Ground-based Network Capacity Analysis for CO2M Cal/Val

Doc.No. : EUM/COPER-CO2M/DOC/21/1215886
Issue : v2
Date : 19 February 2021
WBS/DBS :

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

© EUMETSAT
The copyright of this document is the property of EUMETSAT.

Document Change Record

Issue / Revision	***Date***	***DCR No***	***Changed Pages / Paragraphs***
V0	2020-02-27	N/A	Provided as initial revision.
V1	2020-04-20		Agreed document structure by EUM and LMU.
V1.1	2020-09-17		Draft discussed at technical meeting on 2020-09-17.
V1.2	2020-11-05		Draft prepared for first intermediate presentation on 2020-11-05.
V1.3	2021-01-29		Draft prepared for final review on 2021-02-01.
V1.4	2021-02-19		Final version submitted to EUMETSAT
V2	2021-02-22		Final version accepted by EUMETSAT

Table of Contents

1 INTRODUCTION

1.1 Purpose and Scope

This document is the ground-based network capacity analysis for operational (vicarious) calibration, validation, and monitoring (Cal/Val) of the High-Priority Candidate Copernicus Mission (HPCM) for the monitoring of anthropogenic CO_2 emissions (CO2M) [EC-GR]. The analysis, findings, and conclusions drawn by this document will contribute to the Cal/Val plan of CO2M and will support the planned EUMETSAT study on the "Definition of requirements for continuous Cal/Val and monitoring of level 1 and Level 2 products" [SOW].

The purpose of this work is

1. to identify existent components and infrastructure of the European component of the Total Carbon Column Observing Network (TCCON) as well as the COllaborative Carbon Column Observing Network (COCCON), which are fit-for-purpose to support the operational Cal/Val of CO2M during commissioning, but in particular during the continuous operations phase (Phase E), and
2. to make proposals for required improvements of data-handling, transmission, processing, (software) maintenance, and operations of network data to facilitate the needs of operational Cal/Val for CO2M.

Part 1 of this work shall be carried out based on existing analysis of network capacity [EGHGCV], and by additional gathering of information, as well as by making proposals for additional network analysis to be carried out. Part 2 shall be carried out by gathering the required information from all station providers, the existing protocols and data processing and handling mechanisms and in close cooperation with the EUMETSAT central facility (CF) developments of Multi-Mission Element (MME) Cal/Val functions for CO2M.

Note that the required maintenance of the network(s) station infrastructure, the operations of the station instrumentation, as well as the conduction of the measurements are out-of-scope for this work.

1.2 Relation to other documents

This document is mainly based on existing TCCON and COCCON network analysis [EGHGCV], [EC-GR] as well as available documentation for TCCON [GGG2014].

1.3 Applicable Documents

Acronym	Title	Reference
[SOW]	Scientific Service Support for the Definition of Requirements for an Integrated Function for CalVal and Monitoring of GHG Products	EUM/TSS/SOW/19/1102409

1.4 Reference Documents

Acronym	Authors	Title	Reference
[BAB2020]	A. Baben-hauserheide et al.	Net CO2 fossil fuel emissions of Tokyo estimated directly from measurements of the Tsukuba TCCON site and radiosondes	Atmos. Meas. Tech., 13, 2697–2710, 2020.

			doi:10.5194/amt-13-2697-2020
[BEL2017]	D. A. Belikov et al.	Study of the footprints of short-term variation in XCO_2 observed by TCCON sites using NIES and FLEXPART atmospheric transport models	Atmos. Chem. Phys., 17, 143–157, 2017 doi:10.5194/acp-17-143-2017
[DIET2020]	F. Dietrich at al.	Munich permanent urban greenhouse gas column observing network	Atmos. Meas. Tech. Discuss. [preprint], 2020. doi:10.5194/amt-2020-300
[EGHGCV]	D. G. Feist et al.	Report on European greenhouse gas column Cal/Val network sustainability	EEA/IDM/15/026/LOT1
[FREY2019]	M. Frey et al.	Building the COllaborative Carbon Column Observing Network (COCCON): long-term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer	Atmos. Meas. Tech., 12, 1513–1530, 2019. doi:10.5194/amt-12-1513-2019
[GISI2012]	M. Gisi et al.	XCO_2-measurements with a tabletop FTS using solar absorption spectroscopy	Atmos. Meas. Tech., 5, 2969-2980, 2012. doi:10.5194/amt-5-2969-2012
[HASE2013]	F. Hase et al.	Calibration of sealed HCl cells used for TCCON instrumental line shape monitoring	Atmos. Meas. Tech., 6, 3527–3537, 2013. doi:10.5194/amt-6-3527-2013
[HASE2015]	F. Hase et al.	Application of portable FTIR spectrometers for detecting greenhouse gas emissions of the major city Berlin	Atmos. Meas. Tech., 8, 3059–3068, 2015. doi:10.5194/amt-8-3059-2015
[HASE2016]	F. Hase et al.	Addition of a channel for XCO observations to a portable FTIR spectrometer for greenhouse gas measurements	Atmos. Meas. Tech., 9, 2303-2313, 2016 doi:10.5194/amt-9-2303-2016
[HED2017]	J. K. Hedelius et al.	Intercomparability of XCO_2 and XCH_4 from the United States TCCON sites	Atmos. Meas. Tech., 10, 1481–1493, 2017. doi:10.5194/amt-10-1481-2017
[KOR2014]	E. A. Kort et al.	Four corners: The largest US methane anomaly viewed from space	Geophys. Res. Lett.,41,6898–6903, 2014. doi:10.1002/2014GL061503
[SHA2020]	M. K. Sha et al.	Intercomparison of low- and high-resolution infrared spectrometers for ground-based	Atmos. Meas. Tech., 13, 4791–4839, 2020.

		solar remote sensing measurements of total column concentrations of CO_2, CH_4, and CO	doi:10.5194/amt-13-4791-2020
[EC-GR]	B. Pinty et al.	An operational anthropogenic CO_2 Emissions Monitoring and verification Support Capacity (green report)	EUR 29817 EN doi:10.2760/182790
[ROC2020]	S. Roche et al.	Retrieval of atmospheric CO2 vertical profiles from ground-based near-infrared spectra	Atmos. Meas. Tech. Discuss. [preprint], 2020. doi:10.5194/amt-2020-429
[SIM2002]	A. Simon et al.	Data acquisition and interferogram data treatment in FT-IR spectrometers	Vibrat. Spectrosc., 29, 97-101, 2002. doi:10.1016/S0924-2031(01)00191-6
[TOON2016]	G. C. Toon et al.	HITRAN spectroscopy evaluation using solar occultation FTIR spectra	J. Quant. Spectrosc. Radiat. Transf., 182, 324-336, 2016. doi:10.1016/j.jqsrt.2016.05.021
[TU2020]	Q. Tu et al.	Atmospheric CO_2 and CH_4 abundances on regional scales in boreal areas using CAMS reanalysis, COCCON spectrometers and Sentinel-5 Precursor satellite observations	Atmos. Meas. Tech., 13, 4751–4771, 2020 doi:10.5194/amt-13-4751-2020
[GGG2014]	D. Wunch et al.	The Total Carbon Column Observing Network's GGG2014 Data Version	doi:10.14291/tccon.ggg2014.documentation.R0/1221662
[WUN2019]	D. Wunch et al.	Emissions of methane in Europe inferred by total column measurements	Atmos. Chem. Phys., 19, 3963–3980, 2019. doi:10.5194/acp-19-3963-2019
[ESA-EOPG]		Improved Spectroscopy for Carbon Dioxide, Oxygen, and Water Vapour Satellite Measurements	ESA-EOPG-COP-SOW-1

1.5 Acronyms

Acronym	*Definition*
ADC	Analog Digital Converter
AEMET	Agencia Estatal de Meteorología
AK	Averaging Kernel
AOD	Aerosol Optical Depth
AOI	Area Of Interest

Acronym	***Definition***
ATC	Air Traffic Control
APID	Application Process Identifier
AUTh	Aristotle University of Thessaloniki
BIRA	Belgian Institute for Space Aeronomy
BRDF	Bidirectional Reflectance Distribution Function
CF	Central Facility
CLIM	CLoud IMager
CO2M	CO2 mission
CO2IS	CO2 instrument on the CO2 mission platform
COCCON	Collaborative Carbon Column Observing Network
DUP	Data Use Policy
DLR-IPA	Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre
DOF	Degrees Of Freedom
EARS	EUMETSAT Advanced Retransmission Service
EDGAR	Emission Database for Global Atmospheric Research
EVDC	ESA Atmospheric Validation Data Centre
FMI	Finnish Meteorological Institute
FOV	Field Of View
FRM	Fiducial Reference Measurement
FT	Fourier Transformation
FTIR	Fourier Transform Infrared
FTS	Fourier-Transform Spectrometer
HKTM	House-Keeping Telemetry
HQ	Headquarters
ICD	Interface Control Document
ILS	Instrument line shape
IRF	Institutet för Rymdfysik

Acronym	Definition
IRD	Interface Requirements Document
JPL	Jet Propulsion Laboratory
KIT	Karlsruhe Institute of Technology
LERMA	Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique et Atmosphères
LMU	Ludwig-Maximilians-Universität München
MAP	Multi-Angle Polarimeter
MCC	Mission Control Centre
NDA	Non-Disclosure Agreement
NetCDF	Network Common Data Format
NIR	Near Infra-Red
NMI	National Metrology Institute
NO2IS	NO2 instrument on the CO2 mission platform
NWC	Nowcasting
NWP	Numerical Weather Prediction
NRT	Near Real Time
OPD	Optical Path Difference (of an FTS)
OSSE	Observing System Simulation Experiment
PBL	Planetary Boundary Layer
PFS	Product Format Specification
PM	Person Month
QA	Quality Assurance
QC	Quality Control
RINGO	Readiness of ICOS for Necessities of integrated Global Observations
RT	Real time
SNR	Signal to Noise Ratio
SWR	Short-Wave infra-Red
SZA	Solar Zenith Angle

Acronym	*Definition*
TCCON	Total Carbon Column Observing Network
TM	Telemetry
TUM	Technische Universität München
UoA	University if Alaska
UOL	University of Leicester
UTC	Universal Time Coordinated
VIS	VISible
VSAT	Very Small Aperture Terminal
WMO	World Meteorological Organization

1.6 Document structure

Section Number	*Title*	*Content*
1	Introduction	Background and scope of document.
2		Overview of the ground based networks TCCON and COCCON for satellite GHG observation cal/val.
3		Recommendations for addressing gaps between capabilities of existing ground based cal/val networks and CO2M requirements.
4		Detailed description of data processing and handling in TCCON.
5		Definition of timeliness requirements and proposed network data processing options for future operational CO2M cal/val system.
6		Implementation and maintenance of procedures in order to reduce the inter-station biases and improve the overall network data quality and bias consistency.
7		Existing product quality assurance measures in TCCON and COCCON.
8		Recommendations for addressing the gaps between the current state of the ground based networks and the requirements for an operational CO2M cal/val system.
Appendix A		Background information on TCCON and COCCON data.

2 OVERVIEW OF GLOBAL COLUMN GHG CAL/VAL NETWORKS

2.1 TheTotal Carbon Column Observing Network (TCCON)

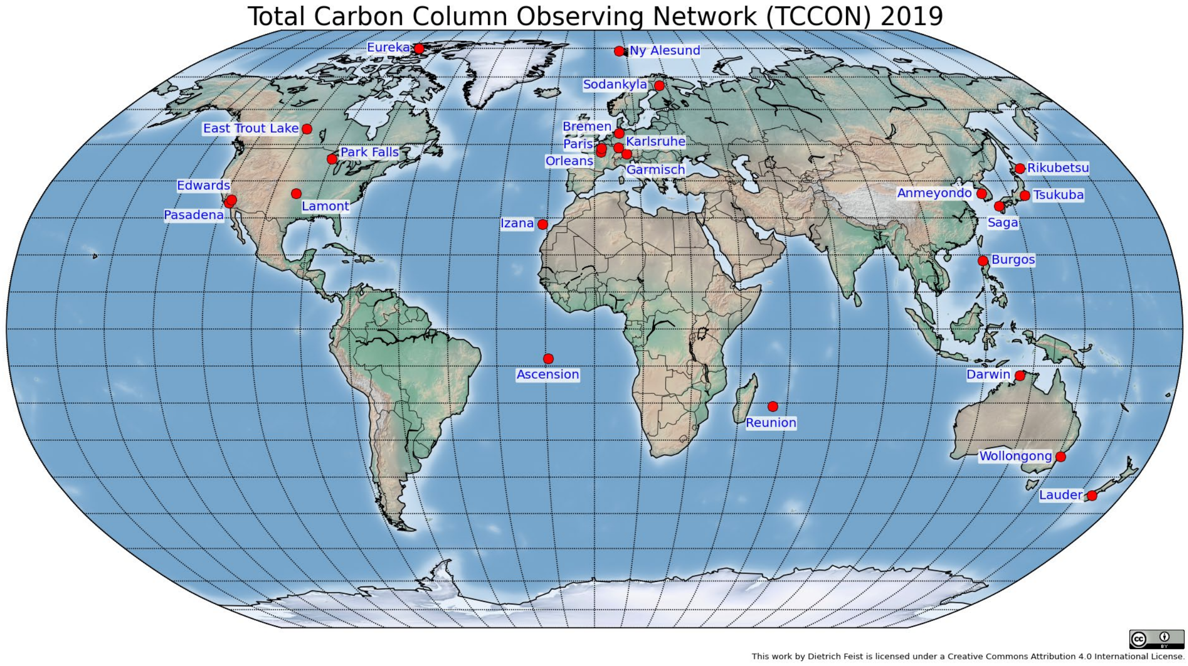

Figure 1: Map of the Total Carbon Column Observing Network (TCCON). There were 27 active TCCON stations in 2019.

2.1.1 The global TCCON network

The cal/val reference for all satellite greenhouse gas observations is the Total Carbon Column Observing Network (TCCON): a global network of ground-based stations that use remote sensing techniques to observe column-averaged greenhouse gas mole fractions. The network is tied against the WMO greenhouse gas standards and provides a link between satellite observations and the ground-based ambient air measurement network.

TCCON was set up in 2004 with four stations supporting NASA's then planned Orbiting Carbon Observatory (OCO) mission. Since then, the TCCON network has grown to 27 stations in 14 countries (see Fig. 1). All TCCON stations use the same type of instrument, a Fourier-Transform Infrared spectrometer type IFS125HR manufactured by Bruker Optics, Ettlingen, Germany. They also share the common retrieval code GGG together with a stringent quality control to give the most precise column-average measurements of CO_2, CH_4, N_2O, HF, CO, H_2O, and HDO. All current (GOSAT, GOSAT-2, OCO-2, OCO-3, TanSat, Sentinel-5P) and planned (MicroCarb, MethaneSAT, Sentinel5, GEOCARB, MERLIN, TanSat-2, GOSAT-3, CO2M) satellite missions observing greenhouse gases rely on TCCON for their cal/val.

2.1.2 The European TCCON network component

The current European TCCON network consists of 12 stations located in 4 European countries on the continent or in the Mediterranean (Finland, France, Germany, Cyprus), 3 islands that are part of a European country (Reunion, Svalbard, Tenerife) and 1 UK overseas territory (Ascension). The respective host institutions are from 4 different countries (Belgium, Finland, France, Germany). Table 1 lists the stations and their respective home institutions.

Table 1: Home institutions and locations of the European TCCON stations.

Home institution of PI	Country (institution)	Station/Location	Country (station)	Established
University of Bremen	Germany	Bremen	Germany	2005
		Ny Ålesund, Svalbard	Norway	2005
		Orléans	France	2009
		Nicosia	Cyprus	2019
KIT	Germany	Garmisch	Germany	2007
		Izaña, Tenerife	Spain	2007
		Karlsruhe	Germany	2010
		Zugspitze	Germany	2015
FMI	Finland	Sodankylä	Finland	2009
BIRA	Belgium	Réunion Island	France	2011
LMU & DLR-IPA	Germany	Ascension Island	St. Helena, Ascension & Tristan da Cunha	2012
LERMA	France	Paris	France	2014

2.1.3 Network policy and organisation

TCCON does not have central funding and thus no formal central organization structure. The main body is the steering committee, where each PI has one vote per station that they are responsible for. The steering committee elects regional and topical chairs for a term of three years which handle the day-to-day affairs on behalf of the PIs. The current chairs are listed in Table 2.

Table 2: TCCON chairs serving from 2019 until 2022.

Role	Name	Institution	Country
TCCON Chair & Co-Chair for the Americas	Debra Wunch	University of Toronto	Canada
Deputy Chair for the Americas	Coleen Roehl	California Institute of Technology	USA
Co-Chair for Africa & Europe	Thorsten Warneke	University of Bremen	Germany
Deputy Chair for Africa & Europe	Dietrich Feist	Ludwig-Maximilians-Universität München & DLR Institut für Physik der Atmosphäre	Germany

Role	Name	Institution	Country
Co-Chair for Asia/Oceania	Nicholas Deutscher	University of Wollongong	Australia
Deputy Chair for Asia/Oceania	(currently vacant)		
Algorithm Co-Chair	Geoffrey Toon	Jet Propulsion Laboratory	USA
Algorithm Deputy Chair	Joshua Laughner	California Institute of Technology	USA

TCCON is a scientific network. Membership is based on accepting the TCCON charter, following the common procedures for data analysis and quality control and sharing their data. The official Level -2 data product is published through the TCCON data archive. The PIs can define an embargo period of up to 12 months after measurement. However, most stations release their data after 3-4 months to allow sufficient time to retrieve the data, process it, and perform quality controls. Level-0 and Level-1 products are not published. These levels are defined in Section **Error! Reference source not found.**.

Everyone is free to use the public TCCON data product internally for any purpose. However, the right to redistribute the original TCCON data or derived data products is not included. The TCCON DUP and licence define rules for co-authorship and proper acknowledgement when TCCON data are used in scientific publications. There are currently no special rules for non-scientific use of the public data, e.g. for operational satellite cal/val activities.

2.2 The COllaborative Carbon Column Observing Network (COCCON)

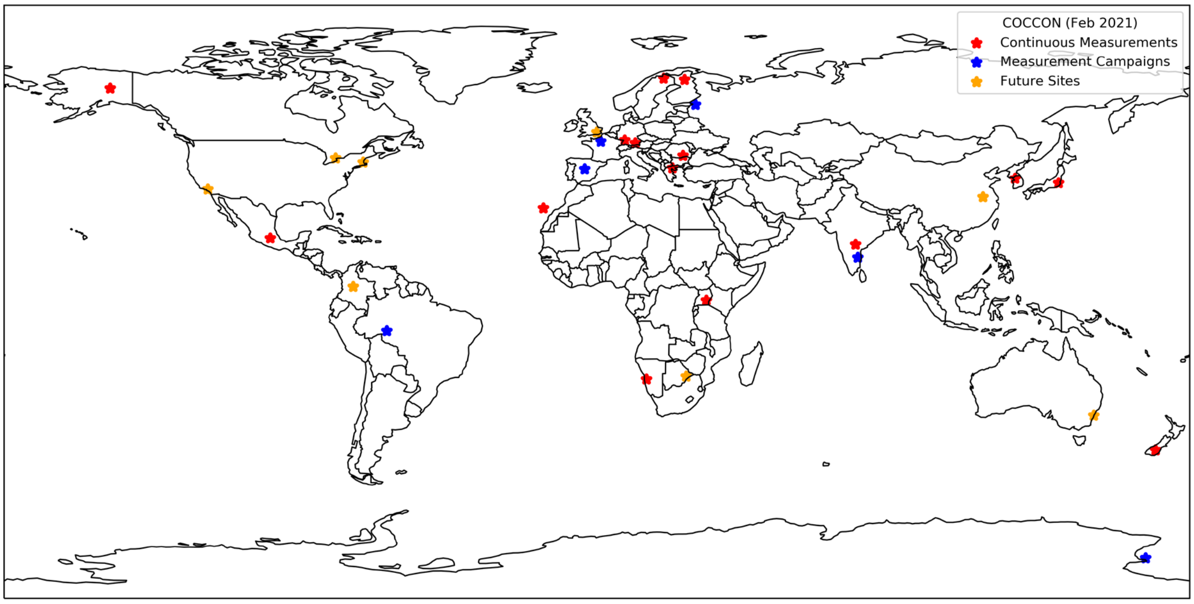

Figure 2: Map of the COllaborative Carbon Column Observing Network (COCCON). There were 15 active COCCON stations in 2020/21. Five stations are located close to TCCON stations. Seven multi-instrument campaigns have been conducted so far. Map provided by D. Dubravica, KIT.

2.2.1 Instrumental concept

COCCON was initiated in 2014. The network relies on the low-resolution Bruker EM27/SUN, a smaller, more mobile and also less expensive FTS [GISI2012]. In the standard configuration, this instrument is capable of observing XCO_2 and XCH_4. With an optional additional channel, XCO can also be retrieved [HASE2016].

Before delivery to the customer by the manufacturer Bruker, each EM27/SUN spectrometer is tested and calibrated by KIT with respect to the TCCON spectrometer operated in Karlsruhe and the reference EM27/SUN spectrometer operated continuously since 2014 collocated to the TCCON spectrometer. These calibrations are repeated when spectrometers are refurbished by Bruker or when required (typical repeating cycle: several years). The calibration has proved to be very stable over time and even during transport of the instrument. For long-term stationary operation, the instruments should be sent back to KIT for recalibration against the TCCON instrument every 1-2 years. With the presented long-term stability, the EM27/SUN qualifies as a useful supplement to the existing TCCON network. [FREY2019].

The mobility of the EM27/SUN instrument makes it ideal for observation campaigns. Multiple instrument setups can be set up for cross-track or along-track satellite validation. Multiple instruments can also be set up around emission hotspots to capture all incoming and outgoing airmasses. This way, the net CO_2 budget of a region can be derived if enough instruments can be set up and operated in a ring around it. The feasibility of this approach has been demonstrated on a campaign basis for the cities of Berlin, Paris, Tokyo, Madrid, Boston (permanent observatory, Harvard University), Munich (permanent observatory, TUM), and St. Petersburg.

2.2.2 Global and European COCCON network

Currently, COCCON consists of 15 stationary sites (see Fig. 2). Table 3 provides an overview of the stations operated by European institutions.

Table 3: Home institutions and locations of the European COCCON stations.

Home institution of PI	Country (institution)	Station/Location	Country (station)	Established
KIT	Germany	Karlsruhe	Germany	2012
		Gobabeb	Namibia	2015
TUM	Germany	Munich	Germany	2015
KIT & UoA	Germany/USA	Fairbanks, Alaska	USA	2016
KIT & NIWA	Germany/New Zealand/	Lauder	New Zealand	2016
FMI	Finland	Sodankylä	Finland	2017
KIT & IRF	Germany/Sweden	Kiruna	Sweden	2017
AEMET	Spain	Izana, Tenerife	Spain	2018
KIT & AUTh	Germany/Greece	Thessaloniki	Greece	2019
UOL	United Kingdom	Jinja	Uganda	2020
INOE	Romania	Magurele	Romania	2020

2.3 Network policy and organisation

A main concept of COCCON is that the individual instruments can also be set up and operated by groups that are not necessarily FTIR specialists. To achieve this goal, COCCON follows a centralized approach where KIT provides (or plans to provide) most of the essential services for the network:

- initial testing, optimization and calibration of new instruments before first deployment.
- routine checks and recalibration of instruments sent back to KIT.
- development of data processing tools an QA/QC procedures.
- central data processing as an optional service for PIs (planned).
- data archival and distribution to data users (planned).

Beyond that, there is currently no formal structure of the network. PIs may use some or all of the services provided by KIT but are not required to do so. Network development is guided by KIT with inputs from PIs and data users.

COCCON is supported by ESA in the framework of the projects COCCON-PROCEEDS, COCCON-PROCEEDS II, and FRM4GHG to develop an easy-to-handle preprocessing tool optimized for the EM27/SUN spectrometer and to demonstrate desired functionalities of a COCCON central facility. The PREPROCESS tool generates quality-checked spectra from raw interferograms, which are then forwarded to a central data analysis facility.

When finally implemented on an operational level, the central facility should remove the trace gas analysis from the instrument operators and ensure the consistency of the trace gas analysis chain for all participating sites. It will also serve as the primary contact and data distribution centre for the data users. Currently, the central processing and HDF file generation for distribution via EVDC is operated in a demonstration mode (contractual framework: COCCON-PROCEEDS II). The current contract supporting this activity will end in autumn 2021.

2.4 Data and software license

The code for COCCON data retrieval, PROFFAST, has been developed in the framework of an ESA project. Similar to the I2S module of GGG, there is a preprocessing step (software PREPROCESS) in PROFFAST which generates spectra from interferograms and is optimized for use on the EM27/SUN spectrometer. PROFFAST performs several specific QCs and the subsequent quantitative trace gas analysis. This second task is broken into two steps: the PCXS code tabulates the daily x-section tables from line lists and atmospheric conditions. In the final step, INVERS retrieves the trace gas amounts from the calibrated spectra, including the post-processing. PROFFAST has been tested extensively and has been identified as the recommended baseline code for EM27/SUN data processing [SHA2020].

PROFFAST is open source and free to use and distribute. The exact software license is currently being discussed. Most likely, the software will be made available to the EM27/SUN community through a Creative Commons licence: CC BY-SA (commercial use permitted) or CC BY-NC-SA (commercial use not permitted). There is no required return service for users of the code. In particular, use of the code does not imply a contribution to the COCCON network.

For the data produced and distributed within COCCON, the plan is to install a data use policy that is similar to the TCCON data use policy. This has been agreed with the TCCON chairs and has been accepted by ESA.

3 METHODS FOR IDENTIFICATION OF NEEDS FOR CO2M CAL/VAL

3.1 Station and network capability gap analysis

Even though providing cal/val reference for satellite GHG observations is a core task for both TCCON and COCCON, the networks were designed for the first and second generation of GHG-observing satellite instruments. Given enough time and resources, it is assumed that the networks can be improved and extended to meet the requirements of the CO2M mission.

To determine potential gaps between the capabilities of TCCON and COCCON and the requirements of the CO2M mission, two tasks need to be completed:

1. The existing capabilities of TCCON and COCCON with respect to CO2M cal/val need to be summarized and put in a form that can be used in observation simulations for the CO2M mission. It is beyond the scope of this report to already provide the results. It does, however, show which parameters are needed to describe these capabilities and how they can be determined.
2. The cal/val requirements of the CO2M mission have to be defined in a form that can be tested against the capabilities of the existing or future improved/extended networks. The suggestion would be to create an OSSE for CO2M that can simulate the effects of various cal/val networks and activities on a station-by station basis – or integrate this feature into an existing OSSE if one already exists. This step should be taken by experienced members of the satellite observation and inverse modelling community with support from the TCCON and COCCON community.

The gaps analysis could then be derived from the OSSE results and translated into concrete steps for the necessary adaption of the ground based networks. This could include upgrades of the instrumentation, improved procedures, data processing and QC as well recommendations for setting up additional stations or relocating existing ones.

A list of parameters that are considered relevant for the first task are listed in Table 4. They are grouped by two parameters: *priority* and *approach*. *Priority* describes how relevant the parameter is for the final gap analysis:

1. **Critical:** the parameter must be determined for a meaningful gap analysis.
2. **Important:** the parameter will considerably improve the quality of the analysis.
3. **Useful:** the parameter should be considered but may be replaced by a reasonable default value if it cannot be determined.

Approach describes the necessary actions to determine the parameter. Some parameters are already publicly known, some are known within the networks but are not public. Other parameters require further analysis or tools that do not exist yet. In particular, the following options appear several times:

- **Station survey:** the parameters are known by the individual PIs but need to be collected for the whole network. A general web survey for updating the metadata for the TCCON data set DOIs is already planned for TCCON in early 2021. This survey could be extended to include many of the station-dependent parameters that are relevant for CO2M. It could also be extended to COCCON.
- **Additional work:** the principal components are available but need some extra work that will likely require additional resources.
- **Dedicated study:** the parameter is not known and not straightforward to determine. It will likely require an effort of several PMs and the tools mentioned in the comment field.

Table 4: List of parameters describing station characteristics of the ground based TCCON and COCCON networks that are considered to be relevant for the CO2M cal/val. The TCCON and

COCCON columns indicate the relevance of this parameter for respective network: ✔→ relevant, (✔) → partially relevant, empty → not relevant.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Cloud cover frequency at CO2M resolution	Critical	✔	✔	Dedicated study	Can be derived from NWP model data or CO2M OSSE if available. Unclear if horizontal resolution would be sufficient.
Emission sources in station vicinity	Critical	✔	✔	Dedicated study	Source type, distance, direction and possibly typical wind direction would be useful. Could use inventories like EDGAR.
Projected CO2M overpass frequency	Critical	✔	✔	Dedicated study	Needs at least a CO2M orbit prediction tool. Optimal results could be derived from CO2M OSSE if that is available.
Surface albedo around station	Critical	✔	✔	Dedicated study	Only coarse data available so far. Need to define wavelength, distance around site and horizontal resolution. Could probably be derived from BRDF satellite product.
Additional CO_2 concentration measurements at surface or in PBL available	Important	✔	✔	Station survey	Could be surface or tower measurements, e.g. from a nearby ICOS station. Would greatly enhance the CO_2 sub layer potential of a station.
Aerosol measurement availability	Important	✔	✔	Station survey	Unclear how many stations have some kind of AOD measurement nearby.
Emission level from station	Important	✔	✔	Station survey	Need to define emission level. Suitable data source unclear. Some stations have ambient-air CO2 concentration measurements nearby.
NO2 levels at station	Important	✔	✔	Dedicated study	Indicator for anthropogenic emissions. Could be estimated from nearby Max-DOAS or GOME-2/S5P data.
Station availability level	Important	✔	✔	Dedicated study	Input from PIs needed to determine reasons for missing observations (e.g. weather vs. downtimes).
Station location and altitude	Important	✔	✔	Well known	Easily accessible for all networks.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Availability of aircraft or aircore calibrations	Useful	✔	✔	Additional work: available from TCCON aircraft profile database.	Similar database may exist for COCCON.
Cloud cover frequency from station data	Useful	✔	✔	Dedicated study	Unclear how many stations have direct solar radiation measurements on site.

Table 5: Like Table 4 but refering to instrumental parameters.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Principal instrument type	Critical	✔	✔	By network affiliation	TCCON requires Bruker IFS125HR, COCCON Bruker EM27/SUN.
Type and quality of on-site pressure measurement	Critical	✔	✔	Station survey	More stringent procedures for pressure measurements and QC are also currently discussed for TCCON.
Instrument firmware	Important	✔	(✔)	Station survey	Firmware may have an influence on Level-0 processing as well on data format.
Instrument hardware configuration	Important	✔		Station survey	TCCON instruments are modular and can be run in different configurations. They can also be upgraded. This includes the main electronics but also choice of detectors and beam splitter.
OPD / spectral resolution	Important	✔		Station survey	Spectral resolution is determined by instrument type and OPD of individual scans. Large range of options available for IFS125HR instruments, including mimicking low-res EM27/SUN observations.
Principal operating mode	Important	✔		Station survey	Some stations do TCCON and NDACC observations with the same instrument. Mostly affects duty cycle for XCO_2 observations.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Scanner speed	Important	✔		Station survey	Scanner speed and OPD determine how long individual samples take. This influences SNR and the ratio of cloud-free observations.
SNR	Important	✔	✔	Station survey	Can be derived from Level-1 spectra by the PIs. Needs a clear definition of how noise and signal should be determined for all stations within each network. In the future, GGG2020 will automatically derive this from the interferograms.
Spectral range	Important	✔	✔	Station survey	Spectral range is determined by choice of detectors and filters. Several options available for both IFS125HR and EM27/SUN.
Sun tracker specifications	Important	✔		Station survey	TCCON stations use a range of different sun trackers that have to fulfil the TCCON standards. COCCON is more uniform as sun tracker is part of the instrument.
Instrument age	Useful	✔	✔	Station survey	Principal hardware components in the instrument series change over the years.

Table 6: Like Table 4 but listing parameters that refer to data products, data quality and network characteristics.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Aerosol background statistics	Critical	✔	✔	Dedicated study	Could be from local AOD measurement, Aeronet, CAMS or other NWP/aerosol model.
Averaging kernels	Critical	✔	✔	Available from public data	In GGG2014, AKs are not stored individually per measurement but can be derived from SZA. Will be the same for GGG2020 data. Procedure should be similar for COCCON.
Data products and quality	Critical	✔	✔	Available from public data	Retrieved species with all attached QC information. Similar for COCCON.

Parameter	Priority	TCCON	COCCON	Approach	Comment
Typical inter-station bias for XCO_2	Critical	✔	✔	Dedicated study	Some literature available for GGG2014. However, station bias expected to be reduced with GGG2020.
Typical noise for XCO_2	Important	✔	✔	Additional work: can be derived from existing time series.	Need clear instrument-independent definition of noise comparable to expected CO2M performance
CO_2 sub layer potential	Useful	✔	✔	Additional work: data processing effort by TCCON PIs required.	Not a standard data product but potentially available at most TCCON stations [ROC2020]. Very limited number of DOF ~2-3 though. Works best if additional surface CO2 concentration measurements are available. GGG2020 will provide additional XCO_2 data products with higher sensitivity in the lower and upper atmosphere, respectively.

3.2 Emission monitoring

It has been already identified by the scientific CO2M Level-02 algorithm providers, that there is a lack of stations at emission sources, and that the XCO_2 columns derived in and around emission plumes show the largest differences between algorithms (as opposed to the background situation). Potential errors in quantifying these plumes have the largest impact on the CO2M mission objective.

In TCCON, some stations are located at or near large emission hotspots: megacities like Tokyo, Mexico City, Los Angeles, Paris Hefei, Xianghe but also in medium-sized cities like Bremen or Karlsruhe. Some are routinely affected by large natural emissions from biomass burning (Ascension, Darwin, Burgos). It has already been demonstrated that the CO_2 emissions from a megacity like Tokyo can be estimated from a single TCCON station and auxiliary meteorological data [BAB2020]. Still, for emission monitoring, it would be very beneficial to have several stations located around the source. This is only the case in the Los Angeles region with two TCCON instruments. On a larger scale, the principal has been demonstrated with TCCON instruments for methane [WUN2019].

COCCON has several advantages when it comes to emission monitoring. There are more available instruments and is much easier to set them up in a ring around an emission source. It should be noted that the planned inversion of GHG emissions of target areas will be based on the observed gradients between different ground pixels of the satellite sensor, so an independent estimation of the emissions from COCCON is not required for the purpose of validation. Instead, primary validation targets are the columnar gradients. These gradients can be sampled with reference accuracy using an array of several COCCON spectrometers in the area under study. The principle has been demonstrated by COCCON campaigns, for example around Berlin [HASE2015]. However, this approach requires a large number of COCCON instruments: ~5 to 10 for a city the size of Berlin. This is one of the reasons why such observations have mostly been done on a campaign basis so far. A permanent city column observation network is currently being set up in Munich [DIET2020] and Toronto.

There has been some prior work on observing emission sources from space [KOR2014]. Combining ground based and satellite GHG observations for the purpose is still a new field [TU2020]. Assessing what is needed to validate and monitor the performance of CO2M retrievals for these emission monitoring conditions using ground-based observations would probably require a dedicated study.

3.3 Network modelling

Network modelling describes a set of skills and tools that are needed to assess the performance of the ground based column networks with respect to CO2M cal/val needs. This includes both the performance of the existing network as well as potential alternative configurations. Ideally, the network modelling framework should be able to address many of the aspects listed in Tabs. 4, 5, and 6. This should also work on different horizontal scales: to assess the global TCCON/COCCON network as well as multi-instrument configurations around emission sources.

Such a network modelling framework does not yet exist for the ground based TCCON and COCCON networks. Some attempts have been to simulate footprints of existing TCCON stations [BEL2017]. Figure 3 shows an example of the global coverage of TCCON in 2017. However, this was a one-time project and the system is not available any more. No such analysis has been made for COCCON so far.

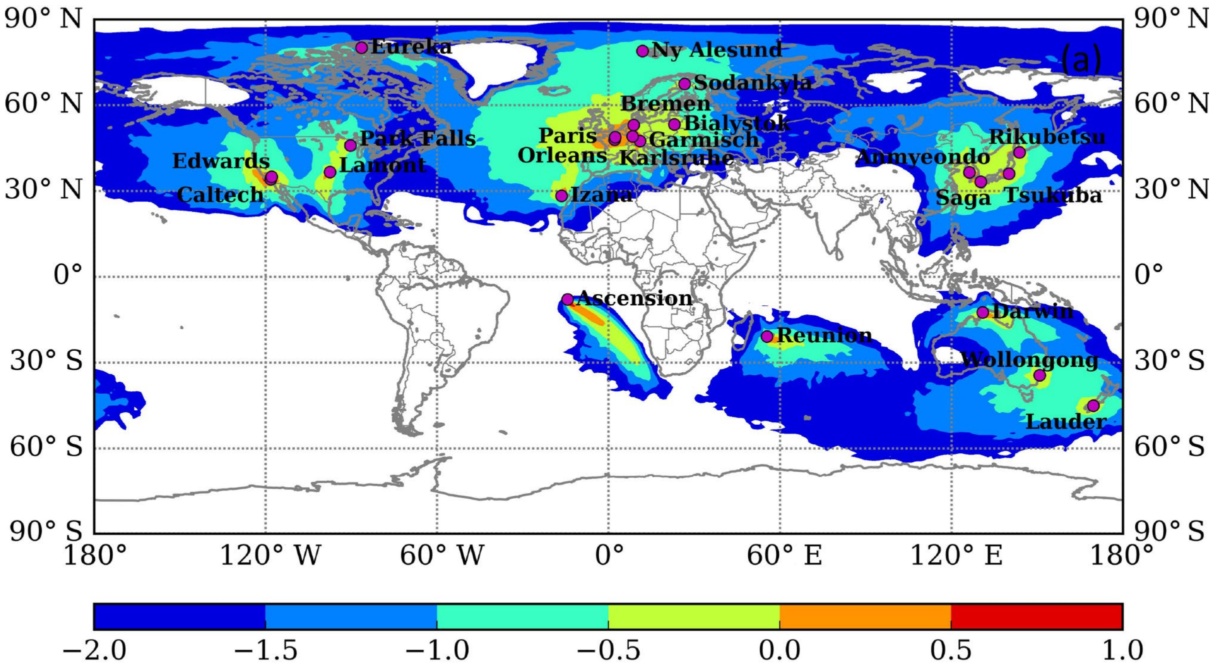

Figure 3: TCCON footprint analysis [BEL2017].The colors indicate where the sampled air in the column above each station came from.

The exact specifications of a network modelling framework should be defined in a follow-up study with groups that have experience in setting up such a system. However, there are some suggestions on critical aspects that should be covered. In particular, the network modelling framework should be able to

- make use of a still-to-be defined data structure that describes many station parameters and is extensible for future use.
- simulate the effects of at least the critical station parameters defined in Tab. 4 which affect the impact and optimal placement of stations.
- simulate the effects of at least the critical instrument parameters defined in Tab. 5 which affect the performance of individual stations.
- simulate the effects of at least the critical parameters defined in Tab. 6 which refer to data products, data quality and network characteristics and affect network performance as a whole.

- simulate both TCCON, COCCON as well as mixed configurations.
- do such simulations also for hypothetical stations for future network planning.
- use realistic input data fields, e.g. from CAMS, to be able to handle a wide range of meteorological and seasonal conditions.
- simulate key aspects of the CO2M instrument like orbit parameters, species, FOV, SNR.
- handle at least basic approaches for plume observation.

Many of these tasks could be handled by individual specialized modules. Creating a fully-integrated OSSE would probably be too much effort.

4 TCCON ALGORITHMS, DATA PROCESSING, AUXILIARY DATA AND SPECTROSCOPY

4.1 Algorithms and software

GGG is the retrieval software for TCCON. The official TCCON data product is the output of the official current GGG release. TCCON PIs may process their spectra with other retrieval codes but the resulting data products may not use the TCCON label.

GGG is developed by a small core team with contributions from other TCCON members. Some of the work of the core team is currently funded by NASA. The software is open source under a royalty-free Caltech licence. Under the terms of the licence, TCCON members are allowed to use and modify the software but not redistribute it to 3rd parties outside the TCCON community. Such parties need to request a licence from Caltech. US export regulations may apply in some cases.

GGG is distributed as Fortran source code with additional shell scripts and tools written in Python and Perl. The users compile binaries from the source code according to recommendations regarding choice of compiler and compilation options. The reference platform is a Linux system without further specification of the actual distribution. GGG2020 has been tested to run on several flavours of Linux, MacOS, and on Windows using CygWin and the Windows Linux Subsystem. It is benchmarked off of Linux RedHat Enterprise Linux Server release 7.7 (Maipo) with gfortran v4.8.5. The GGG distribution contains benchmark input and output files to check that the locally compiled version produces the same results within reasonable numerical limits.

The principal building blocks of GGG are:

- the I2S package that produces spectra from the interferograms.
- the GSETUP code that handles the atmospheric model.
- the GFIT code that handles the complete forward model including spectroscopy and instrument model. It retrieves the column abundances for all species in the retrieval microwindows.
- the postprocessing scripts that derive column-averaged dry-air mole fraction for each target species along with a large number of auxiliary parameters that are needed for QC. The initial automatic QC and the calibration of TCCON to the WMO GHG scale is also handled by the postprocessing scripts.

In the past, new GGG versions were released every 2-3 years. With each new release, all TCCON stations have to reprocess all of their data. This can take weeks to months for the whole network depending on which data levels have to be processed.

After a long period without major updates, a new version is planned to be released before the end of 2020. The current situation looks like this:

- **GGG2014** has been the official operational version since October 2015. Most of the descriptions in the following sections relate to this version. This version is documented in detail in [GGG2014].
- **GGG2020** will replace GGG2014 in 2021. There will be major changes regarding spectroscopy, the sources and handling of auxiliary data, use of priors and the postprocessing concept. This version only requires Free and Open Source Software to be installed on the host system. Early comparisons showed a significant reduction of site-to-site biases with respect to GGG2014.

4.2 Data processing levels and file formats

There is no official processing level numbering scheme in the TCCON community. The following level definitions are the ones used internally for the processing of the Ascension Island TCCON station.

Level-0 and -1 files are in Bruker's proprietary OPUS format. This format is common in the FTIR spectroscopic community. However, the documentation for the OPUS format and software that reads and writes it is not public and only available after signing an NDA. For this reason, COCCON has decided to store its spectra in a new community-defined binary format in the future.

4.2.1 Level 0A: interferogram slices

During the measurement, the Bruker IFS125HR transmits interferogram data through the HTTP protocol to the client software. The interferogram is divided into slices of ~1.5 MB each. Each slice file contains the ADC readout from each detector (up to two) plus additional meta data: e.g. timing information and acquisition parameter settings in OPUS format. The total number of slice files depends on the configured spectral resolution. A typical interferogram consists of ~20 Level-0A files with a total size of ~30 MB.

4.2.2 Level 0B: merged interferograms

The standard output format for interferograms produced by Bruker's OPUS software is a merged file that is constructed from the individual slice files. This file contains two interferograms: one from the forward and one from the backward scan of the interferometer as well as the meta data from the Level-0A files. Some TCCON instruments use Bruker's proprietary OPUS software to run the measurements. In this case, this is the output format for interferograms. There is also Python code that can generate these Level-0B files from Level-0A slice files.

For single scans (the standard configuration), the slices are simply concatenated w/o any additional processing. For multiple scans, the resulting interferogram data are also averaged. For single scans, the total size of a Level-0B file is the same as the sum of the sizes of the corresponding Level-0A files: ~30 MB.

Both Level-0A and 0B files contain interpolated ADC counts from the detectors. The instrument's firmware is responsible for interpolating the constant-time based samples onto the constant-distance grid that is needed for the Fourier Transformation. The procedure is described in detail in [SIM2002]. The underlying raw sampling data are not available from the instrument.

4.2.3 Level 1: spectra

This is the first real processing level where the data are not just recorded from the instrument or simply reformatted. For an FTIR instrument, the spectra have to be generated from the interferograms through a numerical Fourier Transformation. This processing is done by the I2S (interferogram to spectrum) module of GGG. I2S can handle both Level 0A as well as Level 0B files, so the TCCON sites are free to chose how they collect their Level 0 data (with or without using OPUS). The output files from I2S contain the spectra and metadata about the data acquisition parameters (taken from the Level 0 input).

I2S is usually run in batch mode on a large number of spectra. The input file for a batch contains a header with configuration options as well as the list of Level 0 files to be processed. Many configuration options are not critical , e.g. the ones about paths to input and output files, naming convention for the output spectra, time zone correction.

A few parameters have a direct influence on the resulting Level-1 data and their quality:

- **Phase correction for the FT:** There are network-wide recommendations for these parameters but some sites had to adjust them in the past for their specific instrument setup and performance.
- **Quality check based on the solar tracker intensity:** this is mainly a sanity check to put a threshold on the Level-0 files that are processed. The task could as well be handled by the site-specific pre-processing software that prepares the files for Level-1 processing in the first place. Level-2 processing QC later applies the final threshold on the solar tracker intensity and its variation.

- **Lower and upper frequency limits for the FT:** these are defined in a site-specific (but not run-specific) separate configuration file. The numbers depend on the instrument's detector and optical filter configuration. They remain the same as long as the configuration does not change. The numbers are also not very specific: two instruments using the same type of detectors would use the same settings.

The Level-1 processing produces separate output files including one spectrum each. For the typical TCCON setup (two detectors, forward-backward scan), four Level-1 files are generated per Level-0 file:

- One each for detectors A and B for the forward scan of the interferogram.
- One each for detectors A and B for the backward scan of the interferogram.

Note that in the TCCON community, the Level-1 files are considered intermediate data products as they can always be reproduced from the Level-0 files using the same configuration files and the same I2S version. They are also not required to be archived when a new software release requires a Level-1 reprocessing. It is still good practice to keep all the I2S input files. However, the individual PI has to take care of that.

4.2.4 Level-2: Column-averaged dry air mole fractions plus meta and auxiliary data

This is the official TCCON data product. The published version contains the column-averaged dry-air mole fractions of CO_2, CH_4, CO, and other species along with several dozen diagnostic parameters that are generated during the retrieval. The non-public engineering version includes the same along with several hundred additional diagnostic parameters.

The file format and the retrieval are too complex to discuss here in detail. They are described in detail in the GGG2014 dataset documentation [GGG2014]. Some general notes:

- File format for both the public and engineering Level-2 files is netCDF.
- The TCCON Data Archive takes care of long-term archiving of both public and engineering files.
- One row in the Level-2 file corresponds to a single spectrum (one Level-1 file).
- The engineering version contains all rows. Including the ones flagged by the post-processing QC. The public version contains only the unflagged rows (so the quality flag is always zero).
- Several intermediate files in various formats are produced during the Level-2 processing. These are not described here.
- Currently, the NetCDF file production is done at the TCCON Data Archive from intermediate ASCII-based files produced during the local processing by the PI. This will change with the upcoming GGG2020 version as NetCDF files are produced directly.
- The upcoming GGG2020 release employ Unidata naming conventions for the data and metadata in the NetCDF files. This is currently not the case for GGG2014 data.

4.3 Auxiliary data

The current GGG2014 software relies mostly on NCEP Reanalysis data. The atmospheric profiles for GGG are produced from NCEP netCDF data files. Note that these data are only available with a delay of 4 days from the NCEP archive. A script interpolates the NCEP data horizontally to the station's location and vertically onto the TCCON vertical pressure grid. The output format is a simple internal ASCII-based format (mod-file). Caltech runs a central server that produces these mod files for all stations. Alternatively, the station PIs may runs the same script locally for their respective site.

Prior profiles are generated from a climatology that includes latitudinal and seasonal variability as well as trends for all major TCCON species. The format is also a simple internal ASCII-based format (map-file). For both mod and map files, only one file per observation day (representative for local noon) is produced.

For GGG2020, the principal data source for meteorological profiles will be NASA's GEOS-FP-IT, which is available on a 3-hour temporal resolution. Similarly to NCEP, these data are only available with a delay of a few days. Mod and map files will be centrally produced by Caltech and distributed to the TCCON PIs online.

Table 7: Auxiliary data sets required by the current and future GGG versions

Auxiliary data set	GGG2014	GGG2020
Temperature (over pressure)	NCEP Reanalysis (6h)	GEOS FP-IT (3h)
Geopotential height (over pressure)	NCEP Reanalysis (6h)	GEOS FP-IT (3h)
Specific humidity (over pressure)	NCEP Reanalysis (6h)	GEOS FP-IT (3h)
Tropopause pressure	NCEP Reanalysis (6h)	GEOS FP-IT (3h)
Priors (CO_2, CH_4, CO, ...)	TCCON climatology	Updated TCCON climatology

Principally, the mod files used by GGG can be generated from any source. Since 2015, there have been discussions of switching from NCEP Reanalysis to other meteorological data products like the ones provided by ECMWF. The main drivers were an improved temporal and spatial resolution and the removal of the delay enforced by the NCEP data policy. However, it turned out that access to these ECMWF products would not be easy to arrange for all TCCON PIs (especially non-European). So the plans were given up.

The central approach for the production and distribution of mod files taken for GGG2020 would make it much easier to also use other data sources for the whole TCCON network. However, there might be legal issues with some data products that may not allow redistribution to all TCCON members.

In general, TCCON PIs are free to process their data with mod files generated from alternative data sources to generate alternative data products. However, the official TCCON data product has to be processed in the way that has been approved by the TCCON PIs to ensure network-wide quality and consistency.

4.4 Spectroscopy

4.4.1 Line List

TCCON uses a dedicated line list which is compiled from different HITRAN versions and optimized for the TCCON retrieval windows. The whole line list is validated against spectra from the Kitt Peak Observatory to circumvent deficiencies in HITRAN [TOON2016]. The line list is part of the GGG software distribution, so it only changes when a new official version is released. It is also available as a citable digital object:

- Geoffrey C. Toon. Telluric line list for GGG2014. TCCON data archive, hosted by CaltechDATA, 2014. doi:10.14291/tccon.ggg2014.atm.R0/1221656.

The new GGG2020 version will contain a new line list that includes changes proposed during the annual TCCON meetings since 2015. It will also contain additional parameters for non-Voigt line shapes for some species.

4.4.2 Development needs

Spectroscopic parameters from HITRAN typically have accuracies on the order of a few per cent. This limits the accuracy of the TCCON XCO_2 retrievals with respect to the WMO CO_2 scale used for ambient-air measurements by ICOS and other networks. These measurements are accurate on the order of 0.1% or better. A bias in the line strength for key CO_2 or O_2 lines has the effect of scaling the whole network's XCO_2 retrieval results by a constant factor. Such scaling factors have been identified and determined for all TCCON key species. All published TCCON data are corrected with these scaling factors in order to calibrate them to the WMO scale [GGG2014]. This is the basis for the (indirect) calibration of satellite GHG observations to the WMO scale.

The usual way to determine these network-wide scaling factors is to take altitude profiles of the CO_2 concentration with aircraft or balloon instruments (aircore) over a TCCON station while it is making observations. So far, dozens of such profiles have been collected over several TCCON stations. These data can be reused to derive the scaling factors for every new GGG release. However, opportunities for aircraft overpasses of TCCON stations are rare and only a handful of stations fulfil the criteria for aircore launches.

For many reasons, it would be desirable to have a more direct procedure. One suggested way would be to make the ground based remote sensing observations directly traceable to SI. This requires spectroscopic laboratory measurements made in a fully SI traceable way with precision improved by about one order of magnitude. At the same time, a rigorous traceability analysis for the ground based retrieval code would have to be done (in principle, a systematic error propagation analysis based on Monte Carlo methods). The feasibility of both of these steps has already been demonstrated for small sets of spectroscopic parameters and codes not related to GHG retrievals.

Unfortunately, two proposals by the TCCON and European NMI community that aimed at providing SI-traceability for the TCCON spectral parameters and the GGG code have been turned down by the European Metrology Programme for Innovation and Research (EMPIR). Since Ocotber 2020, ESA and NASA are running a joint study that should improve the accuracy of many critical spectral parameters for satellite and ground based GHG retrieval [ESA-EOPG]. However, NMIs that could also provide SI-traceability for the results are not involved in the study.

5 NETWORK DESIGN FOR PROCESSING, DATA TRANSFER AND OPERATIONS

This section describes key aspects of an operational ground based cal/val network that would fulfil the needs of the CO2M mission. Some of these aspects are covered by the existing ground based networks TCCON and COCCON. However, many that are not covered concern the operational level of the networks. The description of this gap does not imply that the networks would endorse all steps of a transition to a higher operational level. This has to be discussed with the TCCON and COCCON PIs in the future.

5.1 Data delivery time scales and applications

For operational users, timely and reliable data delivery is critical. The actual time scale depends on the application. A general classification, was defined in [EGHGCV]:

- **Final product:** <1 year with a data availability of ideally >95%
- **A posteriori:** <1 month
- **Near real time:** <2-3 days
- **Real time:** <3 h

5.1.1 Final product data delivery

Final product data delivery can be met by the existing networks and is actually a requirement for participation in TCCON. These data meet the highest quality standards and are publicly available. For TCCON, all publicly released data have passed automatic QC as well as intensive visual screening of the whole time series of the main and many auxiliary parameters by several PIs. While TCCON PIs are allowed to embargo their data for up to 12 months, most release their station's data within 3-6 months. It is expected that COCCON will set up similar rules and procedures.

For CO2M, the final product is most useful for long-term validation of all data products. It is essential for reaching the highest data quality.

5.1.2 A posteriori data delivery

A posteriori data delivery has been tested by some European TCCON stations for a limited time in an ESA preparation study for S5P. The requirement in the project was delivery of the final TCCON data product to the TCCON Data Archive within 3 months and subsequent delivery to EVDC. To meet that goal, the processing schedule was streamlined so that new data could be processed just in time for the regular QC and upload procedure performed at the TCCON Data Archive. In many cases, this resulted in a data delivery within one month after measurement if no issues showed up during the final QC. The procedure could not be held up over an extended period of time due to limited availability of key personnel.

For CO2M, an a posteriori data product is probably most useful for the validation of higher-level data products that require additional inputs – even if the time scale is slightly larger than one month.

5.1.3 Near real time

Near real time data delivery cannot be achieved within the current network operational setup. The main reason is that the final data product (at least for TCCON) requires human intervention in the final QC process where the data are checked for quality issues that have not been found by the automatic QC – a procedure that has turned out to be crucial for the data quality a number of times. So there is a trade-off between quality and timeliness which can be resolved by an additional data product. Such a rapid delivery data product with reduced QC has already been tested for some TCCON sites and should also be feasible for COCCON as well. However, the required effort for near real time data delivery is large: It requires duplication of the complete data processing chain as well as additional efforts that are described in Section **Error! Reference source not found.**.

For CO2M, such a near real time data product (even with reduced quality) would be most useful for the validation of a future operational data assimilation system that assimilates CO2M data in near real time.

5.1.4 Real time

Real time data delivery is a huge effort because of the large amounts of Level 0 and Level 1 data produced by the FTIR instruments. This is especially true for the TCCON instruments with their high resolution. While the principal approach is not be very different from the near real time case, the additional effort of getting all the required input data products for the processing chain aligned in time should not be underestimated. Especially remote stations will likely not have the bandwidth to transfer Level 0 data in real time. So they will require local pre- or even full processing as lined out in Section **Error! Reference source not found.**. However, data transmission is not the only problem that has to be solved: reaction times to instrument warnings or failures would also have to be pushed from days to hours.

For CO2M, real time data delivery would be most useful for in-flight calibration or health monitoring of the satellite instrument.

5.2 Required resources for different data delivery time scales

The resources required to meet the data delivery time scales defined in Sect. **Error! Reference source not found.** do not change linearly. Different resources become more critical with increasing requirements on timely data delivery. So far, only the standard case of final data product delivery is fully covered.

Figure 4 provides an estimate for the additional resources that are required to meet more demanding data delivery time scales. The numbers are based on experience and general assumptions. For example: at a dedicated upload bandwidth of 256–512 kbit/s and without accounting for downtimes, the average amount of Level 0 data produced by a TCCON instrument during one year can also be transferred within 12 months. However, to be able to transfer the Level 0 data in real time would require a dedicated upload bandwidth of 1024–2048 kbit/s. This may not sound like much but it is an issue for remote sites.

On the personnel side, the work load increases by a factor of 3-4 if the final data product including the human QC has to be produced monthly instead of every 3-4 months. On monthly time scales, leave times of key personnel would have to be covered by additional personnel. This personnel redundancy becomes even more important with NRT or RT data delivery. Even though the data processing would be fully automatic, it would not be enough to monitor instrument health on a weekly basis. Instead, it would have to be monitored daily during 5 days per week for NRT or even hourly during 7 days per week for RT data delivery.

The required redundancy of key equipment components depends on the maximum downtime that would be acceptable for NRT or RT applications. With the assumed redundancy factor of 2 for NRT and RT, typical downtimes could be reduced from weeks to days with trained personnel on site.

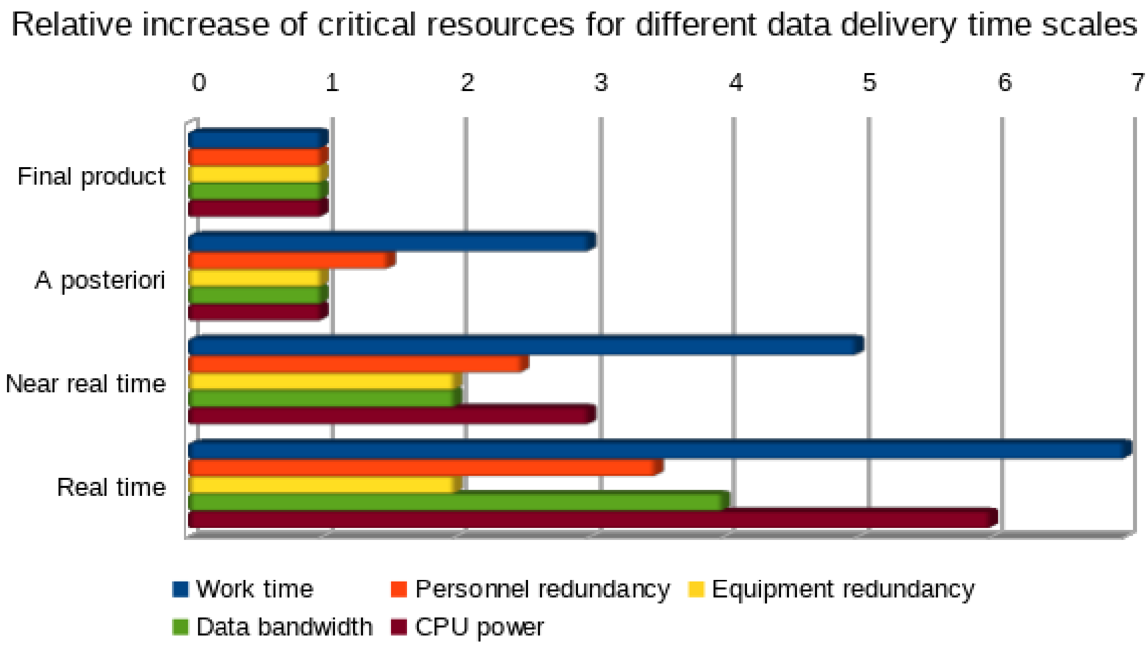

Figure 4: Estimate of required critical resources for different data delivery time scales. All numbers are relative to the final product case which is already covered by the networks.

5.3 Data processing and delivery structures for rapid delivery data products

This section describes data processing options for the ground based networks that would be required for rapid (NRT and RT) data delivery. They would be optional for a posteriori data delivery and mainly useful if non-standard data products (e.g. processed with alternative meteorological fields or different spectroscopy) should be used for the CO2M validation. No changes would be required for use of the standard final product that is routinely delivered by the networks.

As COCCON already works on the development of centralized data processing facility, a straightforward approach in this context would be to provide support for upscaling the already existing CF capacity according to the mission needs. This scheme offers the advantage that the CF would be implemented in a broader than pure mission-related context, therefore with the aim of becoming a permanent body and of serving all COCCON partners independent of their involvement in certain validation tasks. Such prospect might increase the acceptance of the CF services by non-European partners (and the potential of granted co-use of their data in the mission context), which otherwise might prefer to build up own data processing capabilities.

5.3.1 Central processing by EUMETSAT

Central processing means that Level 0 or Level 1 data are transferred from the (European) TCCON and COCCON ground stations to EUMETSAT CF to be processed. At EUMETSAT, the data of all participating stations would be processed using a controlled software release and auxiliary data sets like spectroscopic line list and meteorological fields. This should be possible for both TCCON and COCCON with comparable effort.

The configuration of the processor could be identical to the standard TCCON and COCCON data products. However, it would also be possible to generate custom-made alternative or additional data products specifically for CO2M cal/val purposes. For example, it might be useful to run TCCON or COCCON retrievals using the same spectroscopic line lists and meteorological fields that are used for the CO2M retrievals in order to minimize biases that result from the choice of auxiliary data. All data products will be archived and distributed to the CO2M cal/val community and the PIs according to a data sharing agreement that has yet to be defined.

Central processing has specific pros and cons for EUMETSAT. Note that not all advantages for EUMETSAT will also be viewed as advantages by the networks:

- **Pros:**
  (1) EUMETSAT gets access to Level 0 and Level 1 data which are usually not published. These can be used to produce CO2M-specific cal/val data products. EUMETSAT would have the option to produce alternative data products, for example with different auxiliary input data.

  (2) NRT and RT data products can be fed directly into the CO2M cal/val system.

  (3) All data products, software revisions and auxiliary data sets are fully controlled and traceable for EUMETSAT.

  (4) A centrally administered operational data processing system could relieve the PIs of a part of the burden associated with running their stations operationally.

- **Cons:**
  (1) Central processing requires the largest resources in storage, data transmission and CPU power on EUMETSAT's side.

  (2) The legal issues concerning data ownership, licensing and distribution have to be negotiated with several partners from different countries and institutions. The potential length and complexity of such negotiations should not be underestimated.

  (3) The PIs might not be willing to submit Level 0 and Level 1 data which are usually not published.

  (4) A central processing system at EUMETSAT might be viewed as a duplication of effort by the PIs if they still have to process their data themselves anyway or if they are required to submit their data to more than one central processing facility (see Sec. 5.3.2).

If the setup of the central processing system is identical to the official TCCON or COCCON data product, some PIs may find it useful to have EUMETSAT take care of processing their data instead of doing it themselves. EUMETSAT could offer central processing as a service for the European TCCON and COCCON communities in exchange for access to the lower level data. Whether this is interesting to the PIs will probably depend on how the following questions can be settled:

- Would EUMETSAT be willing to process the whole time series of participating stations or just the time window that is relevant for CO2M cal/val? This is important as every major software update for the final product requires a complete reprocessing of the whole time series for each station.
- Who is responsible for the further QC of the data product? The PIs will likely want to keep this in their hands.
- Who decides which parts and levels of data will be published and in which form? Typically, only Level 2 data are published. Also, TCCON does not support redistribution of the official data product outside the TCCON Data Archive. The rules for COCCON still have to be set up. This use case is outside the standard data license and use policy and will have to be negotiated with the PIs.
- How can updates and changes in the official data processing procedure be implemented into the EUMETSAT processor? Will selected PIs be able to access the system for installation and testing?
- What would be the role of non-European sites? From EUMETSAT's perspective, they would probably be welcome as CO2M cal/val needs will not be limited to European sites. What would be the conditions and potential incentives for them to participate?

Central processing with dedicated data products optimized for CO2M cal/val would offer most control over the whole process to EUMETSAT but would also require a large contribution for setting up and running the central processing infrastructure. A successful implementation would have to be set up in way that it is regarded as a useful service by the TCCON and COCCON communities. As many stations as possible – both European and non-European – would need to be motivated to participate.

5.3.2 Central processing by a third party

EUMETSAT is not the only institution that would be interested in centrally processing TCCON or COCCON data. This might also be considered by other space agencies that run GHG observing satellite instruments. Also, if TCCON can be integrated into ICOS (see Sec. **Error! Reference source not found.**), central processing of the participating stations' data by the ICOS Carbon portal would be required.

At first glance, it might seem attractive to delegate the necessary resources to a third party and negotiate access to the data. This might not even be necessary if the data is published timely and under an open data license. This should be the case for ICOS.

However, in this case EUMETSAT would not be dealing directly with the TCCON and COCCON PIs but rather with the third party producing the data. EUMETSAT's influence on the requirements and configuration of the data products would be limited. Also, EUMETSAT would have only little influence on the set of stations that contribute to this service. Avoiding duplication of efforts would be a major challenge in a scenario where several parties aim for central processing.

In case ICOS would come up with an operational maintenance and operations concept, to be implemented by the participating TCCON stations, the EUMETSAT requirements would need to be checked against the ICOS requirements as these are not necessarily identical. Any deltas would need to be identified and then addressed individually through EUMETSAT/CO2M-specific upgrades of the cal/val reference data needs. This could, for example concern quality assurance levels and latency as well as dedicated data products for CO2M cal/val.

5.3.3 Distributed processing with dedicated hard- and software on site

If transmission of Level 0 data from the stations to EUMETSAT CF is not an option, distributed processing at the stations may be an alternative for NRT and RT data delivery. In a distributed processing system, centrally maintained processor software is installed on dedicated hardware at the stations. The Level 0 data would be transferred locally from the station to the dedicated hardware where the further data reduction would take place.

The hardware and its software can be maintained, monitored and updated remotely by EUMETSAT using centralised distribution and repositories. Maintenance and synchronisation, configuration control and update of hardware, software and provision of common auxiliary data and settings (ECMWF) could be carried out in a controlled way. Customized data products for CO2M cal/val would be possible to improve consistency of results.

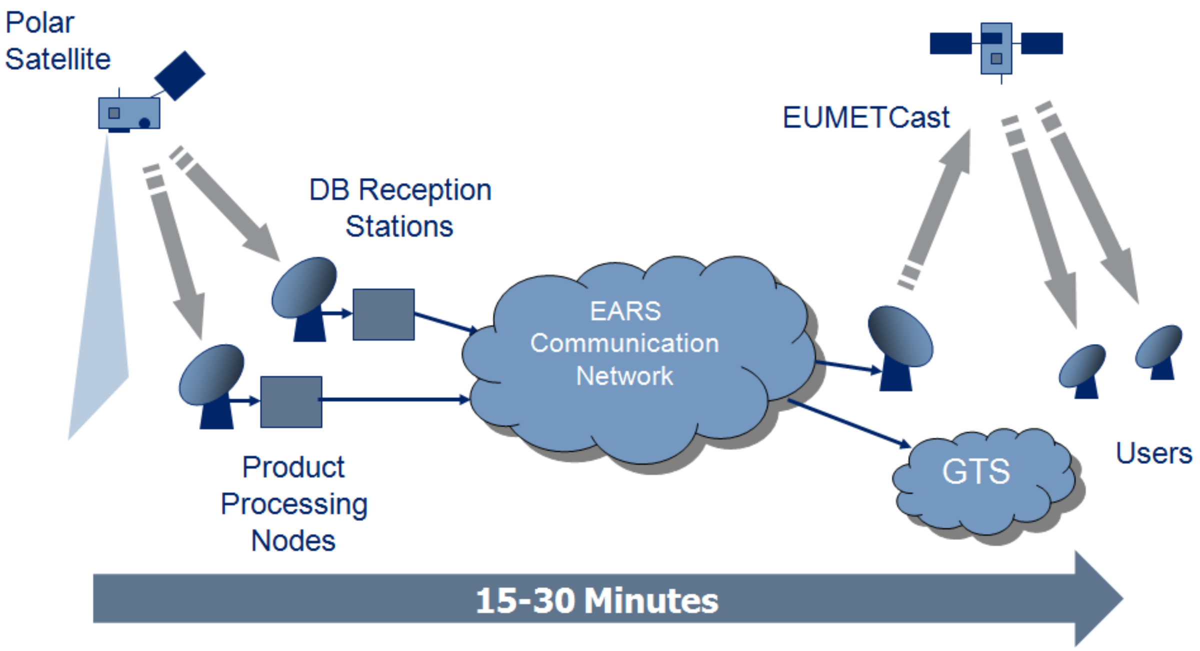

Figure 5: Setup of EUMETSAT's EARS network.
Source: https://www.eumetsat.int/website/home/Data/RegionalDataServiceEARS/

EUMETSAT is already using a similar concept for its EARS network as illustrated in Fig. 5. The different EARS services provide users with high timeliness regional data from polar orbiting meteorological satellites, in support of NWP and NWC applications.

The benefit of distributed processing depends on which levels of data processing would be done on site. For FTIR instruments, Level 1 spectra are not significantly smaller than the Level 0 interferograms. However, Level 2 data products are at least 2-3 orders of magnitude smaller than the Level 0 data but also need a lot of auxiliary data that has to be transferred. So the pros and cons of distrubted processing are

- **Pros:**
  (1) Bandwidth requirements for NRT and RT data transmission are greatly reduced, especially if Level 2 processing can be done on site.
  (2) Data ownership issues can be avoided if the Level 0 data do not have to be transferred.
  (3) EUMETSAT would still have full control over the data processing.
- **Cons:**
  (1) The effort and cost of installing the dedicated hardware might be comparable to acquiring the necessary bandwidth for timely Level0 data transfer, e.g. with a dedicated VSAT satellite communication terminal on site.
  (2) Operation and maintenance of the dedicated hardware could be challenging at remote sites. The remote software administration should be less of a problem.
  (3) With new stations being added to the network over time, it is unlikely that the dedicated hardware can be kept identical across sites.

In general, distributed processing only seems to be useful for sites where an upgrade to a bandwidth that allows at least NRT transmission of Level 0 data is either too expensive or impossible due to technical limitations. Physical installation of distributed hard- and software at stations with sufficient data transmission capability would not be reasonable.

5.3.4 Community processing

An alternative to the concepts above could be dedicated data processing by the TCCON and COCCON communities for CO2M cal/val according to EUMETSAT's requirements. In this case, the TCCON and COCCON data processing and data provision procedures would be basically used as they are. EUMETSAT would only request changes or additional procedures that are necessary to meet its requirements on data provision and latency as well as QC. The networks would be responsible for the implementation. This assumes, of course, that the necessary resources for this effort are provided − also for the non-European partners.

This is similar to the approach that has been used before for the cal/val for OCO-2 and OCO-3 with TCCON. The TCCON PIs are informed about OCO-2/3 target mode observations and prioritize their data processing accordingly. The target mode observations are processed with a shorter latency and the PIs also submit Level 1 data to the OCO-2/3 teams. However, since no additional resources are available for this service, the special handling of OCO-2/3 target mode data by the TCCON PIs is on a best-effort basis only. Providing dedicated cal/val data on an operational level would require much larger efforts and additional resources comparable to the estimates in Fig. 4.

For EUMETSAT, the pros and cons of the community processing concept would be:

- **Pros:**
  (1) Direct cost and effort for EUMETSAT would be small compared to the central and distributed processing options. Only the infrastructure for receiving and handling the dedicated cal/val data would have to be set up and maintained by EUMETSAT. The bulk of the work would be handled by the TCCON and COCCON communities.
  (2) This model would be a good basis for a collaborative partnership between EUMETSAT and the TCCON and COCCON communities.
  (3) Legal issues concerning data ownership and licenses as well as intellectual property can be avoided.
- **Cons:**
  (1) EUMETSAT would have less control over the whole process compared to the central and distributed processing options. There will be a higher coordination effort with the networks partners and the implementation of changes will likely take longer.
  (2) Data delivery and timeliness depend on the resources available to the networks and the individual stations.

For the community processing option, the necessary investment into the setup and maintenance of a data processing infrastructure would be comparably small on EUMETSAT's side. However, for reaching the same data quality and timeliness requirements that EUMETSAT needs, the coordination effort would be higher. EUMETSAT and the TCCON and COCCON communities have many shared interests in the future CO2M cal/val programme. The best approach should be a long-term partnership with a close collaboration.

6 REFERENCE STANDARD (NETWORK MONITORING)

In the context of operational cal/val and monitoring of CO2M, the reference standard and network monitoring refers to the implementation and maintenance of certain procedures in order to reduce the inter-station biases and improve the overall network data quality and bias consistency.

Note that in this context, the network reference standard does not refer to the absolute accuracy level that has been or should be achieved. The improvement of instruments, processing, quality of auxiliary data, and spectroscopic accuracy is implemented elsewhere: e.g. via improvement of network infra-structure and instruments, through improvement of retrieval software, or progress in the field of spectroscopy.

The following procedures for increasing and maintaining network product quality consistency have already been implemented or are planned.

6.1 Initial hardware requirements

Both TCCON and COCCON have specific hardware requirements for their member sites. The primary instrument type and model are clearly defined. However, a variety of instrument options exist on the manufacturer side, especially for the high-resolution Bruker IFS125HR used by TCCON but also for the EM27/Sun used by COCCON. Examples are internal filters, detectors, or beam splitters which can be exchanged easily even during operation. Especially for the IFS125HR, a number of different modules are available for the initial configuration. These can be used to increase the spectral resolution of the instrument or for putting liquid or gas samples into the light beam. The choice depends on what other tasks the instrument may be used for or how much space or funding is available.

Additional components like the solar tracker or supporting hardware like PCs, vacuum pumps, or environmental sensors may vary even more as the choice depends on the site conditions. These are typically not standardized as long as they meet reasonable requirements.

- **Implementation by TCCON**
  - Bruker IFS125HR instrument with defined set of spectral range and resolution required. Additional detectors, spectral range or higher resolution (e.g. for NDACC observations) are optional.
  - Solar tracker is site-dependent and a number of designs exist. They all have to meet the pointing accuracy requirements defined by TCCON.
  - Environmental sensors have to meet defined precision and accuracy requirements.
  - Most instruments are evacuated during operation but this is not required.
- **Implementation by COCCON**
  - Bruker EM27/Sun instrument has fewer options, so hardware is more uniform across the network. However, there are options like an additional CO channel which are not yet implemented in a few instruments.
  - Solar tracker is an integral part of the instrument.
  - Requirements for environmental sensors not yet clearly defined.
- **Frequency:** once at initial site configuration and setup

6.2 Instrument alignment

Proper optical alignment of all components of the interferometer is critical for the performance of the instrument. The results of poor alignment are loss of light that is available for the interferometer as well as spectral artifacts. Especially at high spectral resolution, the spectral quality depends heavily on the ILS which is directly affected by the alignment quality. Alignment may change over time, so it should be checked and repeated regularly.

- **Implementation by TCCON**
  - The alignment of the IFS125HR is complicated as the TCCON instruments operate near or even beyond the standard specifications of the manufacturer. Therefore, the TCCON community has developed its own recommendations, procedures, and tools for the optical alignment of the instrument. These are distributed within the community through documents, videos and hands-on workshops.
  - The alignment changes whenever an instrument is moved or after major modifications. It may also change slowly over time due to wear of some internal components of the interferometer. Such changes are monitored by doing regular ILS measurements. If necessary, the alignment procedure has to be repeated on site.
  - For avoiding site-specific ILS biases, TCCON is using centrally calibrated HCl gas cells [HASE2013].
- **Implementation by COCCON**
  - Due to its lower resolution, the EM27/Sun interferometer is much more compact and rigid than the IFS125HR interferometer. The initial alignment typically is stable over several years even when the instrument is transported.
  - Alignment re-checking is typically done at KIT when instruments are returned for maintenance. If a re-alignment is needed, all calibration measurements are repeated before re-deployment. Re-alignment on site is not recommended.
  - Due to the lower spectral resolution, the ILS parameters can be derived from open-path measurements using water vapour lines. Such measurements can be performed also on site without the need for calibrated cells. In addition, the atmospheric observations themselves are exploited: a drift in the derived Xair serves as sensitive indicator for a possible drift of instrumental characteristics.
- **Frequency:** Once at initial setup for all instruments. For TCCON, alignment should be checked at least on a yearly basis. Depending on site conditions, some instruments require yearly re-alignment, some are stable over several years.

6.3 Maintenance of instrument-specific hardware components and status

Even for the same instrument type and model, the hardware configuration in the network will become more heterogeneous over time due to changes and updates by the manufacturer. Due to the long life time, the age difference between instruments in the networks may be years or even decades. Newer instruments will include different components due to improvements by the manufacturer or because previously installed components are not produced any more.

- **Implementation by TCCON**
  - The PIs are responsible for keeping track of the hardware components in their instruments. There is no central register for the age or type of components inside the spectrometers.
  - Most newer components are also available as upgrades for older instruments. However, most sites lack the funding for continuous upgrades of their instruments to the latest hardware components.
- **Implementation by COCCON**
  - So far, most of the COCCON instruments have been purchased during a short time window and are therefore more uniform. This may change over time.
  - KIT has access to most COCCON instruments at initial installation and during later maintenance and re-calibration. So, in principle, KIT would be able to keep track of the hardware status of the network.

- **Frequency:** Several years

6.4 Maintenance of instrument-specific software

Besides hardware upgrades, the instrument firmware is also updated by the manufacturer from time to time. Usually, Bruker provides firmware upgrades to its customers. However, newer firmware versions may also require the latest instrument electronics hardware version. So, older instruments cannot follow all firmware updates unless they are also able to update their hardware components.

- **Implementation by TCCON**
  - No special procedures for firmware upgrades. Usually, instrument firmware is updated during major maintenance acitivities and should be logged by the PI. There is no central register for the current firmware status of each instrument.
- **Implementation by COCCON**
  - So far, all EM27/SUN instruments share the same hardware and are all able to upgrade to the latest firmware. This homogeneity might change over years.
- **Frequency:** Several years

6.5 Maintenance of instrument configuration

Instrument configuration is typically set as a list of parameters for each measurement. Most of the configuration parameters are also stored in the Level 0 data files, so they are logged automatically. The principal measurement configuration only changes rarely, typically after principal discussions at the annual meetings.

- **Implementation by TCCON**
  - Configuration stored in Level 0 files for every single measurement.
- **Implementation by COCCON**
  - Same as TCCON.
- **Frequency:** Every measurement

6.6 ILS monitoring

The ILS is a very good indicator for both the quality of the optical alignment as well as the stability of the instrument. The ILS is very sensitive even to small changes of the instrument. Therefore, it is a good indicator for all kinds of sudden as well as gradual changes.

ILS can be monitored by observing the spectra of very narrow spectral lines in low-pressure gas cells. The partial pressure of the calibration gas inside the cell as well as the temperature of the cell have to be known as accurately as possible. The cells also have to be stable over many years, which has to be verified separately.

A cell measurement needs a much longer integration time than a solar measurement and a stable artificial reference light source.

- **Implementation by TCCON**
  - All TCCON sites use reference gas cells filled with HCl at low pressure. The stability of the cells has been checked every few years at KIT.
  - The IFS125HR instruments have several built-in reference light sources which can be used for the cell measurements. At many sites, the cell remains in the beam path all the time, so the cell measurements could at least be done automatically. At other sites, the HCl cell has to be put into the beam manually and removed afterwards for routine measurements.
- **Implementation by COCCON**

  - The EM27/SUN does not have enough space inside the instrument for a reference cell, nor does it have a suitable internal reference light source. Cell measurements are possible but have to be set up in the lab.
  - Instead of .cell measurements, open path measurements with H_2O are used.
- **Frequency:** ILS measurements should be done at least monthly. However, at many remote sites, they can only be done during maintenance visits.

6.7 Maintenance of Level 1 and 2 data processing software

The Level-1 and -2 processing software is the most critical component for the standard data product of both networks. The details are described in Sec. 4.

- **Implementation by TCCON**
  - Software is maintained by a team of TCCON PIs and the Algorithm Chairs.
  - Revision control is employed throughout the whole development process. All changes are documented.
  - Software is distributed to the PIs from a central software repository. Benchmark data sets are provided to verify the network-wide consistency of the results.
- **Implementation by COCCON**
  - Software is developed and maintained centrally at KIT. For the future, employment of a formal revision control system is planned.
  - KIT will provide central processing as a service in the future.
- **Frequency:** Major versions are typically stable for several years.

6.8 Monitoring Xair consistency between sites

The oxygen concentration in the atmosphere is very constant over altitudes up to 100 km and all regions. Therefore, Xair, the column-averaged dry-air mole fraction of oxygen divided by 0.2095 should be an ideal calibration reference for column measurements to check site-to-site consistency. However, the usefulness of Xair for this purpose has so far been limited by oxygen spectroscopy. This may change with newer spectroscopic line models that are implemented into the Level-2 processing software.

- **Implementation by TCCON**
  - Xair is determined routinely and part of the QC procedure.
- **Implementation by COCCON**
  - Xair is determined routinely and part of the QC procedure.
- **Frequency:** Every measurement

6.9 Level 2 calibration with airborne in-situ profiles

Aircraft and aircore calibrations are used for calibration of TCCON and COCCON to the WMO in-situ GHG scale. Aircraft observations have to be done by jet aircraft that are able to reach the top of the troposphere and beyond. Such observations are expensive and not possible at all sites due to ATC regulations.

Aircore measurements from meteorological balloons reach up into the stratosphere. However, due to the weight and size of the payload, these balloon launches are only possible in certain regions with limited air traffic. The payloads have to be recovered for analysis, so island sites and sites in densely settled areas are difficult for aircore calibration.

- **Implementation by TCCON**

  - All sites need at least one aircraft or aircore profile measurement to reach full TCCON status.
  - Over the past years, many TCCON sites had aircraft or aircore overpasses. All the results are included in the in-situ calibration of the most recent Level 2 software.
- **Implementation by COCCON**
  - COCCON instruments have routinely taken part in several campaigns with aircore or aircraft calibration flights.
- **Frequency:** Irregular due to high cost and difficult logistics

6.10 Travelling reference standard

A travelling standard would be a dedicated instrument that is moved between sites to check site-to-site consistency. Shipping a TCCON instrument in a standard ISO container is possible and has been done before. However, it is a rather expensive solution and takes 2-3 weeks to set up at a new site. Also, there are no spare instruments available. Instead, an EM27/SUN instrument can be calibrated against a running TCCON site and then sent to another site. Due to the high stability, the calibration of the EM27/SUN should remain stable during transport. The EM27/SUN is also easy to set up and operate during such a campaign. However, intercomparisons can only be done at the lower resolution of the EM27/SUN. High-resolution effects in the spectra cannot be compared.

- **Implementation by TCCON**
  - Currently, there is no travelling standard for TCCON. Individual site visits with EM27/SUN have been performed in the past [HED2017].
  - In the planning for TCCON integration into ICOS, a travelling standard EM27/SUN is included.
- **Implementation by COCCON**
  - Usually, instruments are sent back to KIT for recalibration, so a travelling standard is not strictly necessary.
- **Frequency:** With the current size of TCCON and a single travelling standard, it would take 6-8 years to visit all sites − assuming that a new site can be visited every 3 months. This might not be possible for remote sites where the shipping can take longer than 3 months. So, realistically, several travelling standards would be needed.

7 PRODUCT QUALITY ASSURANCE

TCCON and COCCON are scientific networks. The data quality of their products is their top priority. QA procedures have been established by the respective communities. They consist of best practices for the setup and maintenance of the instruments, operational procedures, use of a common revision-controlled and documented software, automatic and manual QC checks and the continuous discussion and exchange of ideas on how the data quality can be improved.

7.1 Procedures

The existing QA procedures are mostly best practices that the community has agreed on and that every PI understands. Formal QA documents are uncommon. External QA standards which are common in business applications, are not used at all.

In the case of TCCON, the internal Wiki contains most of the documented procedures for standard procedures. The key procedures are described in short articles, graphics, code and video sequences. COCCON is building up a similar platform. Dedicated workshops and annual community meeting are crucial for distributing knowledge about the best practices in the community.

Examples for already established QA procedures are

- a set of recommended and required hardware components and auxiliary devices.
- instructions for achieving and maintaining optimal optical alignment of the FTIR instruments.
- the regular use of reference gas cells to monitor the quality of the alignment over time.
- well-documented and revision-controlled development of the data analysis software.
- operation of the instruments with well-defined settings.
- use of the community-developed data analysis software.
- automatic QC procedure.
- manual QC before submission to the data archive.

Some of these procedures for TCCON have been documented in the course of an ESA study in 2017.

7.2 Quality assessment, availability and monitoring

To meet the overall requirement set for the operational service of the network, as delivered to CO2M cal/val and monitoring, it is expected that more formal QA procedures will have to be established. Also, several aspects will have to be monitored and reported that are currently handled at the discretion of the individual PIs.

It is expected that the following tasks will have to be handled and reported.

- Availability of data vs downtime of service: current operational status for TCCON stations reported on web site, past statistics for individual stations currently not available.
- Successful processing within expected quality limits: handled by data archive.
- Other reporting on infrastructure affecting the service: currently not reported.
- Degradation monitoring: handled at the discretion of the PIs.
- Maintenance downtime: currently not reported.

8 REQUIRED NETWORK UPGRADES: TOWARDS OPERATIONAL CO2M CAL/VAL AND MONITORING

There is a gap between the cal/val needs of CO2M and what the current TCCON and COCCON networks already provide. Table 8 summarizes key aspects of an operational cal/val system: what CO2M will likely need, what the current networks TCCON and COCCON can provide and recommendations for addressing the gap from EUMETSAT's point of view. The list is not necessarily complete.

Gaps do not only exist because something is not provided by the networks but also because not all needs for CO2M have been well defined yet. The situation is especially difficult for cases where the effort and cost for the required updates of the network are high: e.g. installation of new stations or moving to an NRT or RT data delivery scheme. Before such steps are taken, it has to be clear that the costs are well justified.

Table 8: Key aspects of a future operational CO2M cal/val network and recommendations for its realization

Key aspect	CO2M cal/val needs	Provided by TCCON/COCCON	Recommendations
Network design	Current and future station locations should be optimized for CO2M cal/val.	Station locations are a compromise of scientific interest, opportunity and logistical constraints.	Develop pre-launch OSSE as a tool for assessment of current status and future extensions of the network.
International collaboration	CO2M needs global data for cal/val.	Networks are global but non-European stations cannot be funded through European instruments.	Provide services and infrastructure that are also attractive for non-European stations.
Geographic coverage	All continents and regions.	North America, Europe, East Asia, Australia covered. Africa, South America, Indian subcontinent, central Asia missing	Support setup of stations in uncovered regions.
Latitude coverage	All latitudes.	80 °N to 45 °S, limited coverage of Southern hemisphere.	Support setup of additional stations on Southern hemisphere.
Parameter space coverage	Full albedo range, aerosol, clouds, precipitation, ...	Only partial coverage of albedo range. Coverage of other parameters has to be assessed	Use CO2M OSSE to investigate optimal locations for stations
Precision and accuracy	XCO_2 precision: 0.7 ppm XCO_2 systematic error: <0.5 ppm	TCCON: XCO_2: 0.8 ppm precision and accuracy, inter-station bias <1 ppm. TBD for GGG2020.	Support activities to improve network precision, accuracy and inter-station bias.

Key aspect	CO2M cal/val needs	Provided by TCCON/COCCON	Recommendations
Data levels	EUMETSAT would prefer access to all data levels (Level 0 to 2).	Networks typically only publish Level 2 data. Lower levels are not covered by data licenses.	Define needs for lower level data. Negotiate conditions of data use with PIs to avoid data ownership issues.
Column data products	Standard network data products as well as custom products (e.g. alternative meteo or spectroscopy).	Standard data product only.	Set up/support local or central infrastructure for data processing.
Profile data products	Need for CO2M has to be assessed.	Technically possible for TCCON but currently not available.	Assess usefulness for CO2M cal/val and support implementation if needed.
Data product with complete QC	Needed for final and a posteriori validation.	Final product delivery (<12 months) on a routine basis. Delivery on time scales <3 months currently not available.	Set up/support infrastructure for data delivery within 1 month.
NRT and RT data delivery	Need for CO2M has to be assessed.	Currently not available.	Assess usefulness for CO2M cal/val and set up/support implementation of data processing infrastructure if needed.
Emission monitoring	Key proficiency for CO2M	Most stations are located far from emission sources.	Set up networks of COCCON instruments around reference sources.
Operational level	Networks should be fully operational.	Scientific networks only. Operational level of stations depends on availability of resources.	Find solution together with Copernicus and ICOS member states.
Long-term availability	Networks should stay operational throughout CO2M mission.	Not guaranteed.	Find solution together with Copernicus and ICOS member states.

The most useful tool that could be developed would be a pre-launch OSSE for the CO2M mission. Ideally, it should have the following capabilities:

- Simulate critical aspects of the CO2M sensor that are relevant to cal/val: spectroscopy, sensor calibration, artifacts and degradation.
- Simulate orbit, ground (albedo) and atmospheric parameters (clouds, aerosol) that affect the availability of ground station overpasses.
- Provide atmospheric CO_2 maps and profiles from inventories and atmospheric transport.

- Simulate the properties of existing and potential ground based stations (as in Tab. 4) and run a retrieval with the original TCCON or COCCON code.
- Evaluate the effect of different QC strategies (e.g. final product vs. NRT/RT products) on the CO2M results.
- If possible, make the system user-friendly enough that the network PIs can run their own simulations.

Even if a full-scale integrated OSSE cannot be provided, individual solutions for the tasks listed above would be very useful. For example, there is currently no network design tool for the column networks. Even a footprint simulation tool for existing and potential stations would already be a big help.

Describing the necessary steps and upgrades in more detail is beyond the scope of this report. The evaluation of the recommendations in Tab. 8 should be addressed in a follow-up study.

Appendix A DATA PROPERTIES

A.1 Data sizes (maximum per day)

TCCON:

- Level 0: 4-8 GB (depending on measurement schedule and length of local day)
- Level 1: 1-2 GB
- Level 2: 1.5-3 MB (engineering output)

COCCON:

- Data sizes are similar. The effect of the lower spectral resolution is compensated by the higher number of measurements over time.

A.2 Data formats and granularity

TCCON data files are in netCDF format. There are two versions:

- engineering output files which contain more than 1500 parameters
- public data files which only contain a subset of parameters that is considered to be useful for external users.

In the netCDF files, each measurement is represented by one row.

This is an example of the TCCON public data format for GGG2014. The future GGG2020 format should follow CF metadata conventions:

netcdf ae20120522_20161221.public {

```
dimensions:
     a32 = 32 ;
     a4 = 4 ;
     a21 = 21 ;
     ak_P_hPa = 71 ;
     ak_zenith = 16 ;
     time = 41688 ;
     prior_date = 750 ;
     prior_Height = 71 ;
variables:
     float zobs_km(time) ;
          zobs_km:description = "Geometric Altitude (km)" ;
          zobs_km:units = "km" ;
          zobs_km:missing_value = -1.e+38f ;
          zobs_km:name = "zobs_km" ;
     float zmin_km(time) ;
          zmin_km:description = "Pressure Altitude (km)" ;
          zmin_km:units = "km" ;
          zmin_km:missing_value = -1.e+38f ;
          zmin_km:name = "zmin_km" ;
     int year(time) ;
          year:missing_value = -32768 ;
          year:description = "Year (e.g. (2009))" ;
          year:name = "year" ;
```

```
float xn2o_ppb_error(time) ;
     xn2o_ppb_error:description = "one-sigma precision" ;
     xn2o_ppb_error:units = "ppb" ;
     xn2o_ppb_error:missing_value = -1.e+38f ;
     xn2o_ppb_error:name = "xn2o_ppb_error" ;
float xn2o_ppb(time) ;
     xn2o_ppb:description = "0.2095*column_n2o/column_o2" ;
     xn2o_ppb:units = "ppb" ;
     xn2o_ppb:missing_value = -1.e+38f ;
     xn2o_ppb:name = "xn2o_ppb" ;
float xhf_ppt_error(time) ;
     xhf_ppt_error:description = "one-sigma precision" ;
     xhf_ppt_error:units = "ppt" ;
     xhf_ppt_error:missing_value = -1.e+38f ;
     xhf_ppt_error:name = "xhf_ppt_error" ;
float xhf_ppt(time) ;
     xhf_ppt:description = "0.2095*column_hf/column_o2" ;
     xhf_ppt:units = "ppt" ;
     xhf_ppt:missing_value = -1.e+38f ;
     xhf_ppt:name = "xhf_ppt" ;
float xhdo_ppm_error(time) ;
     xhdo_ppm_error:description = "one-sigma precision" ;
     xhdo_ppm_error:units = "ppm" ;
     xhdo_ppm_error:missing_value = -1.e+38f ;
     xhdo_ppm_error:name = "xhdo_ppm_error" ;
float xhdo_ppm(time) ;
     xhdo_ppm:description = "0.2095*column_hdo/column_o2" ;
     xhdo_ppm:units = "ppm" ;
     xhdo_ppm:missing_value = -1.e+38f ;
     xhdo_ppm:name = "xhdo_ppm" ;
float xh2o_ppm_error(time) ;
     xh2o_ppm_error:description = "one-sigma precision" ;
     xh2o_ppm_error:units = "ppm" ;
     xh2o_ppm_error:missing_value = -1.e+38f ;
     xh2o_ppm_error:name = "xh2o_ppm_error" ;
float xh2o_ppm(time) ;
     xh2o_ppm:description = "0.2095*column_h2o/column_o2" ;
     xh2o_ppm:units = "ppm" ;
     xh2o_ppm:missing_value = -1.e+38f ;
     xh2o_ppm:name = "xh2o_ppm" ;
float xco_ppb_error(time) ;
     xco_ppb_error:description = "one-sigma precision" ;
     xco_ppb_error:units = "ppb" ;
     xco_ppb_error:missing_value = -1.e+38f ;
     xco_ppb_error:name = "xco_ppb_error" ;
float xco_ppb(time) ;
     xco_ppb:description = "0.2095*column_co/column_o2" ;
     xco_ppb:units = "ppb" ;
```

```
        xco_ppb:missing_value = -1.e+38f ;
        xco_ppb:name = "xco_ppb" ;
    float xco2_ppm_error(time) ;
        xco2_ppm_error:description = "one-sigma precision" ;
        xco2_ppm_error:units = "ppm" ;
        xco2_ppm_error:missing_value = -1.e+38f ;
        xco2_ppm_error:name = "xco2_ppm_error" ;
    float xco2_ppm(time) ;
        xco2_ppm:description = "0.2095*column_co2/column_o2" ;
        xco2_ppm:units = "ppm" ;
        xco2_ppm:missing_value = -1.e+38f ;
        xco2_ppm:name = "xco2_ppm" ;
    float xch4_ppm_error(time) ;
        xch4_ppm_error:description = "one-sigma precision" ;
        xch4_ppm_error:units = "ppm" ;
        xch4_ppm_error:missing_value = -1.e+38f ;
        xch4_ppm_error:name = "xch4_ppm_error" ;
    float xch4_ppm(time) ;
        xch4_ppm:description = "0.2095*column_ch4/column_o2" ;
        xch4_ppm:units = "ppm" ;
        xch4_ppm:missing_value = -1.e+38f ;
        xch4_ppm:name = "xch4_ppm" ;
    float wspd_m_s(time) ;
        wspd_m_s:description = "Wind Speed (m/s)" ;
        wspd_m_s:units = "m/s" ;
        wspd_m_s:missing_value = -1.e+38f ;
        wspd_m_s:name = "wspd_m/s" ;
    float wdir_deg(time) ;
        wdir_deg:description = "Wind Direction (deg.)" ;
        wdir_deg:units = "deg" ;
        wdir_deg:missing_value = -1.e+38f ;
        wdir_deg:name = "wdir_deg" ;
    float tout_C(time) ;
        tout_C:description = "External Temperature (C)" ;
        tout_C:units = "C" ;
        tout_C:missing_value = -1.e+38f ;
        tout_C:name = "tout_C" ;
    double time(time) ;
        time:units = "days" ;
        time:description = "Fractional days since 1970/1/1 00:00:00" ;
        time:name = "time" ;
    float sia_AU(time) ;
        sia_AU:description = "Solar Intensity (Average)" ;
        sia_AU:units = "AU" ;
        sia_AU:missing_value = -1.e+38f ;
        sia_AU:name = "sia_AU" ;
    float pout_hPa(time) ;
        pout_hPa:description = "External Surface Pressure (hPa)" ;
```

```
        pout_hPa:units = "hPa" ;
        pout_hPa:missing_value = -1.e+38f ;
        pout_hPa:name = "pout_hPa" ;
    float long_deg(time) ;
        long_deg:description = "Longitude (deg.)" ;
        long_deg:units = "deg" ;
        long_deg:missing_value = -1.e+38f ;
        long_deg:name = "long_deg" ;
    float lat_deg(time) ;
        lat_deg:description = "Latitude (deg.)" ;
        lat_deg:units = "deg" ;
        lat_deg:missing_value = -1.e+38f ;
        lat_deg:name = "lat_deg" ;
    float hout_RH(time) ;
        hout_RH:description = "External Humidity (%)" ;
        hout_RH:units = "%RH" ;
        hout_RH:missing_value = -1.e+38f ;
        hout_RH:name = "hout_%RH" ;
    float hour(time) ;
        hour:missing_value = -1.e+38f ;
        hour:description = "Fractional UT Hour" ;
        hour:name = "hour" ;
    float fvsi(time) ;
        fvsi:description = "Fractional Variation in Solar Intensity (%)"
;
        fvsi:units = "%" ;
        fvsi:missing_value = -1.e+38f ;
        fvsi:name = "fvsi_%" ;
    int day(time) ;
        day:missing_value = -32768 ;
        day:description = "Day of the year (1-366)" ;
        day:name = "day" ;
    float azim_deg(time) ;
        azim_deg:description = "Solar Azimuth Angle (deg)" ;
        azim_deg:units = "deg" ;
        azim_deg:missing_value = -1.e+38f ;
        azim_deg:name = "azim_deg" ;
    float asza_deg(time) ;
        asza_deg:description = "Solar Zenith Angle (deg)" ;
        asza_deg:units = "deg" ;
        asza_deg:missing_value = -1.e+38f ;
        asza_deg:name = "asza_deg" ;
    char GSETUP_Version(time, a4) ;
        GSETUP_Version:description = "GSETUP Version Number" ;
        GSETUP_Version:name = "GSETUP_Version" ;
    char GFIT_Version(time, a4) ;
        GFIT_Version:description = "GFIT Version Number" ;
        GFIT_Version:name = "GFIT_Version" ;
```

```
    int prior_date_index(time) ;
        prior_date_index:units = "Index starting at 0" ;
        prior_date_index:missing_value = -32768 ;
        prior_date_index:description = "Index of date in main array" ;
        prior_date_index:name = "prior_date_index" ;
    double prior_date(prior_date) ;
        prior_date:units = "Fractional days since 1970/1/1 00:00:00" ;
        prior_date:name = "prior_date" ;
        prior_date:description = "Date of the a priori profile" ;
    double prior_Height(prior_Height) ;
        prior_Height:units = "km" ;
        prior_Height:name = "prior_Height" ;
        prior_Height:description = "A priori altitude profile" ;
    int prior_year(prior_date) ;
        prior_year:name = "prior_year" ;
        prior_year:description = "Year of the a priori profile (e.g.
2009)" ;
    double prior_n2o(prior_date, prior_Height) ;
        prior_n2o:units = "ppb" ;
        prior_n2o:name = "prior_n2o" ;
        prior_n2o:description = "A priori N2O profile" ;
    int prior_month(prior_date) ;
        prior_month:name = "prior_month" ;
        prior_month:description = "Month of the a priori profile (e.g.
March is 03)" ;
    double prior_hf(prior_date, prior_Height) ;
        prior_hf:units = "ppt" ;
        prior_hf:name = "prior_hf" ;
        prior_hf:description = "A priori HF profile" ;
    double prior_hdo(prior_date, prior_Height) ;
        prior_hdo:units = "parts" ;
        prior_hdo:name = "prior_hdo" ;
        prior_hdo:description = "A priori HDO profile" ;
    double prior_h2o(prior_date, prior_Height) ;
        prior_h2o:units = "parts" ;
        prior_h2o:name = "prior_h2o" ;
        prior_h2o:description = "A priori H2O profile" ;
    double prior_gravity(prior_date, prior_Height) ;
        prior_gravity:units = "m/s^2" ;
        prior_gravity:name = "prior_gravity" ;
        prior_gravity:description = "Gravitational acceleration" ;
    int prior_day(prior_date) ;
        prior_day:name = "prior_day" ;
        prior_day:description = "Day of the month of the a priori
profile" ;
    double prior_co2(prior_date, prior_Height) ;
        prior_co2:units = "ppm" ;
        prior_co2:name = "prior_co2" ;
        prior_co2:description = "A priori CO2 profile" ;
```

```
    double prior_co(prior_date, prior_Height) ;
        prior_co:units = "ppb" ;
        prior_co:name = "prior_co" ;
        prior_co:description = "A priori CO profile" ;
    double prior_ch4(prior_date, prior_Height) ;
        prior_ch4:units = "ppb" ;
        prior_ch4:name = "prior_ch4" ;
        prior_ch4:description = "A priori CH4 profile" ;
    double prior_Temp(prior_date, prior_Height) ;
        prior_Temp:units = "K" ;
        prior_Temp:name = "prior_Temp" ;
        prior_Temp:description = "A priori temperature profile" ;
    double prior_Pressure(prior_date, prior_Height) ;
        prior_Pressure:units = "hPa" ;
        prior_Pressure:name = "prior_Pressure" ;
        prior_Pressure:description = "A priori pressure profile" ;
    double prior_Density(prior_date, prior_Height) ;
        prior_Density:units = "molecules/cm^3" ;
        prior_Density:name = "prior_Density" ;
        prior_Density:description = "A priori density profile" ;
    double ak_zenith(ak_zenith) ;
        ak_zenith:units = "degrees" ;
        ak_zenith:name = "ak_zenith" ;
        ak_zenith:description = "Column averaging kernel solar zenith
angles" ;
    double ak_n2o(ak_P_hPa, ak_zenith) ;
        ak_n2o:name = "ak_n2o" ;
        ak_n2o:description = "N2O column averaging kernel profile" ;
    double ak_hf(ak_P_hPa, ak_zenith) ;
        ak_hf:name = "ak_hf" ;
        ak_hf:description = "HF column averaging kernel profile" ;
    double ak_hdo(ak_P_hPa, ak_zenith) ;
        ak_hdo:name = "ak_hdo" ;
        ak_hdo:description = "HDO column averaging kernel profile" ;
    double ak_h2o(ak_P_hPa, ak_zenith) ;
        ak_h2o:name = "ak_h2o" ;
        ak_h2o:description = "H2O column averaging kernel profile" ;
    double ak_co2(ak_P_hPa, ak_zenith) ;
        ak_co2:name = "ak_co2" ;
        ak_co2:description = "CO2 column averaging kernel profile" ;
    double ak_co(ak_P_hPa, ak_zenith) ;
        ak_co:name = "ak_co" ;
        ak_co:description = "CO column averaging kernel profile" ;
    double ak_ch4(ak_P_hPa, ak_zenith) ;
        ak_ch4:name = "ak_ch4" ;
        ak_ch4:description = "CH4 column averaging kernel profile" ;
    double ak_P_hPa(ak_P_hPa) ;
        ak_P_hPa:name = "ak_P_hPa" ;
```

```
        ak_P_hPa:description = "Column averaging kernel pressure" ;

// global attributes:
        :More_Information = "https://tccon-wiki.caltech.edu" ;
        :TCCON_Reference = "Wunch, D., G. C. Toon, J.-F. L. Blavier, R.
A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock,
and P. O. Wennberg (2011), The total carbon column observing network,
Philosophical Transactions of the Royal Society - Series A: Mathematical,
Physical and Engineering Sciences, 369(1943), 2087-2112,
doi:10.1098/rsta.2010.0240. Available from:
http://dx.doi.org/10.1098/rsta.2010.0240" ;
        :Data_Use_Policy = "https://tccon-
wiki.caltech.edu/Network_Policy/Data_Use_Policy" ;
        :Auxiliary_Data_Description = "https://tccon-
wiki.caltech.edu/Network_Policy/Data_Use_Policy/Auxiliary_Data" ;
        :Software_Version = "GGG2014" ;
        :id = "ae" ;
        :longName = "ascension01" ;
        :Contact = "Dietrich Feist <dfeist@bgc-jena.mpg.de>" ;
        :Location = "Ascension Island, Saint Helena, Ascension and
Tristan da Cunha" ;
        :Site_Reference = "Geibel, M. C., C. Gerbig, and D. G. Feist
(2010), A new fully automated FTIR system for total column measurements of
greenhouse gases, Atmospheric Measurement Techniques, 3(5), 1363-1375,
doi:10.5194/amt-3-1363-2010. Available from: http://www.atmos-meas-
tech.net/3/1363/2010/" ;
        :Data_DOI = "10.14291/tccon.ggg2014.ascension01.R0/1149285" ;
        :Data_Reference = "Feist, D. G., S. G. Arnold, N. John, M. C.
Geibel. 2014. TCCON data from Ascension Island, Saint Helena, Ascension and
Tristan da Cunha, Release GGG2014R0. TCCON data archive, hosted by the
Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory,
Oak Ridge, Tennessee, U.S.A.
http://dx.doi.org/10.14291/tccon.ggg2014.ascension01.R0/1149285" ;
        :Data_Revision = "R0" ;
        :creation_date = "2017/03/01" ;
        :start_date = "2012/05/22" ;
        :end_date = "2016/12/21" ;
}
```

A.3 Data transmission cost

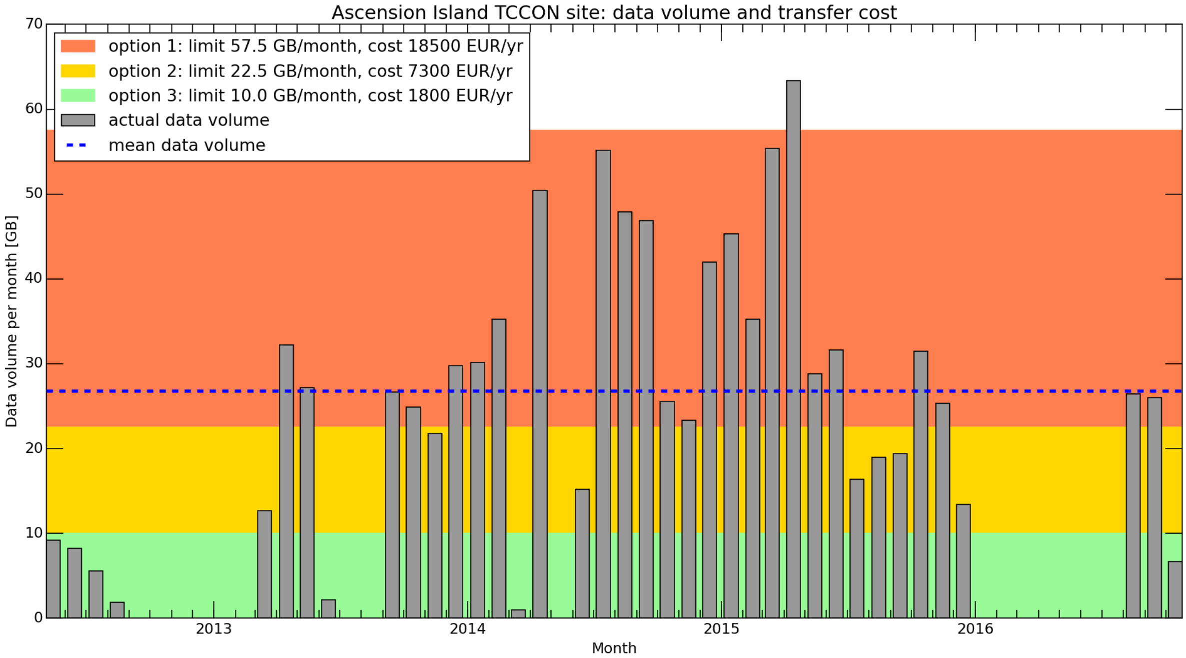

Figure A.1: Example for different data transmission cost options and actual transferred data volume for the Ascension Island TCCON station.