurde. Dazu kann eine Angabe zum Aussehen der Verteilung der Schnee-Indikator-Werte gemacht werden.

6.3.2

Der pixelweise vorliegende Schnee-Indikator selbst stellt ein Bestimmtheitsmaß für die Schneekontamination dar, deshalb ist er Teil des QS-Metadatenlayers der Primärkomponente.

6.4.1

Quicklooks sind z.B. für die Vorab-Auswahl von Datenprodukten nutzbar, etwa für das Auffinden von Szenen, in denen bestimmte Gebiete wolkenfrei sind. Daher ist es hilfreich, im inneren QS-Metadatenatz der Quicklook-Komponente zu vermerken, ob Quicklooks überhaupt (erfolgreich) generiert wurden.

7.1.1

Im äußeren QS-Metadatenatz der Prozessierungs-Kontext-Komponente kann vermerkt werden, ob während der Nachprozessierung für die Atmosphärenprodukte eine Extrapolation der wolken-optischen Dicke für teilbewölkte Pixel durchgeführt wurde. Dies macht natürlich nur Sinn, wenn dies nicht schon während der APOLLO-Prozessierung vorgenommen wurde (vgl. 5.5.3).

7.2.1

Die Plausibilität des Säulen-Wasserdampfgehalts (TWC) kann vom Postprozessor für die Atmosphärenprodukte im QS-Metadatenlayer der Primärkomponente gekennzeichnet werden.

7.3.1

Das vollständige Vorliegen aller (notwendigen) QS-Metadaten wird vom Postprozessor für die Atmosphärenprodukte festgestellt und im inneren Datensatz der QS-Metadaten-Komponente vermerkt.

7.3.2

Ebenfalls im inneren Datensatz der QS-Metadaten-Komponente kann eine Zusammenfassung aller erhobenen und in Metadaten transformierten QS-Information gegeben werden.

Die Abbildung 2-4 zeigt, dem Farbkomposit eines SEVIRI-Datenausschnittes gegenüber gestellt, einen Auszug aus dem QS-Metadaten-Layer der Primär-Komponente mit Aussagen zur Vollständigkeit der Daten und Bestimmtheit der Wolkentyp-Klassifikation. Rot bedeutet keine bzw.

fehlerhafte Daten, Weiß steht für Wolken, die typisiert werden konnten und Blau zeigt die als wolkenkontaminiert erkannten Pixel an, für die keine Wolkentyp-Klassifikation möglich war.

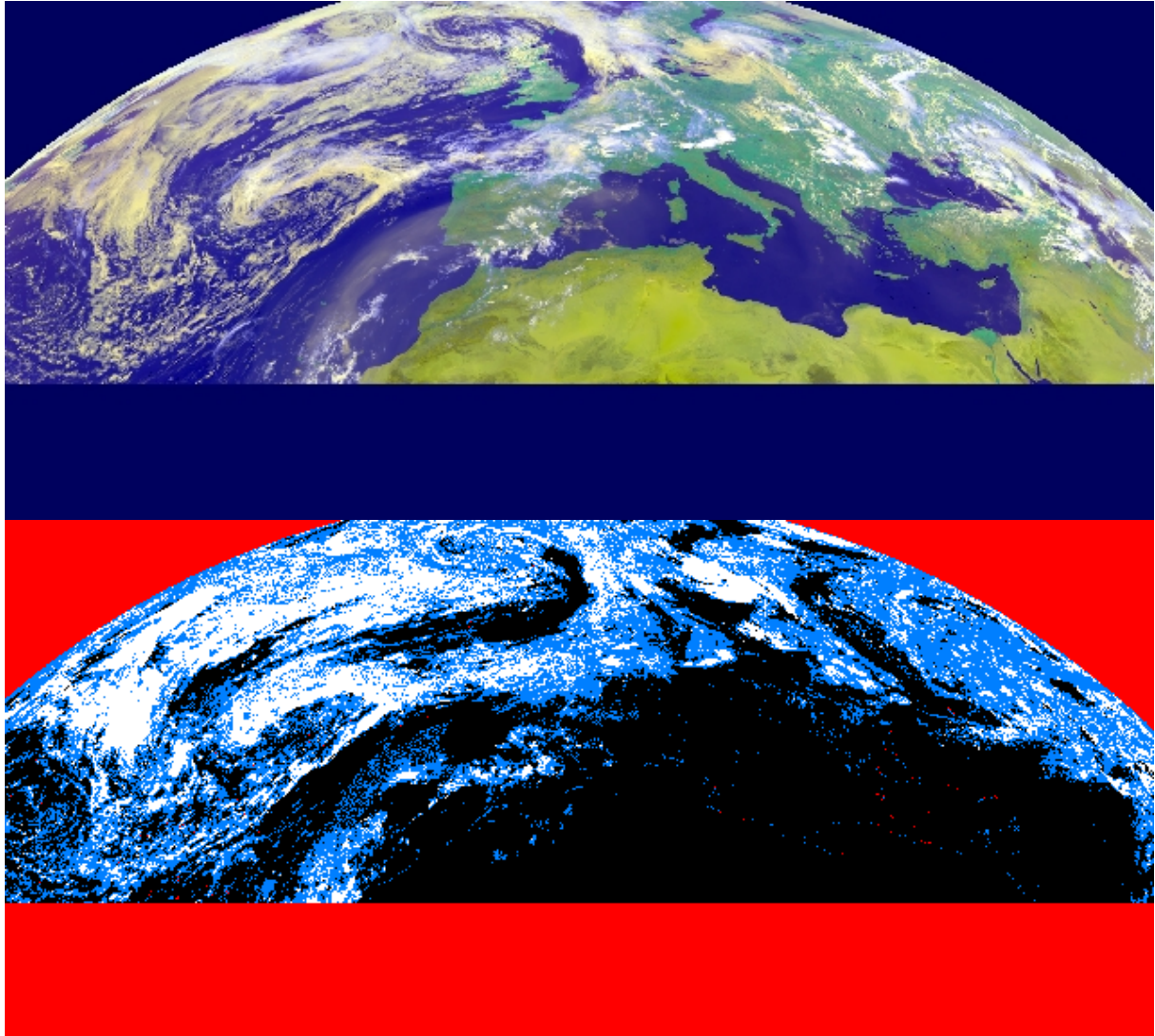


Abbildung 2-4: Beispiel für einen Auszug aus [^]QP zur Vollständigkeit und Bestimmtheit

3. Transformation der QS-Metadaten in Nutzerinformation

Es ist offensichtlich, dass der Nutzer der endgültigen Datenprodukte die QS-Informationen und QS-Metadaten in der Vielfalt und Form, so wie sie der Provider während der Produkt-Prozessierung und Produkt-Zusammenstellung generiert, nicht unmittelbar anwenden kann. Der Nutzer hat, zumal wenn es sich um ein kommerzielles Unternehmen handelt, den Wunsch, seinen Kunden übersichtliche und leicht verständliche Informationen über die Qualität der Datenprodukte zu geben. Deshalb ist es notwendig, dass der Provider, teils in Absprache mit dem Nutzer, die QS-Metadateninhalte bündelt (dieser Abschnitt) und die Belastbarkeit der daraus resultierenden Aussagen klar stellt (siehe Abschnitt 4).

Die Bündelung der QS-Metadateninhalte umfasst die Berücksichtigung mindestens der notwendigen (plus evtl. einiger optionalen) QS-Metadaten und möglichst aller Gruppen von vorliegenden Metadaten-Informationen und -Klassen. Dabei möchte der Nutzer stets drei grundsätzliche Dinge wissen, nämlich

- wie genau sind die gelieferten Datenprodukte,
- sind die Datenprodukte vollständig, so dass sie weiter prozessiert werden können und
- welche Grenzen haben die Qualitätsaussagen.

Die nutzergerechte Transformation der QS-Metadaten-Information sollte also so erfolgen, dass eine Übersicht über die Genauigkeit und die Vollständigkeit des Datenproduktes entsteht. Dabei muss noch berücksichtigt werden, dass sich die Genauigkeit eines Datenproduktes in der Regel nicht komplett analytisch-physikalisch angeben lässt. Meist ist ein Verfahren nur in Teilen, also in Bezug auf einige der in der Methode genutzten physikalischen Größen über eine Validierung oder validierungs-ähnliche Genauigkeitsabschätzung abgesichert. Über weite Teile eines Verfahrens und viele der verwendeten Zwischengrößen lassen sich dagegen nur plausible Genauigkeits-Annahmen machen. Es ist daher notwendig, gegenüber dem Nutzer die Aussagen über die Produktgenauigkeit aufzuteilen in einen analytisch-physikalisch gesicherten Teil und in einen empirischen, evtl. sogar spekulativen Teil. Es ergeben sich also drei Zielbereiche bei der Transformation der QS-Metadaten in Nutzerinformationen, nämlich die

- Produkt-Vollständigkeit,
- die analytisch-physikalische Produkt-Genauigkeit und
- die empirische Produkt-Genauigkeit.

Aussagen über Risiken und Grenzen der Nutzung der Qualitätsinformation sind genereller Natur und beziehen sich jeweils auf Produktgruppen mit kompatibler QS-Metadatenstruktur, die mit vergleichbaren Methoden generiert wurden. Sie werden daher in Abschnitt 4 dargelegt.

Mit Hilfe der im Abschnitt 2 beispielhaft entwickelten QS-Metadaten für den DFD-Teil der HELIOSAT-Prozesskette wird in **Tabelle 3-1** gezeigt, wie daraus die QS-Informationen für den Nutzer in den drei Zielbereichen zusammengestellt werden. Im grün unterlegten Teil der Tabelle sind die notwendigen QS-Metadaten aufgelistet, d.h. diejenigen, die in die Informations-Zusammenfassung einfließen müssen. Im blau unterlegten Tabellenteil sind dann noch die optionalen QS-Metadaten genannt. Schließlich werden im gelb unterlegten Tabellenteil noch die QS-Metadaten eingeordnet, die sich nicht direkt den (vom Nutzer erwarteten) Klassen „Vollständigkeit“ und „Genauigkeit“ zuordnen lassen, aber in der jeweiligen Spalte eine Ergänzung oder Präzisierung darstellen. Im gelben Teil wird zwischen notwendigen und optionalen QS-Metadaten unterschieden durch Kennzeichnung der optionalen mit einem „(o)“.

Vollständigkeit	analytische Genauigkeit	empirische Genauigkeit
DQ_CompletenessScenesLv15	DQ_ThematicAccuracyOzone	DQ_ThematicAccuracyScenesCal
DQ_CompletenessScenesBRDF	DQ_ThematicAccuracyAerosol	DQ_ThematicAccuracyApolloOdp
DQ_CompletenessScenesLSM		
DQ_CompletenessScenesGeo		
DQ_CompletenessScenesSubset		
DQ_CompletenessApolloAreas		
DQ_CompletenessApolloAlgorithms		
DQ_CompletenessAtmoProdMDSet		
DQ_CompletenessApolloOdp		DQ_ThematicAccuracyApolloGeo
DQ_CompletenessApolloQL		DQ_ThematicAccuracyApolloBRDF
DQ_CompletenessAtmoProdOdp		DQ_PositionalAccuracyApolloBRDF
		DQ_PositionalAccuracyApolloOdp
	DQ_TemporalValidityOzone	DQ_PlausibilityApolloGeo
	DQ_OzoneDefault (o)	DQ_PlausibilityApolloBRDF
	DQ_TemporalValidityAerosol	DQ_PlausibilityApolloAreas
	DQ_AerosolDefault (o)	DQ_PositionalValidityApolloGlint (o)
		DQ_PositionalValidityApolloCoast (o)
		DQ_ApolloDefaultAlgorithms (o)
		DQ_CertaintyApolloCloudLevel (o)
		DQ_PlausibilityApolloOdp
		DQ_ApolloDefaultCth (o)
		DQ_PlausibilityApolloThinIndex (o)
		DQ_CertaintyApolloCloudType
		DQ_CertaintyApolloCloudClass (o)
		DQ_CertaintyApolloSnowIndexThr
		DQ_CertaintyApolloSnowIndex
		DQ_PlausibilityAtmoProdTWC (o)
		DQ_CorrectnessAtmoProdMDSet

Tabelle 3-1: Transformationstabelle QS-Metadaten zu Nutzerinformation

Um ein für den Nutzer brauchbares Datenprodukt zur Verfügung zu stellen, werden also für den DFD-Teil der HELIOSAT-Prozesskette mindestens die im grünen Tabellenbereich genannten QS-Metadaten benötigt, ergänzt durch die im gelben Tabellenbereich genannten notwendigen QS-Metadaten (ohne „(o)“). Dabei sichern die QS-Metadaten der Klassen Gültigkeit (validity), Plausibilität (plausibility) und Bestimmtheit (certainty), dass die Genauigkeitsangaben ohne zusätzliche, nachträgliche Einschränkungen übernommen werden können. Die Einzelgenauigkeiten der Ausgangsdaten, physikalischen Zwischengrößen und Zusatzdaten müssen nach den Regeln der Fehlerfortpflanzung zusammengefasst werden, wobei in der Regel angenommen werden kann, dass sie untereinander unabhängig sind.

Die im blauen und gelben Tabellenteil genannten optionalen QS-Metadaten bieten eine Ergänzung hinsichtlich inhaltlicher Details oder Details einiger Zusatzdaten.

4. Risiken und Grenzen der QS mit Metadaten

Die Annahme, vom Provider mitgelieferte QS-Metadaten könnten ein Datenprodukt dem Nutzer gegenüber hinsichtlich dessen Qualitätskategorien „Vollständigkeit“ und „Güte“ ohne zusätzliche Kommunikation sofort und hinreichend beschreiben, ist falsch. Grundsätzlich sind zwischen Provider und Nutzer Absprachen notwendig, die am Anfang, wenn Produkte definiert werden, intensiver sein müssen als später in der Routinephase. Die Inhalte dieser Absprachen sind die Folgenden:

- Der Provider definiert den Produktinhalt nach den Anforderungen des Nutzers.
- Der Provider definiert eine sinnvolle Produktstruktur einschließlich (QS-)Metadaten, die sich an den Erfordernissen der Inhaltsstruktur, der Archivierung und des Transfers orientiert und dokumentiert dies dem Nutzer gegenüber.
- Der Nutzer informiert den Provider darüber, mit welchen Verfahren die Datenprodukte bei ihm weiterprozessiert werden und welche Produkteigenschaften vor allem hinsichtlich der Qualität hierfür notwendig sind.
- Dabei genügt es nicht, wenn sich Provider und Nutzer über den Inhalt und die Struktur der Datenprodukte absprechen. Vielmehr müssen auch konkrete Schwellenkriterien vereinbart werden, nach denen bei der Produkt-Generierung die zu prüfenden Qualitätsparameter abgefragt und beurteilt werden.
- Der Provider erläutert dem Nutzer, ob Qualitätsaussagen auf gesicherten physikalischen Analysen beruhen oder ob es sich um Abschätzungen und daraus resultierend bloße Empfehlungen handelt.
- Der Nutzer informiert den Provider darüber, welche Qualitätserwartungen seine Kunden haben und ob unerfüllte Qualitätserwartungen der Kunden oder nicht genügend belastbare Qualitätsaussagen juristische Konsequenzen haben können und wenn ja, welche.

Ein großes Risiko, dass QS-Metadaten ihren Nutzen verfehlen, besteht in einem Ausbleiben oder Vernachlässigen dieser Kommunikation.

Aus Tabelle 3-1 folgt nun sofort, dass zumindest für den DFD-Teil der HELIOSAT-Prozesskette die QS-Informationen, die echte physikalische Genauigkeitsangaben erlauben, gegenüber rein empirischen QS-Abschätzungen in der Minderzahl sind. Das bedeutet, dass sich für die Produkte MSG.SEVIRI.CLOUDS und MSG.SEVIRI.TWC nur wenig quantitative Genauigkeitsangaben machen lassen. Dagegen gibt es viele plausible empirische Qualitätskriterien, die der vom Provider in langer Zeit erworbenen Erfahrung bei der Generierung energiemeteorologischer Datenprodukte entstammen. Dieses Unverhältnis lässt sich ein wenig dadurch ausgleichen, dass möglichst alle QS-Metadaten, auch die optionalen, vom Provider zum Nutzer transferiert werden und dem Nutzer Werkzeuge in die Hand gegeben werden, wie alle Details dieser Informationen gegenüber dem Kunden zu brauchbaren QS-Aussagen weiterverdichtet werden können. Hierfür lassen sich für die QS-Metadaten-Informationen der sieben relevanten Qualitätsklassen folgende einfache Regeln stichwortartig aufstellen, um das Risiko der Fehlinterpretation der Metadaten und damit des ganzen Produktes zu reduzieren (vgl. auch mit Tabelle 2-6 bis Tabelle 2-12):

Vollständigkeit/Verfügbarkeit:

- Einordnung in einschränkende und kritische Vollständigkeitsdefizite durchführen
- Durchziehen der einschränkenden Vollständigkeitsdefizite bei der weiteren Prozessierung und in Zeitreihen
- Das Fehlen von notwendigen QS-Metadaten sollte definitionsgemäß immer kritisch sein

Konsistenz:

- Niemals Prozessor-Versionen mischen (Versionstreue)
- Quellentreue bei Eingangs- und Zusatzdaten
- Regelmäßige, ausreichend häufige Aktualisierung von Eingangs- und Zusatzdaten
- Stets die Repräsentativität von Interpretationen abfragen, d.h. möglichst breite Datengrundlage anstreben

Genauigkeit:

- Stets klare Trennung zwischen analytisch-physikalischen und empirischen Genauigkeitsangaben
- Unterscheidung zwischen Gesamt- und Einzelwertgenauigkeit

Gültigkeit:

- Vermeiden von pauschaler Nutzung aller Daten zugunsten konsequenter Beachtung von Geltungsbedingungen
- Erzeugung von Gültigkeitstabellen zum Einordnen von Einzelwerten in einen passenden Geltungsraum (wann, wo und unter welchen Voraus- und Randbedingungen)

Plausibilität:

- Verfolgen einer Bottom-up-Strategie, d.h. ein aus einem unplausiblen Zwischenergebnis erzeugtes Ergebnis ist ebenfalls als unplausibel anzusehen.
- Anwendung raum-zeitlich arbeitender Filter, um unplausible Werte durch raum-zeitlich kompatible zu ersetzen (Ausreißerfilter)

Überprüftheit:

- Sicherstellung der Überprüftheit auch der Eingangs- und Zusatzdaten auf jeder Zwischenstufe der Prozessierung
- Sicherstellung der Abprüfung aller notwendigen QS-Kriterien durch Protokollierung
- Lückenlose Protokollierung des Prozessierungsverlaufs inkl. Fehlerdiagnostik

Bestimmtheit:

- Diskussion alternativer Interpretationen von Ergebnissen
- Durchreichen von Unbestimmtheiten durch den Rest der Prozessierungskette

Diese Auflistung kann in anderen multilateralen Prozessierungs-Szenarien weitere Empfehlungen zur Risiko-Minimierung enthalten. In jedem Fall sollten Postprozessoren an verschiedenen Stellen des Prozessierungsverlaufes eingebaut werden, die durch Kombination von verfügbaren QS-Metadaten-Inhalten das Risiko von Fehlinterpretationen oder falschen Qualitätserwartungen reduzieren.

5. Zusammenfassung

Im Kontext der Erstellung und Nutzung energiemeteorologischer Datenprodukte bezweckt dieses Dokument

- die Klarstellung eines Begriffs von Metadaten zur Qualitätssicherung,
- die Abgrenzung solcher Metadaten gegenüber anderen Daten,
- strukturelle und inhaltliche Definitionen,
- die Identifikation methodischer Notwendigkeiten und methodischer Empfehlungen und
- die Verdeutlichung von Risiken und Grenzen im Umgang mit solchen Metadaten.

Obwohl dieses Dokument speziell anhand der und aufbauend auf die Produktstruktur des DFD die Erzeugung der QS-Metadaten beschreibt, haben die im Folgenden kurz dargestellten Aussagen generelle Gültigkeit für energiemeteorologische Datenprodukte.

Erstellte Definitionen:

Zweck und Abgrenzung gegenüber anderen Metadaten:

Außer zur Qualitätssicherung gibt es natürlich andere Metadaten, die der Beschreibung sowie räumlichen und zeitlichen Orientierbarkeit sowohl aus Provider- als auch aus Nutzersicht dienen.

Abgrenzung gegen Zusatzdaten:

Zusatzdaten sind alle Daten, die zusätzlich zur Initiierung, Erzeugung oder Verteilung von Datenprodukten benötigt werden, ohne diese weiter zu beschreiben. Die Qualität von Zusatzdaten kann dennoch Auswirkungen auf die Qualität des Datenproduktes haben. Beispiel hierfür sind etwa die Kalibrationskoeffizienten.

Einteilung in innere und äußere QS-Metadaten, je aus Nutzer- und Providersicht:

Sind QS-Informationen a priori bekannt oder lassen sie sich aus Rand- und Startbedingungen bei der Produkterzeugung ableiten, werden die daraus generierten QS-Metadaten als äußere QS-Metadaten bezeichnet. Dem gegenüber führen QS-Informationen, die erst im Verlauf der Prozessierung bekannt werden bzw. ermittelbar sind, zu inneren QS-Metadaten.

Erstellung einer Produktstruktur auf Basis der DFD-DIMS-Produktstruktur:

Die bereits im DFD-DIMS-Archiv- und Prozessierungsumfeld existierende Struktur von Datenprodukten wird um die in diesem Dokument aufgeführten QS-Metadaten-Strukturen erweitert.

Identifizierung von sieben erforderlichen Qualitätsklassen mit inhaltlicher Beschreibung, davon 4 Qualitätsklassen als Standard beschrieben in ISO19115 (Completeness, Accuracy, Validity, Correctness):

Die beim Provider anfallenden und vom Nutzer angeforderten QS-Informationen lassen sich in sieben Qualitätsklassen einteilen. Es stellt sich heraus, dass vier dieser Klassen sich in der Norm ISO19115 wiederfinden.

Konkrete Definition aller für den DFD-Teil der HELIOSAT-Prozesskette möglichen QS-Metadaten (d.h. Namens- und Inhaltsdefinition, wo möglich angelehnt an ISO19115):

Bei der hier durchgeführten Definition von QS-Metadaten beschränken wir uns auf den DFD-Teil der HELIOSAT-Prozessierungskette, da wir hierfür der Provider und damit der Experte sind. Die anderen Teile können analog von den zuständigen Providern bearbeitet werden.

Vorgehensweise zur QS-Metadatengewinnung:

Erhebung der gesamten verfügbaren QS-Informationen unter Berücksichtigung der Nutzeranforderungen (konkret gezeigt am DFD-Teil der HELIOSAT-Prozesskette):

Der erste Schritt ist, festzustellen, welche QS-Informationen im Voraus, im Umfeld und während der Prozessierung ableitbar sind.

Beschreibung der Quantifizierung der QS-Informationen allgemein und konkret für den DFD-Teil der HELIOSAT-Prozesskette:

Die QS-Informationen sollten quantifizierbar sein oder es muss ein Verfahren zur Quantifizierung beschrieben werden.

Abbildung der verfügbaren QS-Informationen auf definierte QS-Metadaten-Einheiten, wo möglich angelehnt an ISO19115:

Die QS-Metadaten-Einheiten, die die QS-Informationen beinhalten und transportieren sollen, werden mit möglichst weitgehender Anlehnung an ISO19115 definiert. Die QS-Informationen werden zuvor den sieben Qualitätsklassen zugeordnet.

Festlegung der Priorität (notwendig oder optional):

Die Festlegung der Priorität der QS-Metadaten geschieht nach dem Prinzip der Dominanz des Notwendigen. D.h. QS-Metadaten sind nur dann als optional einzustufen, wenn sie sowohl für den Provider als auch für den Nutzer optional sind, d.h. für beide Parteien nicht unbedingt notwendig sind zur ausreichenden Beurteilung der Produktqualität.

Ableitung von Nutzerinformationen, Behandlung von Risiken:

Festlegung von nur drei QS-Informationsklassen für den Nutzer: Vollständigkeit, analytische Genauigkeit, empirische Genauigkeit:

Da der Nutzer eher eine komprimierte, übersichtlichere QS-Beurteilung des Datenproduktes wünscht, muss die Vielfalt der QS-Metadaten reduziert werden. Dies geschieht durch eine Abbildung aller QS-Metadaten der sieben Qualitätsklassen in die zwei für den Nutzer bestimmten QS-Aussageklassen „Vollständigkeit“ und „Genauigkeit“. Die Genauigkeitsaussagen müssen noch differenziert werden in Abschätzungen auf Grund empirischer Annahmen und in analytisch-physikalisch ermittelte Genauigkeiten, so dass dann effektiv drei Aussageklassen dem Nutzer vermittelt werden.

Verdichtung der QS-Metadaten-Informationen zu Aussagen in diesen drei Klassen auf der Basis von Vereinbarungen zwischen Provider und Nutzer (Bewertung der Gesamt-Vollständigkeit, quantitative Angaben zu Gesamtgenauigkeiten in Form von MEAN, RMSE, Konfidenzintervallen):

Für diese Verdichtung der QS-Metadateninformation sind Bedeutungs- und Bewertungs-Vereinbarungen zwischen Nutzer und Provider notwendig.

Inhaltliche Beschreibung der notwendigen Kommunikation zwischen Provider und Nutzer zur Risiko-Minimierung und Aufstellung von Regeln zur Vermeidung von Fehlinterpretationen und falschen Güte-Erwartungen bzgl. der QS-Nutzerinformationen:

Um das Risiko einer unzureichenden Nutzbarkeit der QS-Aussagen durch den Nutzer zu minimieren, sowie Fehl-, Falsch- und Überbewertungen der Aussagen zu vermeiden, bedarf es einiger weiterer Kommunikationsregeln zwischen Provider und Nutzer.

Referenzen

A. von Bargaen, T. Schröder, A. Doicu, K. Kretschel, C. Lerot, M. Van Roozendaal, A. Kokhanovsky, M. Vountas, H. Bovensmann, M. Hess, B. Aberle & F. Schreier:

“SCIAMACHY level 1b-2 data processing: Update of off-line data processor to version 3.0”, in: Proceedings of the Second Workshop on the Atmospheric Chemistry Validation of ENVISAT (ACVE-2), 3-7 May 2004, ESA ESRIN, Frascati, Italy, edited by: Danesy, D., ESA Publications Division, Noordwijk, Niederlande, SP-562, 2004

Kriebel, K. T., Saunders, R. W. and Gesell, G., 1989, „Optical properties of clouds derived from fully cloudy AVHRR pixels“. Beiträge zur Physik der Atmosphäre, Vol. 62, No. 3, 165-171

Kriebel, K. T., Gesell, G., Kästner, M. and Mannstein, H., „The cloud analysis tool APOLLO: improvements and validations“, Int. J. Remote Sensing, 2003, Vol. 24, No. 12, 2389-2408

Lambert, J.-C., Alaart, M., Andersen, S. B., Blumenstock, T., Bodeker, G., Brinksma, E., Cambridge, C., de Maziere, M., Demoulin, P., Gerard, P., Gil, M., Goutail, F., Granville, J., Ionov, D. V., KyrÅ“o, E., Navarro-Comas, M., PETERS, A., Pommereau, J.-P., Richter, A., Roscoe, H. K., Schets, H., Shanklin, J. D., Suortti, T., Sussmann, R., Van Roozendaal, M., Varostos, C., Wagner, T., Wood, S., and Yela, M.: “First ground-based validation of SCIAMACHY v5.01 ozone column”, in: Proceedings of the Second Workshop on the Atmospheric Chemistry Validation of ENVISAT (ACVE-2), 3-7 May 2004, ESA ESRIN, Frascati, Italy, edited by: Danesy, D., ESA Publications Division, Noordwijk, Niederlande, SP-562, 2004.