GAIA-CLIM Report / Deliverable D1.10

Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:

Report on the scientific assessment of gaps using a statistical approach based on heteroskedastic functional regression

A Horizon 2020 project - Grant agreement: 640276

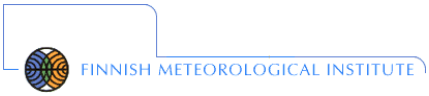
Date: 12 February 2018

Lead Beneficiary: University of Bergamo, Italy

Nature: R

Status: Final

Dissemination level: PU



Work Package	**WP1 - Geographical capabilities mapping**
Deliverable	D1.10
Title	Report on the scientific assessment of gaps using a statistical approach based on heteroskedastic functional regression
Nature	R
Dissemination	PU
Lead Beneficiary	University of Bergamo, Italy
Date	12 February 2018
Status	Final
Editors	Alessandro Fassò (UNIBG) and Viktoria Sofieva (FMI)
Contributors	Alessandro Fassò, Ilia Negri, Francesco Finazzi and Nil Venet (UNIBG), Fabio Madonna (CNR), Heather Lawrence (ECMWF), Viktoria Sofieva (FMI)
Reviewers	Peter Thorne (NUIM), Anna Mikalsen(NERSC)
Contacts	Alessandro.Fasso at unibg.it
URL	http://www.gaia-clim.eu/

This document has been produced in the context of the GAIA-CLIM project. The research leading to these results has received funding from the European Union's Horizon 2020 programme under grant agreement n° 640276.

Table of content

Executive summary

The quantitative assessment of geographical gaps in the current global non-satellite observing system is a key issue to improve our ability to use sub-orbital observations for characterising satellite data records of a number of atmospheric Essential Climate Variables (ECVs). An observational gap assessment needs to deal both with the measurement uncertainties associated with each individual measurement (investigated under GAIA-CLIM WP2), and those additional uncertainty contributions that appear when comparing different sounding techniques for the observation of the inhomogeneous and variable atmosphere, that is, when comparing data sets characterized with different sampling and smoothing properties, both in space and time (investigated under GAIA-CLIM WP3).

The present GAIA-CLIM deliverable, D1.10, describes a statistical approach for the identification of the geographical gaps, as well as the methods for analyses of small-scale variability of atmospheric trace gases. Specifically, Part 1 of D1.10 focuses on temperature and humidity vertical profiles provided by the RAOB radiosonde program. ERA-interim reanalysis has been used as the background information. Observation-Background (O-B) is calculated on 5-days RAOB averages (meteorological scale) with respect to the ERA-interim 90-day moving average. Consideration is limited to the RAOB mandatory levels, where a Gaussian process model is used to understand the O-B spatio-temporal variations, which are estimated via a kriging method. Based on this model, the standard deviation of the kriging estimator is used to define the network uncertainty map at each pressure level. Hence, the gap region is defined as the geographical area where the uncertainty exceeds the 85% percentile of the kriging residual variability distribution.

Next, Part 2 of D1.10 focuses on the analyses of small-scale natural variability field of atmospheric constituents (horizontal distances from tens to hundreds of kilometres). This information is of high importance for data validation, in order to define a time-space window, where the variations are small. At the same time, this information provides also the data representativeness, i.e., the spatial and temporal window, where a measurement represents the atmospheric state. In our report, we consider two atmospheric constituents: (i) stratospheric ozone profiles and (ii) methane total column. For stratospheric ozone, we estimate the small-scale variability using satellite measurements. For methane, we use the Carbon Tracker model. We indicate those geographical regions where the variability is high, and thus more frequent reference measurements are required in these regions for informative validation.

In Section 1, statistical methods for large spatio-temporal data sets are reviewed. Part 1 is dedicated to temperature and humidity measurements: in Section 2, the data from RAOB project and ERA-interim reanalysis are presented; Section 3 reviews the Gaussian Process approach used for RAOB data; Section 4 and 5 report the results and the identified gap regions for temperature and humidity; in Section 6, the conclusions for Part 1 are given. Part 2 considers ozone and methane: in Section 7, a variogram analysis is applied to stratospheric ozone profiles obtained by satellite instruments GOMOS and MIPAS; in Section 8, CH_4 total column data based on CarbonTracker Europe-CH_4 model outputs are analysed. Detailed conclusions are presented in each part, and the overall conclusions are presented in Section 9.

1 Introduction on statistical methods for geographical uncertainty

Recent decades have seen great improvements in statistical techniques related to the use of kriging applied to large datasets, such as those of interest in atmospheric sciences and global climate change. The idea of using kriging to study atmospheric ECVs is not new. For example, Sherwood (2000a and 2000b) introduced Iterated Universal Kriging (IUK) for time series homogenisation, and Sherwood et al. (2008) used an integrated approach for homogenisation and climate change detection. Homogenisation, consists essentially of change-point detection and adjustment of data for any kind of known and quantifiable inhomogeneities (bias, change of sensors, calibration drift, local environment changes etc.). Homogenisation methods developed for radiosondes have a long history, see for example Thorne et al. (2011). They may be divided into 'self-homogenisation', where changes are defined with respect to spatial and/or temporal trends, and 'background-based' where changes are defined using the difference of the observed time series respect to some background expectation time series (O-B). For example, Haimberger et al. (2012) use a comparison with both the reanalysis background and the neighbour data (RAOBCORE-RICH method). Although in this deliverable we are not focusing on homogenisation, we similarly make use of the ERA-interim background introduced in the next Section.

The term kriging refers generally to spatial interpolation or spatial forecasting. In particular, it is given by a conditional expectation formula which makes use of a spatial covariance function or structure function. Various spatial covariance models have been considered recently for large datasets.

In this document, we assume that $y(s)$ is a Gaussian Process (GP) with s in a geographical space, which may be Euclidean or a sphere or a spherical shell. A set of n georeferenced observations from a GP will be denoted by $Y=(y(s_1), \dots, y(s_n))$ and by Σ its variance-covariance matrix.

1.1 Fixed-Rank Kriging

Cressie and Johannesson (2008) consider fixed rank kriging, which is based on the following linear mixed effect model:

$$Y = X\beta + v + \varepsilon \quad (1)$$

Here Xβ is the fixed effect component, which includes a spatial regression. The random effect component v has a spatial covariance $C(s, s')$, which can be factorized as follows:

$$C(s, s') = S(s)'KS(s'). \quad (2)$$

Here K is an $r \times r$ matrix and S(s) is an r-dimensional basis functions vector. As a result, the kriging formula for the interpolated $\hat{y}(s)$ has a computational complexity which is $O(nr^2)$ instead of $O(n^3)$. This makes the computation considerably more tractable.

Using this approach, the same authors considered the Total Ozone Column (TOC) level 2 product of the Nimbus-7 polar orbiting NASA satellite. Specifically, they computed smoothed maps at the same resolution as the level-3 product, namely at 1°×1,25° resolution.

1.2 Spatial partitioning

In this approach, the spatial domain is divided into D sub-regions and data in different sub-regions are assumed conditionally independent. If sub-regions have the same number of elements, say r, then the variance-covariance matrix Σ is approximated by a block diagonal matrix with $r \times r$ dimensional blocks. As a result, likelihood and kriging computations have a $O(nr^2)$ complexity.

Various partitioning methods are available, see for example Sang et al. (2011), Knorr-Held and Raßer (2000), Kim et al. (2005) Anderson et al. (2014), Heaton et al. (2017), Konomi et al. 2014) and Neelon et al. (2014). The simplest approaches, based on sub-regions of equal areas and sub-regions with equal number of observations, perform well in terms of computational effort.

1.3 Nearest neighbour process

Suppose that Y is an n-dimensional spatial observation of a GP as in Equation (1). Then the related Nearest Neighbour Gaussian Process (NNGP) is derived by factorizing the conditional likelihoods of

$$y(s_i)|Y_{i:r}$$

where $Y_{i:r}$ is the set of the r nearest neighbors of $y(s_i)$ with r usually quite small, see e.g. Datta et al (2016).

It turns out that the kriging computation has the form

$$\hat{y}(s) = A(s)Y_{i:r}$$

and the computational complexity is $O(r^3)$. Similarly, the likelihood computation has complexity $O(nr^2)$.

1.4 Covariance tapering

The approach used here is based on the idea to approximate the dense covariance matrix Σ by a sparse matrix obtained by zeroing the small values of Σ. In doing this positive definiteness of the resulting covariance must be ensured. The solution (e.g. Bevilaqua et al., 2016) consists in Schur multiplication of Σ with a matrix K, namely $\Sigma \otimes \mathrm{K}$, where K is obtained by a finite support correlation function. If $\Sigma \otimes \mathrm{K}$ is sparse, then computational complexity is reduced. Nonetheless, either model estimation or kriging may be biased, especially for sparse networks.

PART 1: TEMPERATURE AND HUMIDITY

2 Temperature and humidity global data

In this section, we introduce the data used for studying geographic gaps: the RAOB radiosonde temperature and humidity vertical profile, and the data used for the background: the ECMWF ERA-interim reanalysis data.

2.1 RAOB description

A radiosonde is a battery-powered telemetry instrument package carried into the atmosphere usually by a weather balloon that measures various atmospheric parameters and transmits them by radio to a ground receiver. The radiosonde contains instruments capable of making direct in-situ measurements of air temperature, humidity and pressure with height, typically up to altitudes of approximately 30-35 km, depending on several factors (including location and season).

Conventional radiosonde observations (RAOBs) have been used historically as a commonly *de facto* accepted comparison dataset in satellite measurement and derived product validation. Worldwide there are more than 800 operational radiosonde launch sites. Most countries share data with the rest of the world through international agreements. Nearly all routine radiosonde launches occur 45 minutes before the official observation time of 0000 UTC and 1200 UTC, and more rarely at 6 or 18 UTC. There are also several launches that occur on demand to monitor specific events or in campaigns.

The radiosonde transmits temperature and relative humidity data at each pressure level. The altitudes of these levels are calculated using the Global Positioning System (GPS). Prior to the advent of GPS, and still to this day in some older models, this is done using the hypsometric equation that relates the vertical height of a layer from the mean layer temperature, the humidity of the layer and the air pressure at top and bottom of the layer. Significant levels are reported where the vertical profiles of the temperature or the dew point undergo a change determined by the observer from the sounding. The height of the troposphere and stability indices are calculated.

RAOB observations are processed, tabulated and encoded for transmission over communication networks. While the radiosonde transmits an essentially continuous stream of temperature and humidity information back to the station (each 5-10 m of altitude, measured each 1-2 s), traditionally only subsets of this information have been able to be encoded and transmitted. The transmitted RAOB data include temperature, humidity, and wind profiles at 22 mandatory levels (including surface) and various significant levels:

- **MANDATORY LEVELS:** By international convention, the following 22 specific pressure levels must be reported in the RAOB message (if attained): the surface, 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1 hPa. Note that many sites do not routinely measure to the highest levels owing to balloon burst.

- **SIGNIFICANT LEVELS:** In addition to the mandatory levels, pressure levels at ***significant*** or abrupt changes and extrema in the vertical temperature and/or dew point temperature profiles are included as part of the RAOB message. In a radiosonde observation, a significant level is a level other than a mandatory level, for which values of pressure, temperature, and humidity are reported either because they are sufficiently important or unusual to warrant the attention of the forecaster, or these values are required for the reasonably accurate reproduction of the radiosonde observation. There are defined rules governing the selection of significant levels, set forth in the Manual of Radiosonde Observations[1]. As a result, significant levels change among different profiles and, on average, there were 28 significant levels per profile in the data sets considered. By assuming that the temperature and dew point profiles change linearly with height between significant levels (*i.e.,* constant environmental lapse rate), a reasonably accurate reproduction of the RAOB sounding can be made from the sequence of RAOB message information at significant levels and supplemented by the mandatory levels. However, inspection of available archives suggests a strong trend towards a greater number of significant levels being reported. It is extremely unlikely that the atmosphere has become more vertically structured. Rather, it is likely that reporting of significant levels has become more widespread with a propensity to assign more such levels over time.

More recently, with the transmission to BUFR encoded messages, full profile data is being transmitted by an increasing number of sites. This should ameliorate issues with significant level reporting practices going forwards. Given issues with reporting significant levels and that only mandatory levels are constantly present in the current analysis subsequent work is limited to the mandatory levels only.

2.2 ERA-interim description

ERA-Interim is a global atmospheric reanalysis which is available from 1979 onwards and continuously updated in real time. The data assimilation system used to produce ERA-Interim is based on the year 2006 release of the Integrated Forecast System (IFS - Cy31r2). This system includes a 4-dimensional variational analysis (4D-Var) with a 12-hour analysis window (Dee et al., 2011). The spectral resolution is T255 (T255 spherical-harmonic representation of the basic dynamic fields), while the tentative horizontal resolution is ~80 km (reduced Gaussian grid N128); ~83km/0.75 ° when interpolated to a regular lat/lon grid.

Each single analysed model-output is described as instantaneous, though it represents an average over the model time step (30 minutes for ERA-Interim). Depending on the parameter, forecast data in ERA-Interim is either instantaneous or accumulated from the beginning of the forecast (twice daily forecasts starting at 00 and 12 UTC). Parameters such as precipitation and radiation are accumulated. The accumulated parameters in ERA-Interim are listed in Tables 9, 13-15 of the ERA-Interim archive document[2]. Analyses are available every 6 hours (0, 6, 12, 18 UTC) and the twice daily forecasts (from 00 and 12 UTC) provide output for surface and pressure level parameters, at forecast steps every 3

[1] WBAN Circular P, 7th ed. rev., 1957.

[2] http://www.ecmwf.int/en/elibrary/8174-era-interim-archive-version-20

hours to 24 hours, then with decreasing frequency to 10 days.

There are 60 vertical levels from the surface up to 0.1 hPa. The list of ERA-Interim fields (parameters) available for downloading can be accessed in the ECMWF data server. Note that while ERA-Interim "runs" in near-real-time, data is published with a few months delay. Note also that several quality issues are known with ERA-Interim data and ERA-Interim daily retrieval efficiency documentation gives indications for best practices in retrieving ERA-Interim data.

2.3 Datasets used

RAOB radiosonde data have been obtained from the NOAA/ESRL radiosonde database[3] and provided as NetCDF files containing one radiosonde launch per file. The analysis is based on data for the year 2015. The original RAOB dataset contained 503,735 soundings from 896 stations with regular WMO code. We focus on data at synoptic times 00:00 and 12:00 UTC. To this end, we kept only the two launches (whenever available) closest to these two synoptic times. Moreover, soundings with fewer than 10 measurements were also deleted. Although not all stations have two measurements per day, the average coverage is about 1.4 observations per day. Moreover, observations at 00:00 and 12:00 UTC are relatively balanced (about 47% vs 53%). Hence, we conclude that the analysis should not have serious day/night biases.

The analysis of humidity has been based on specific humidity in g/kg, which is routinely provided within the ERA-interim dataset in kg/kg. For the RAOB dataset, the specific humidity has been computed starting from the dew point depression and using the empirical formula provided in Bolton (1980).

The analysis of Sections 4 and 5 is based on 338,640 soundings related to 684 RAOB stations remaining after filtering (Figure 1). Considering data availability and quality, for temperature, H=12 mandatory levels between 925 and 50 hPa were used. Similarly, for humidity, H=6 mandatory levels between 925 and 300 hPa were included in the analysis (humidity sensors are known to have issues above circa 300hPa where the air is very dry). The ERA-interim data corresponding to these levels and with a horizontal resolution of 0.75°, as discussed above, have been obtained from the ECMWF data server[4]. Each level being populated by N_2=241×480 pixels. The resulting uncertainty and gap maps are based on N_2xH = 1,388,160 pixels each time step t, with t=1,…,T=730 half days in one year.

[3] https://ruc.noaa.gov/raobs/
[4] http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=pl/

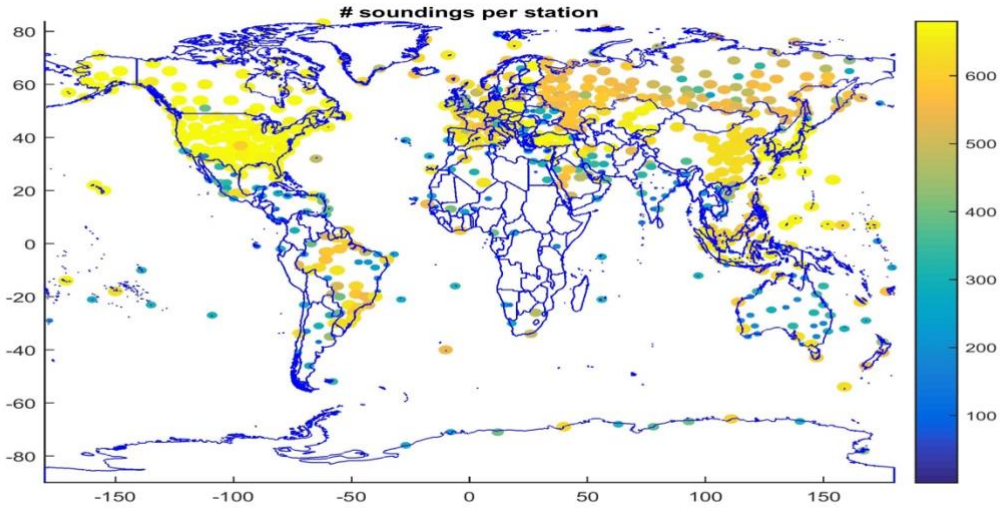


Figure 1 - Geographical distribution of the 684 RAOB stations used in the analysis. The colour represents the number of 00+12 UTC soundings available as per the colour bar.

3 Methods for temperature and humidity

This section gives some details about the statistical methods used in the gap analysis of temperature in Section 4 and humidity in Section 5. For the sake of simplicity, in this section, we refer only to temperature, but all methods considered here have been applied, in exactly the same manner, to humidity.

The approach based on comparing measurements with a background is used here. Usually this approach is referred to as

$$O - B, \tag{3}$$

Where *O* is the observations and *B* some background expectation field, because it is based on some statistics of this difference. In this study, the difference in (3) is not restricted to be a perfect white noise, nor auto-correlated in space and time, but, due to some co-location mismatches, it is admitted to have a spatio-temporal structure. To see this, consider the radiosonde reading of temperature at location station $s_i = (lat_i, lon_i)$, $i = 1,\ldots,N = 684$, time step $t = 1,\ldots,T = 729$, and pressure level p_j, $j = 1,\ldots, q = 12$, and denote them by $y_p(s,t)$. The spatio-temporal structure of $O - B$ is modelled generalising the approach of Calculli et al. (2015) and Finazzi and Fassò (2014) to large datasets.

Let us consider the statistical model given by

$$y_p(s,t) = x_p(s,t)'\beta_p + GB_p(s,t) + \varepsilon_p(s,t). \tag{4}$$

The vector x may include the ERA-interim background and other meteorological and local variables, such as humidity, wind, latitude, longitude etc. We will comment on this in Sections 4 and 5.

The coefficient vector β is unknown and has to be estimated. In the original $O-B$ approach $\beta=1$, but we avoid this assumption here because, the spatial and temporal resolution of ERA-interim is smoothed to represented single point measurements (see D3.4 and application Sections below). For this reason, the Gaussian Process $GP_p(s,t)$ adjusts for the co-location mismatch and representativeness error. Hence, following Fassò et al. (2014), the above GP is modelled as:

$$GP_p(s,t)=\alpha_p z_p(s,t),$$

where $\alpha=(\alpha_1,\ldots,\alpha_q)$ is a set of scale coefficients to be estimated. The k-dimensional Gaussian Process $Z(s,t)=(z_1(s,t),\ldots,z_q(s,t))$ is defined over the sphere $\times$ time by the following independent Markovian models:

$$z_p(s,t)=g_p z_p(s,t-1)+\eta_p(s,t).$$

Here the innovations $\eta_p(s,t)$ are standardized Gaussian white noise processes with unit variance and independent in time. Moreover, the spatial structure is given by q co-regionalization spatial covariance functions satisfying:

$$E(\eta_p(s,t)\eta_p(s',t))=\rho(\|s-s'\|,\theta_p).$$

In the isotropic case, the spatial correlation function ρ is based on geodesic distance $\|s-s'\|$ and θ is the correlation range, to be estimated on the data. For example, in the application, we use the exponential model:[5]

$$\rho(h,\theta)=\exp\left(-\frac{\|s-s'\|}{\theta}\right).$$

The isotropy assumption may be questioned for atmospheric data. Hence, we consider also anisotropic alternatives in the class of axially symmetric covariance functions (Stein, 2007, and Ma, 2016). In particular, following Venet et al. (2017), we consider an exp-exp covariance function with a two-dimensional parameter $\theta=(\theta_1,\theta_2)$ and correlation function given by

$$\rho(s,s')=\exp\left(-\frac{\|s-s'\|}{\theta_1}\right)\exp\left(-\frac{\|lat-lat'\|}{\theta_2}\right),$$

which modulates the longitudinal correlation according to latitudinal displacement.

[55] With some abuse of notation, for simplicity, we use $\|s-s'\|$ also for the geodesic distance.

Finally, the last term of the model (4), $\varepsilon_p(s,t)$, is a Gaussian white noise process, uncorrelated in space and time, with zero mean and variance $\sigma^2_{\varepsilon,p}$.

3.1 Model estimation

In order to use the model (4) for gap analysis, the parameter set

$$\Psi = (\beta, \sigma^2_\varepsilon, \alpha, \theta)$$

needs to be estimated. In this study, the maximum likelihood estimates of Ψ are computed using an Expectation-Maximization (EM) algorithm, which extends Calculli et al. (2015).

In general, the computation of the likelihood function of a spatio-temporal model requires the inversion of $NT \times NT$ variance-covariance matrices. This has a computational complexity of $O(T^3N^3)$. In this study, thanks to separability of spatio-temporal covariance terms, the expectation step, needed both in model fitting and in kriging computation, requires a Kalman filtering and smoothing to be executed. Hence at each iteration of the latter, the inversion of T $N \times N$ matrices is required. As a result, the computational complexity of each E step is $O(TN^3)$ for full matrices, representing a substantial optimisation challenge for large problem sets. In contrast to Calculli et al. (2015), our dataset is quite large, and hence results in an unfeasible problem using standard computing facilities. Hence, in estimation of model (4), spatial partitioning of Section 1.2 is applied, reducing the complexity to $O(TNr_1^2)$ where r_1 is the size of the sub-regions. Note that spatial partitioning gives unbiased estimates of the model parameter set Ψ but may results is some discontinuities of the estimated temperature field $\hat{y}(t,s)$ at the sub-regions borders. To avoid, once the model is estimated, we use a nearest neighbour approach in kriging as described in the next section.

3.2 Kriging and geographic gaps

Model (4) allows radiosonde temperature to be estimated at an unobserved point (s_0, t_0) by the kriging formula

$$\hat{y}_p(s_0,t_0) = E(y_p(s_0,t_0) \mid Y)$$

where Y is the set of all available temperature measurements.

The kriging method gives also the uncertainty of this estimate, namely the estimation standard deviation:

$$uk_p(s_0,t_0) = Var(y_p(s_0,t_0) \mid Y)^{1/2}.$$

In this document, uk is named kriging uncertainty and gives information on the capability of the radiosonde network to estimate the difference $O-B$ in an un-sampled point (s_0,t_0). Hence, if the kriging estimate and uncertainty are computed on a fine grid for s_0, and for a time interval T_0, then the average kriging uncertainty in T_0 is $\bar{u}_p(s,T_0)=\frac{1}{|T_0|}\sum_{t\in T_0} uk_p(s,t)$.

Eventually, a geographic gap for time lapse T_0 and pressure p_0 is defined as the geographic region $G(T_0,p_0)$ where the kriging uncertainty exceeds a threshold λ :

$$G(T_0,p_0)=\{s\ :\ \bar{u}_p(s,T_0)>\lambda\}.$$

This kriging computation has to be repeated for all pixels s_0 in the fine grid given by ERA-interim which has $N_2=241\times 480=115,680$ such pixels and T time steps. Hence, the computing time is $N_2\times O(TN^3)$ for a full matrices approach, which is not feasible for practical use here.

Different from the estimation of the preceding subsection, kriging computation is implemented here using the nearest neighbour approach of Section 1.3. This reduces the computational complexity to $O(TNr_2^2)$ where r_2 is the number of neighbours used and avoids the possible artefact caused by spatial partitioning at borders between adjacent sub-regions.

4 Gap analysis for temperature

As explained in the executive summary, the capability of the RAOB network to understand the variations of the meteorological field with respect to seasonal variations is considered to assess the observational gaps using an entire year of data. Hence, in the following, RAOB data will denote 5-day averages of observed profiles and ERA-interim or seasonal trend will denote 90-day moving averages of ERA-interim reanalysis at 00:00 and 12:00 UTC.

4.1 Preliminary analysis and modelling

Firstly, outliers have been filtered out based upon the residuals of the linear regression of RAOB vs ERA-interim. Measurements have been excluded if the corresponding residual in absolute value exceeded the threshold given by 4.5 × residual standard deviation. Then, the spatial correlation has been analysed by means of the variogram (or structure function) approach. In Figure 2, the isotropic variogram is reported together with the directional variograms at selected pressure levels. As expected, a faster decorrelation is evident in the North-South direction than the East-West direction, although the scale changes with altitude. An exception is the layer around 200 hPa (just above the climatological tropopause), where reduced total variability, reduced spatial correlation and little North-South East-West difference is found. The 300- to 200-hPa stratum is also the layer in which the core of the jet stream is usually located. It is here that the cumulative effects of the mean temperature field of the troposphere produce the sharpest horizontal contrasts in the wind field.

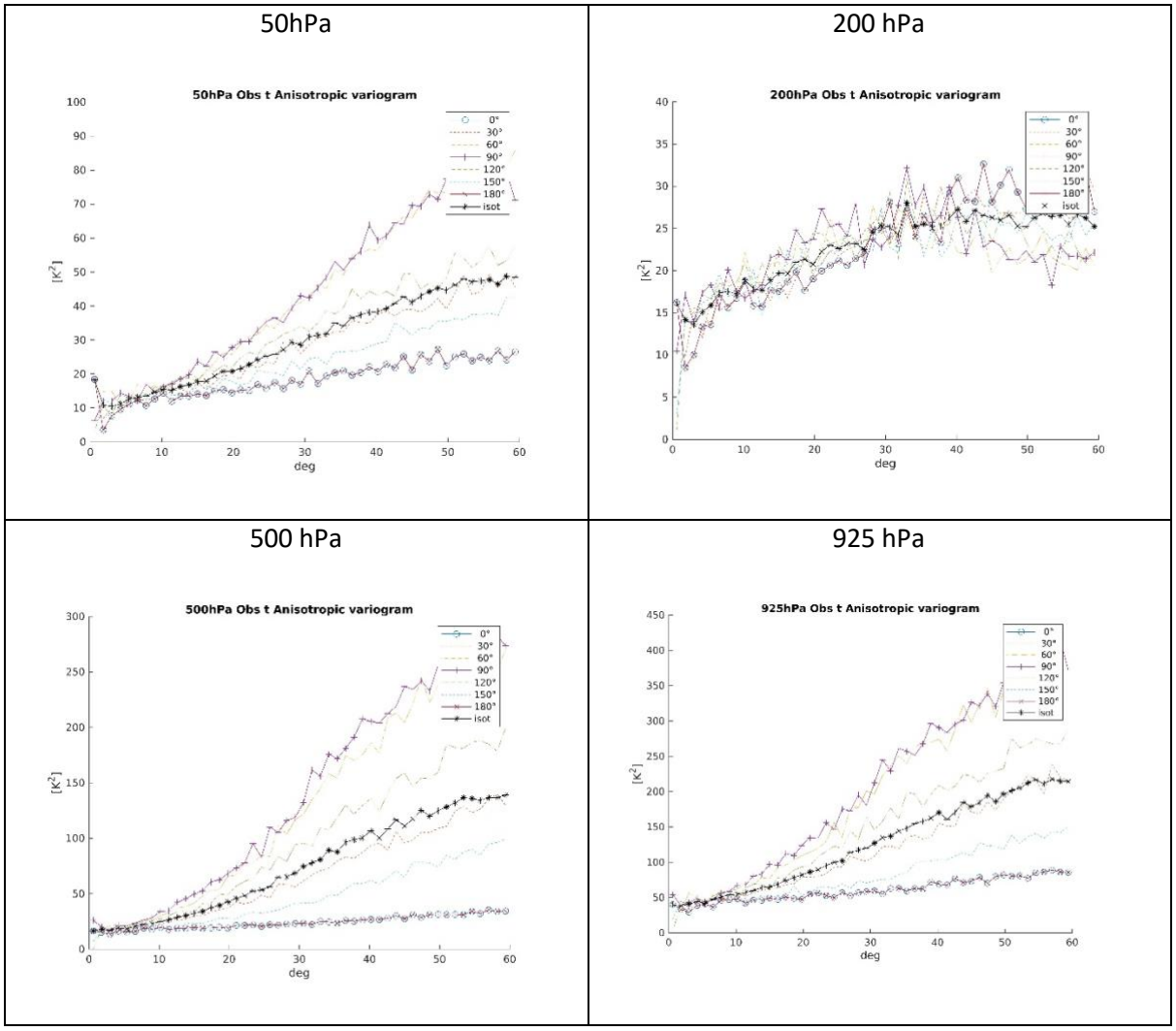


Figure 2 – Temperature analysis: anisotropic variograms of RAOB data at selected pressure levels. Isotropic variograms: Black solid line with stars; Longitudinal variogram: purple solid line with crosses and cyan circles; Latitudinal variograms: purple solid line with plus signs; Variograms in intermediate directions: dashed lines and dotted lines. Note that the vertical axis range varies in each panel.

Considering O-B, the spatial structure has a minimum around the tropopause as shown by the estimated model coefficients in Table 1. Overall, the spatial range coefficient (theta1) decreases as pressure increases from about 2200 km to about 1100 km. Note that, despite the fact that the latitudinal component (theta2) has a u-shaped behaviour, with faster decorrelation at mid-range altitudes, the resulting spatial correlations correspond to quite mild anisotropy in O-B. It may also be noted that at 200 hPa, in connection to the particular behaviour of Figure 2, model uncertainty (sigma2_eps) is larger and the temporal dynamics (g) is stronger. Also, the spatial range (theta1) has a drop. The coefficient for ERA-interim is very close to one, confirming the general good agreement between radiosonde and the reanalysis for temperature.

Table 1 - Temperature analysis: estimated global coefficients (mle) and standard deviations (std) of the GP model defined in Equation (4).

Pressure level		ERA-interim	Error	GP scale			spatial correlation	
			sigma2_eps	g	alpha	sigma2_Z	theta1	theta2
		[K/K]	[K^2]	[K/K]	[K/K]	[K^2]	[km]	[°]
50	mle	1,03	0,34	0,42	1,56	2,96	2172,1	128,8
	se	*0,011*	*0,005*	*0,005*	*0,004*		*16,5*	*4,9*
70	mle	1,01	0,38	0,41	1,72	3,55	1977,7	160,7
	se	*0,008*	*0,006*	*0,005*	*0,004*		*11,4*	*6,2*
100	mle	0,99	0,44	0,59	2,38	8,68	1294,5	56,0
	se	*0,011*	*0,015*	*0,006*	*0,003*		*4,8*	*0,9*
150	mle	1,00	0,33	0,33	2,01	4,53	1483,1	70,7
	se	*0,011*	*0,007*	*0,004*	*0,004*		*4,4*	*0,9*
200	mle	1,02	0,72	0,57	2,62	10,18	1308,8	27,9
	se	*0,026*	*0,020*	*0,005*	*0,003*		*3,9*	*0,2*
250	mle	1,02	1,57	0,85	1,38	6,93	1347,1	32,3
	se	*0,017*	*0,015*	*0,003*	*0,005*		*25,4*	*2,0*
300	mle	1,01	0,34	0,57	1,93	5,51	1241,6	27,1
	se	*0,009*	*0,011*	*0,006*	*0,003*		*6,7*	*0,4*
400	mle	1,00	0,23	0,25	1,83	3,58	1463,1	82,1
	se	*0,006*	*0,005*	*0,004*	*0,003*		*3,6*	*1,0*
500	mle	1,00	0,24	0,24	1,90	3,83	1427,0	114,7
	se	*0,006*	*0,005*	*0,003*	*0,003*		*3,2*	*1,6*
700	mle	1,01	0,22	0,25	1,94	4,01	1416,2	183,3
	se	*0,006*	*0,005*	*0,003*	*0,003*		*3,1*	*3,8*
850	mle	1,01	0,48	0,29	2,25	5,50	1274,2	490,3
	se	*0,006*	*0,010*	*0,004*	*0,004*		*2,7*	*31,5*
925	mle	1,01	0,75	0,40	2,35	6,56	1138,4	164,0
	se	*0,007*	*0,017*	*0,005*	*0,004*		*3,4*	*6,4*

The model has been tested in cross validation by removing 25% of the stations and using the corresponding data to check the spatial forecast capability. The results are given in Table 2, where the Root Mean Square Uncertainty (R-MSU) is about 1K. Note again that, around 200 hPa, the fitting (R^2=1-MSU/VAR) is lower and R-MSU is higher.

Table 2 - Temperature analysis: model and estimation performance. Number of iterations to convergence in EM algorithm with an exiting tolerance τ=0.001; fitting is given by Root Mean Squared Uncertainty (R-MSU) and R^2=1-MSU/VAR.

Pressure level	Iterations #	R^2 % of variance	R-MSU [K]
50	34	0.973	1.00
70	32	0.985	1.07
100	31	0.978	1.58
150	30	0.961	1.34
200	26	0.881	1.62
250	8	0.941	1.66
300	29	0.982	1.16
400	30	0.988	1.10
500	30	0.986	1.17
700	32	0.987	1.18
850	27	0.980	1.62
925	25	0.980	1.74

4.2 Mapping and gaps

Using the estimated model reported in Table 1 and the kriging formulas of Section 3, we get the RAOB temperature uncertainty maps as given in Figure 3. The maps are similar but not identical since the theta parameter is not the same at each level. More importantly, as detailed in Table 3, the scale is different at different pressure levels. In Figure 4, the gap regions are defined as those regions where there is highest uncertainty. In particular, we considered the 85% uncertainty percentile such that the identified gap region is the 15% of the global domain at the considered pressure level where the sampling related uncertainty resulting from the uneven distribution of RAOB stations is highest. The shape of the gap regions is similar at different pressure levels. Differences arise between standard reporting levels with respect to the size of Oceanic tropical gap, and a South-eastern African gap in the lower atmosphere. As shown in Table 3, the 85% threshold level is equivalent to a temperature uncertainty (level dependent) between 1.4 and 3K, with a maximum at 200 hPa, consistent with findings in Section 4.1.

Table 3 - Temperature analysis: Sampling (kriging) uncertainty statistics at pressure levels 50, 200, 500 and 925hPa.

pressure level	Sampling uncertainty [K]		
	min	Max	gap threshold (85%)
50	0.41	1.63	1.43
200	0.73	3.15	2.98
500	0.43	1.93	1.78
925	0.69	2.54	2.42

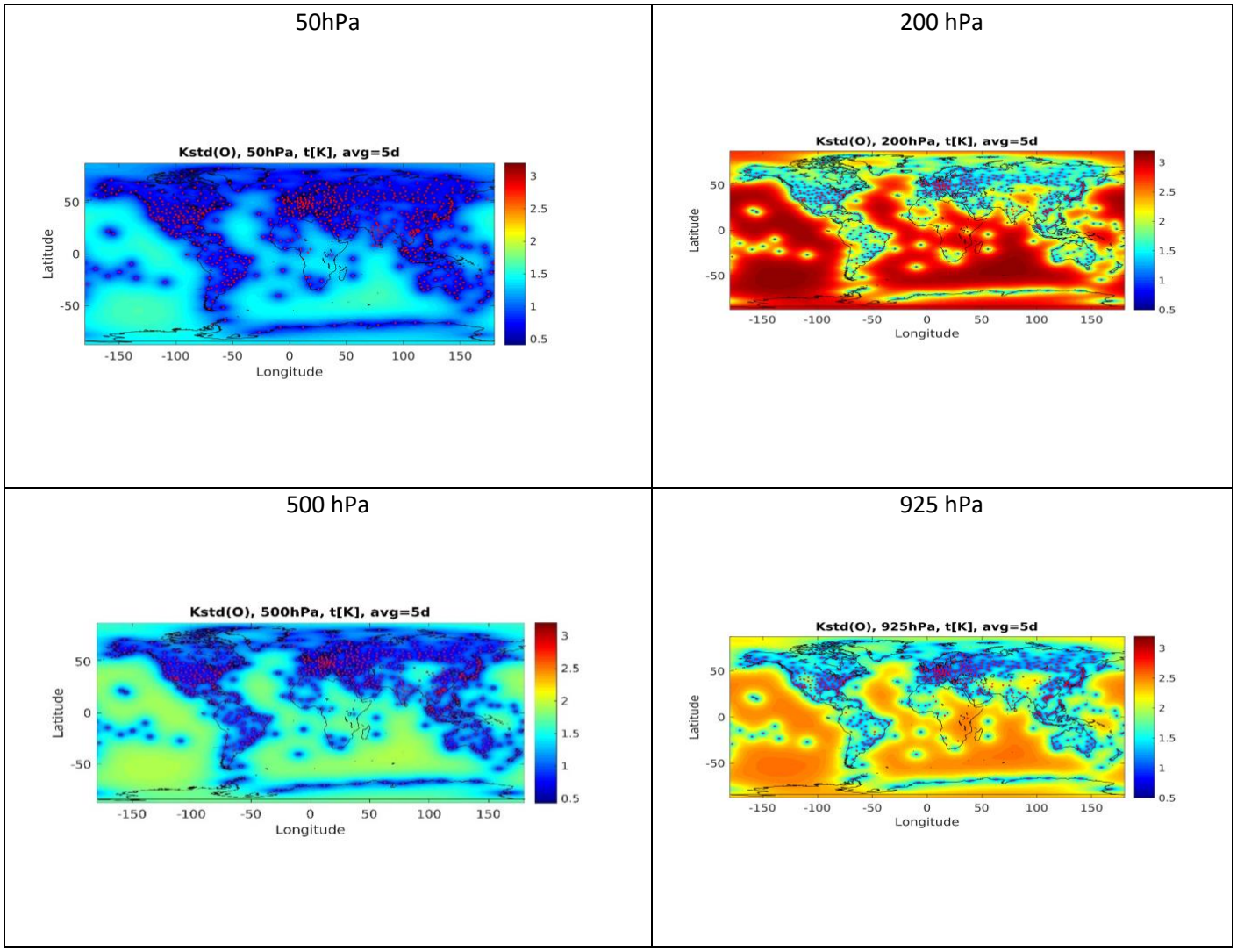


Figure 3 - Temperature analysis: kriging uncertainty at 50, 200, 500 and 925 hPa. The different scales of uncertainty are reported in Table 3.

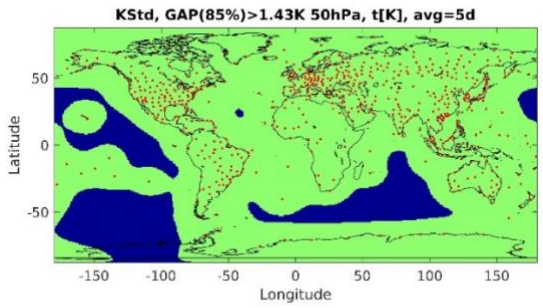
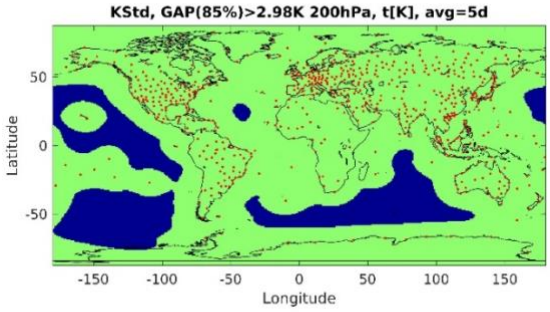
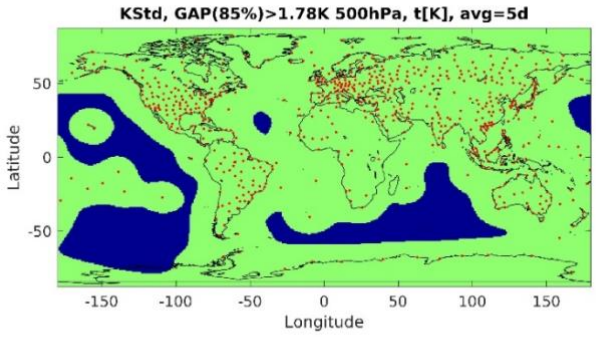
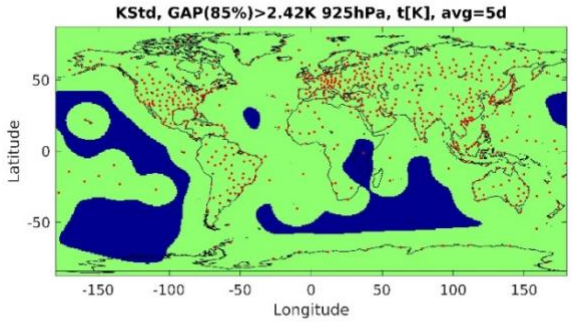
50hPa	200 hPa
	
500 hPa	925 hPa
	

Figure 4 - Temperature analysis: geographic gaps (dark blue) at 50, 200, 500 and 925 hPa defined as the region where the kriging uncertainty exceeds the 85th percentile of all values (e.g. the 15% of the global domain with highest sampling related uncertainty). The red dots are the RAOB stations.

5 Gap analysis for humidity

Following the same approach as in Section 4, we compare the meteorological variation of the specific humidity obtained by RAOB with the ERA-interim seasonal trend. Due to the radiosonde limited

capability of precise humidity measurement in the higher atmosphere, we focus solely on the pressure range between 925 and 300 hPa.

5.1 Preliminary analysis and modelling

The spatial correlation is first analysed by the variogram (or structure function) analysis. In Figure 5, the isotropic variogram is reported together with the directional variograms at the selected pressure levels. As expected they show a smaller correlation range than temperature data. In particular, at 300 hPa, even the isotropic variogram shows little spatial structure, but from the scale in g^2/kg^2, we see that the variation is quite small. At lower altitudes, the isotropic variogram reveals a reasonable spatial structure but with an increasing variability with increasing pressure (decreasing height) from 1 g^2/kg^2 at 500 hPa up to 19 g^2/kg^2 at 925 hPa.

Hence, the overall expectation on the performance of model (4) for humidity is lower than is the case for temperature.

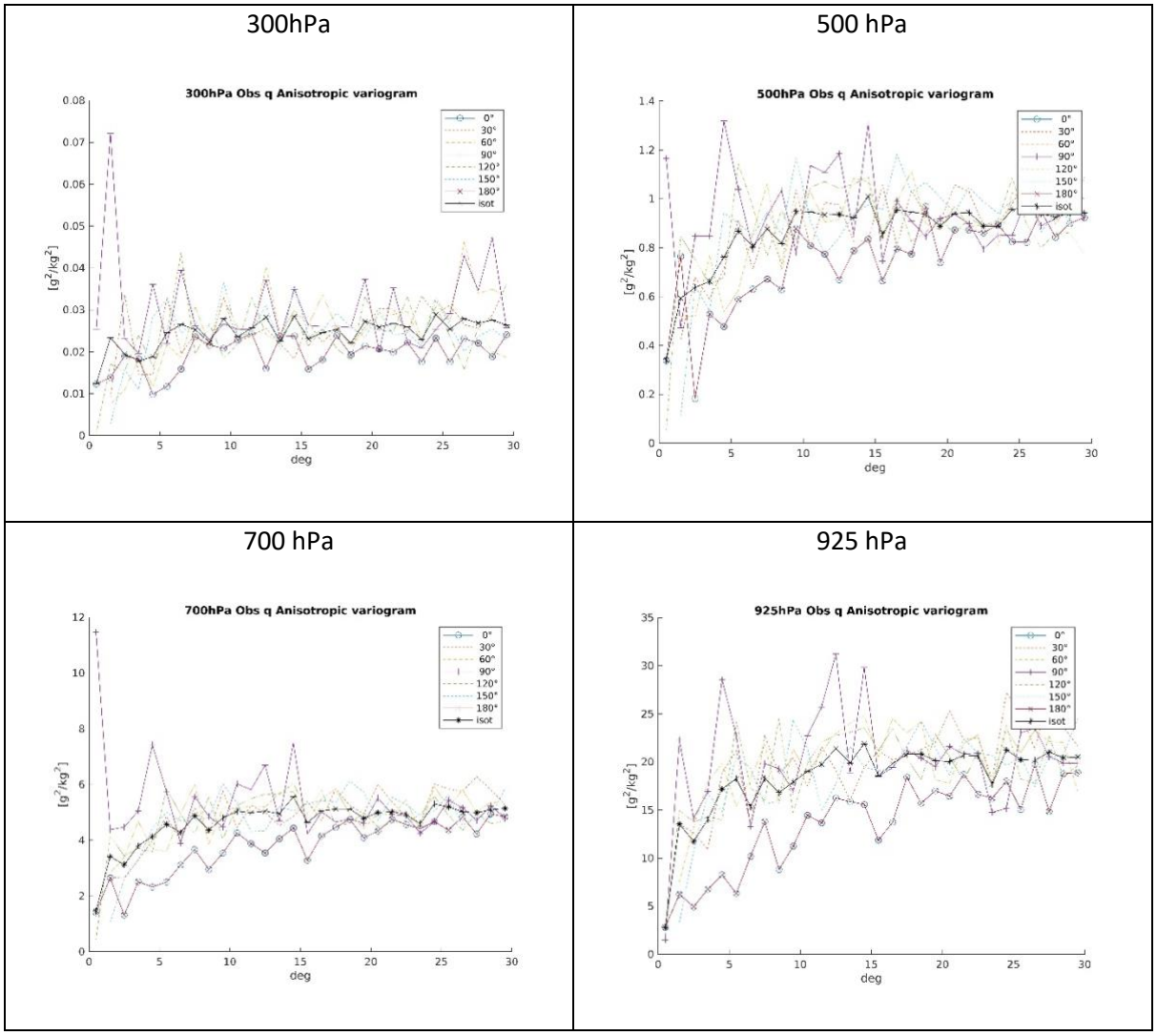


Figure 5 - Humidity analysis: anisotropic variograms of RAOB data at selected pressure levels. Isotropic variograms: Black solid line with stars; Longitudinal variogram: purple solid line with crosses and cyan circles; Latitudinal variograms: purple solid line with plus signs; Variograms at intermediate directions: dashed lines and dotted lines. Note that y-axis ranges vary between panels.

The estimated model (4) is reported in Table 4. At 300 hPa, the large standard deviations for theta1 and theta2 highlight the poor identifiability of the spatial structure. Also, the estimated relation with ERA-interim is lower ($\beta = 0.16$) than at remaining pressure levels. In this atmospheric layer, various elements may contribute to this increased uncertainty: atmospheric variability, instrumental limitations in measuring low levels of humidity, ERA-interim model limitations, the statistical approach used, and day and night data being affected by different solar radiation biases.

For the remaining pressure levels, the model is consistently estimated with a spatial range decreasing from 1500 to 1000 km. Although the temporal autocorrelation is only moderate, as indicated by the small value of g, the spatio-temporal random component (Z) has an important role as shown by the variance sigma2_Z, which is quite a bit larger than the corresponding term sigma2_eps.

The cross-validation analysis, analogous to Section 4.1, is reported in Table 5. The interpolation properties of model (4) now increase in the lower atmosphere with the percentage of the total variance, which increases from 0.76 to 0.96 and R-MSU which decreases from 0,92 to 0.06 g/kg.

Table 4 - Humidity analysis: estimated coefficients (mle) and standard deviations (std) of the GP model defined in Equation (4).

Pressure level		ERA-interim	White noise	GP scale			Spatial correlation	
			sigma2_eps	g	alpha	sigma2_Z	theta1	theta2
		[K/K]	[K^2]	[K/K]	[K/K]	[K^2]	[km]	[°]
300	mle	0.916	0.003	0.398	0.074	0.006	1553.2	20.5
	std	*0.015*	*0.000*	*0.008*	*0.006*		*7462.4*	*190.2*
400	mle	1.019	0.017	0.306	0.254	0.071	1552.1	18.1
	std	*0.016*	*0.000*	*0.006*	*0.005*		*419.6*	*8.5*
500	mle	1.009	0.049	0.273	0.509	0.280	1390.9	18.3
	std	*0.014*	*0.001*	*0.005*	*0.004*		*75.4*	*1.8*
700	mle	0.995	0.182	0.273	0.947	0.969	1225.9	20.7
	std	*0.011*	*0.004*	*0.005*	*0.004*		*19.6*	*0.7*
850	mle	0.992	0.275	0.345	1.018	1.177	1081.6	32.0
	std	*0.008*	*0.006*	*0.006*	*0.005*		*20.0*	*2.0*
925	mle	0.995	0.250	0.392	1.012	1.210	1054.0	39.7
	std	*0.007*	*0.005*	*0.006*	*0.004*		*21.3*	*3.4*

Table 5 - Humidity analysis: model and estimation performance. Number of iterations to convergence in EM algorithm with an exiting tolerance τ=0.001; fitting is given by Root Mean Squared Uncertainty (R-MSU) and R^2=1-MSU/VAR.

Pressure level	Iterations [#]	R^2 [% of var]	RMSE [g/kg]
300	20	0.78	0.92
400	22	0.79	0.76
500	30	0.80	0.53
700	27	0.87	0.14
850	22	0.93	0.02
925	22	0.96	0.06

5.2 Mapping and gaps

Using the estimated model of Table 4 and the kriging formulas of Section 3, we get the RAOB humidity uncertainty maps of Figure 6. As for the temperature analysis, the maps are similar but not identical since the theta parameter increases with altitude. More importantly the scale is different at different pressure levels, see Table 6. As in Section 4.2, Figure 7 highlights the 15% of the global domain exhibiting highest uncertainty from spatial sampling effects under model (4). Again, the gaps are similar at different pressure levels and differences concern mainly the South-eastern African gap in the lower atmosphere, whilst the size of Oceanic tropical gap has a smaller variation for humidity with respect to temperature. As shown in Table 6, the threshold level increases from 0.08 g/kg at 300 hPa to 1.1 g/kg at 925 hPa. This is consistent with the increase of humidity in the lower atmosphere.

Table 6 – Humidity analysis: Sampling (kriging) uncertainty statistics at pressure levels 300, 500, 700 and 925hPa.

pressure level	Kriging uncertainty [g/kg]		
	min	max	gap threshold (85%)
300	0.03	0.08	0.08
500	0.17	0.53	0.52
700	0.33	0.99	0.96
925	0.37	1.10	1.08

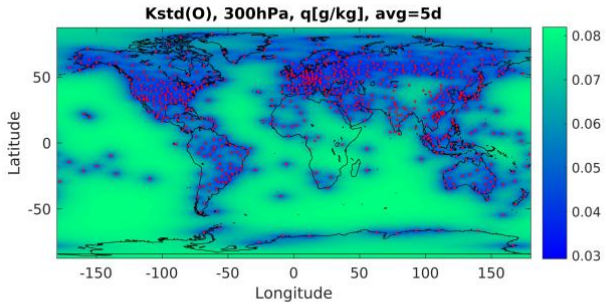
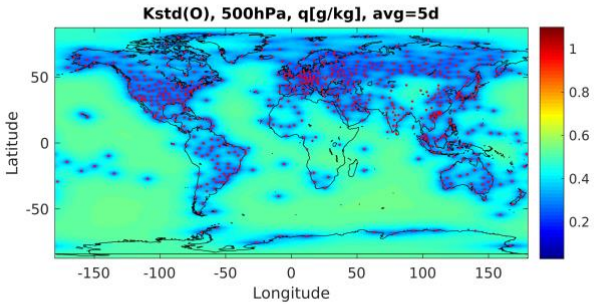
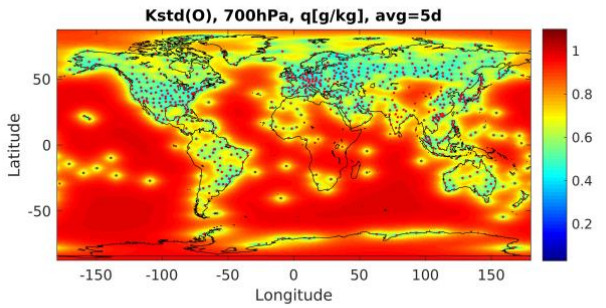
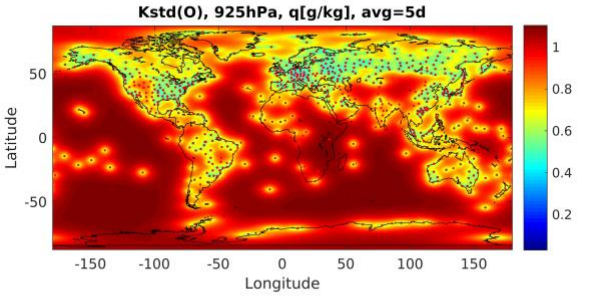
300hPa	500 hPa
	
700 hPa	925 hPa
	

Figure 6 – Humidity analysis: kriging uncertainty at 300, 500, 7000 and 925 hPa. Note that the colour scale of 300 hPa is different from the other panels. Given that the absolute ranges (Table 6) vary over an order of magnitude this is unavoidable.

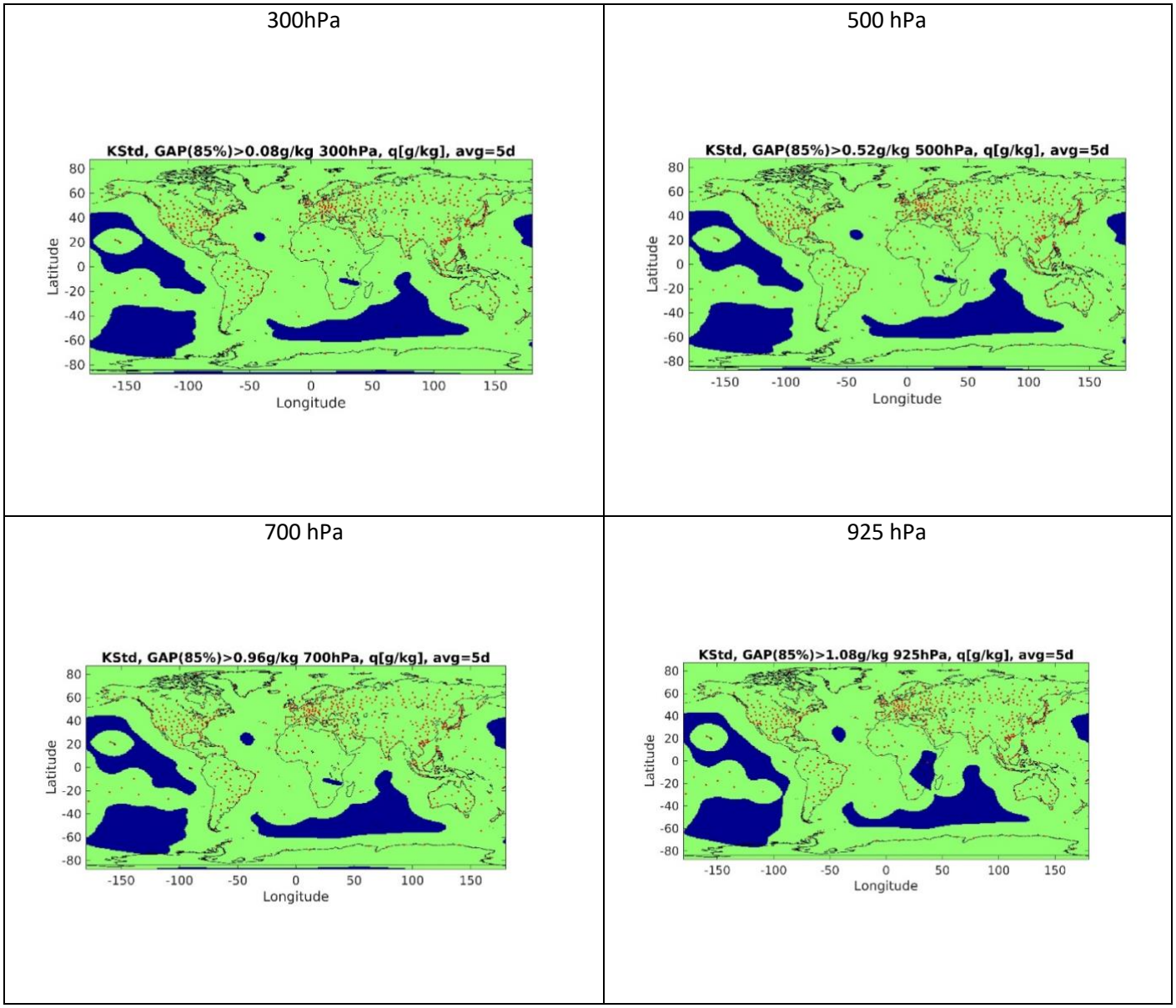


Figure 7 – Humidity analysis: geographic gaps (dark blue) at 50, 200, 500 and 925 hPa defined as the region where the kriging uncertainty exceeds the 85% percentile of all values calculated (i.e. shows the 15% of the global domain where sampling uncertainty is greatest). Red dots are RAOB stations.

6 Conclusions for temperature and humidity

This document considers a Gaussian process approach for identifying geographical gaps of temperature and humidity RAOB meteorological variations with respect to a seasonal background given by ERA-interim reanalysis. The gap regions are defined as being those regions of highest kriging uncertainty. It turns out that these regions depend on the (possibly anisotropic) correlation distance from the RAOB network. The gap regions are similar at different altitudes and concentrated primarily

over the oceans of the Southern Hemisphere. In principle this gap could be covered by a more intensive radiosonde observation from ships or from sustained operations on remote islands.

As expected, temperature has a larger spatial and temporal correlation, whilst, at high altitudes, it is difficult to fit a good GP model to humidity due to the very small values of specific humidity. In both cases, the gap regions have been identified as the 15% of sphere with largest kriging uncertainty. For temperature, the uncertainty threshold increases from 1.4 K at 50 hPa to 3K at 925 hPa. For humidity, it increases from 0.08 g/kg at 300 hPa to 1.1 g/kg at 925 hPa.

The model used has a good interpolating performance as assessed by a cross-validation exercise. In particular, the root mean square uncertainty ranges between 1 and 1.74 K for temperature with percentage of variance (R^2) generally above 0.98. For humidity the latter fitting index is lower, being between 0.78 and 0.96. Correspondingly, the root mean squared uncertainty increases from 0.06 g/kg at 925 hPa to 0.92 g/kg at 300 hPa.

The approach could be extended in future. For example gap maps could be produced based on physical thresholds instead of the percentile criterion used here. Further research may also be developed to include the effect of the sensor type, which could be of interest for example in data homogenisation, and the effect of solar radiation. However, development is needed to optimise the computational aspects. In particular, we would like to extend the study at single measurement scale, instead of the meteorological one given by the five days average used in this deliverable, and the functional modelling approach, which is capable to deal with the entire profile in a global manner. Such innovations would require either substantial optimisation, substantially increased computational capacity, or both, to achieve.

PART 2: OZONE AND METHANE

7 Small-scale natural variability of ozone profiles in the stratosphere

7.1 General remarks

While variability of zonally averaged ozone climatologies is well studied and reported [*Fortuin and Kelder*, 1998; *McPeters et al.*, 2007], the small-scale ozone variability (at distances up to a few hundred kilometres and over time intervals up to a few hours) is nearly unexplored. This is not surprising: to get this information, very accurate and closely co-located measurements are required. The variability of ozone in the stratosphere in most locations and seasons is very small, ~5% , even for a relatively large time interval of one month [*Fortuin and Kelder*, 1998; *Kyrölä et al.*, 2006; *McPeters et al.*, 2007]. Therefore, to get estimates of the ozone natural variability, the uncertainty of individual measurements should not exceed 1-2%. Such uncertainty can be achieved with satellite solar occultation or stellar occultation measurements (for the latter, sufficiently bright stars can be used). The coverage by solar occultation measurements is insufficient to get close time-space co-locations, while this can be achieved with stellar occultation measurements. A similar precision can be achieved for infrared limb sounders. In our analysis, we use the data from GOMOS (Global Ozone Monitoring by Occultation of Stars) and MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) satellite instruments.

7.2 Satellite Data

7.2.1 GOMOS

GOMOS is a stellar occultation instrument that operated on board of Envisat over 2002-2012 [*Bertaux et al.*, 2010; *Kyrölä et al.*, 2010]. Ozone profiles are retrieved from the ultraviolet (UV) and visible spectrometer measurements at wavelengths between 250 and 692 nm. In this work, we use night-time ozone profiles (with solar zenith angle larger than 105°), which are retrieved from atmospheric transmittance spectra. The GOMOS ozone profiles are obtained with the ALGOM2s v1.0 processor [*Sofieva et al.*, 2017]. ALGOM2s is nearly identical to the ESA IPF v6 processor (GOMOS IPF v6 data were used in the WMO-2014 ozone assessment) in the stratosphere, but has improved data quality in the UTLS.

GOMOS provides stratospheric ozone profiles with a vertical resolution of 2 km below 30 km, and 3 km above 30 km, with a linear transition between [*Tamminen et al.*, 2010]. The vertical resolution of the GOMOS ozone profiles is the same for all occultations due to the Tikhonov-type [*Tikhonov*, 1963] target-resolution regularization [*Sofieva et al.*, 2004; *Kyrölä et al.*, 2010]. The stellar flux recorded by GOMOS, and thus signal-to-noise ratio and uncertainty of retrieved profiles, depends on stellar

magnitude and spectral class. The estimated random uncertainty of GOMOS ozone profiles in the stratosphere is 0.5-5 % [*Tamminen et al.*, 2010]. Validation of estimated uncertainties for ozone profiles in the stratosphere has shown that they are realistic except for cases when ozone profiles are derived from occultations of dim stars [*Sofieva et al.*, 2014].

7.2.2 MIPAS

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) is an infrared limb emission spectrometer that was also flown on the Envisat platform [*Fischer et al.*, 2008]. In 2002-2004, the instrument operated at full spectral resolution. Due to a failure of the instrument's mirror slide in 2004, the operations were suspended for almost a year and were resumed in 2005 with reduced spectral, but improved vertical, resolution. These operations continued until the loss of communications with the ENVISAT platform in April 2012.

Stratospheric ozone profiles are retrieved from MIPAS/ENVISAT limb emission spectra. In this work, we use the scientific MIPAS processor IMK/IAA V5R_O3_220/221 developed at Karlsruhe Institute of Technology. The retrieval is performed via constrained inverse modelling of limb radiances. In the stratospheric/tropospheric retrievals, local thermodynamic equilibrium (LTE) is assumed. A detailed description can be found in *von Clarmann et al. (2003; 2009)*. The data are available in the user-friendly form as a HARMOZ data set [*Sofieva et al.*, 2013]. This dataset we used in our analyses.

Due to their different data characteristics, the two MIPAS measurement periods are usually treated as two independent datasets. Their processing schemes are different, and the vertical resolution of the early MIPAS period is lower than that of the later period: 3.5–6 km vs 2–5 km for retrieved ozone. In our work, we solely use MIPAS data in the period 2005-2012. The mean uncertainty of retrieved MIPAS ozone profiles in 2005-2012 is 0.05-0.1 ppmv (1-5 %). Validation of MIPAS random uncertainty estimates has been performed via evaluation of structure functions [*Laeng et al.*, 2015].

7.2.3 Self-co-located data

MIPAS has a relatively dense sampling pattern with ~1000 profiles each day, which is shown in Figure 8 (top). The MIPAS self-co-located data can be found in polar regions above 80° latitude. MIPAS consecutive measurements have temporal difference 2 min and they are separated by a distance of ~400 km, mainly in latitudinal direction. In our analyses, we selected consecutive pairs with separation <4°in latitude and <1.5° in longitude. The number of consecutive pairs is large, about 2/3 of the total number of measurements.

GOMOS has a significantly lower data sampling, ~100-200 profiles per day. The number of self-co-located GOMOS measurements which are separated by less than 500 km in space and by less than 3 h in time is very limited. A sufficient amount (~200 per year) of co-located occultations of the same star are available only for one star, namely star number 30 in the GOMOS catalogue. These co-locations occur during winter close to the North Pole.

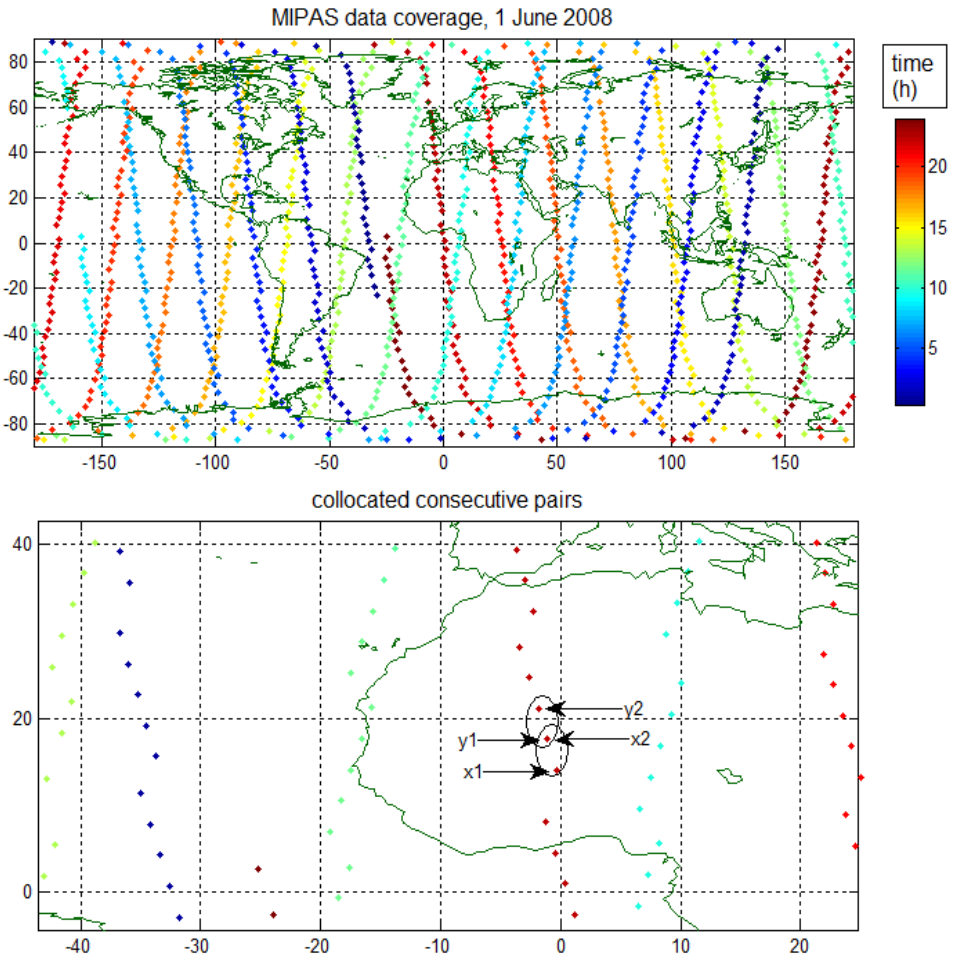


Figure 8. Top: MIPAS data coverage in one day, 1 June 2008, color indicates the time of measurements. Bottom: creating pair of consecutive measurements, see text for explanation.

7.3 Data analysis

7.3.1 Evaluation of structure functions

The degree of spatial dependence of a random field (or a stochastic process) $f(\mathbf{r})$ is usually characterized by the structure function [e.g., *Tatarskii*, 1961]:

$$D(\boldsymbol{\rho}) = D(\mathbf{r}_1 - \mathbf{r}_2) = \left\langle \left[f(\mathbf{r}_1) - f(\mathbf{r}_2) \right]^2 \right\rangle, \qquad (5)$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are two locations and $\boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2$. This concept assumes that the random field is locally homogeneous, which is the spatial equivalence of a random process with stationary increments. In spatial statistics, $D(\boldsymbol{\rho})$ is called the variogram [*Wackernagel*, 2003].

When using experimental profiles, the difference of ozone between two points is defined not only by the natural variability of the ozone field but also by the uncertainty of the measurements. Therefore, in order to get experimental estimates of the structure function, estimated mean data variance should be subtracted:

$$D_0(\boldsymbol{\rho}) = D(\boldsymbol{\rho}) - \sigma_1^2 - \sigma_2^2 . \qquad (6)$$

The estimates of profile uncertainty are provided with the inversion algorithm. For the MIPAS and GOMOS measurements considered here, $\sigma_1 = \sigma_2 = \sigma_{err}$, so that Equation (6) is reduced to:

$$D_0(\boldsymbol{\rho}) = D(\boldsymbol{\rho}) - 2\sigma_{err}^2$$

For small separation distances, the structure functions are nearly isotropic in longitudinal and latitudinal directions, therefore an isotropic 1D structure function can be considered (i.e., rms ozone difference is dependent on separation distance only). In the computation of the separation distance, we take into account the advection of air masses. In our computations, the wind profiles are taken from the ERA Interim reanalysis at the locations of satellite measurements.

To reduce uncertainty in evaluation of the structure function using the experimental data (which is inversely proportional to the number of measurements), we present the cumulative structure function $D_0(\rho_c)$, where the mean in Equation (5) is evaluated over all values $\rho < \rho_c$. Figure 9 shows $\sqrt{D_0(\rho_c)}$ near the North Pole in summer (left) and winter (right). For summer, only MIPAS data are available (no night-time measurements from GOMOS). As observed in Figure 9, the small-scale ozone variability is significantly larger in winter than in summer (as expected). In most cases, the rms of ozone differences grows slowly with increasing separation distance, for the considered range of distances. Ozone variability also grows with altitude, especially in winter.

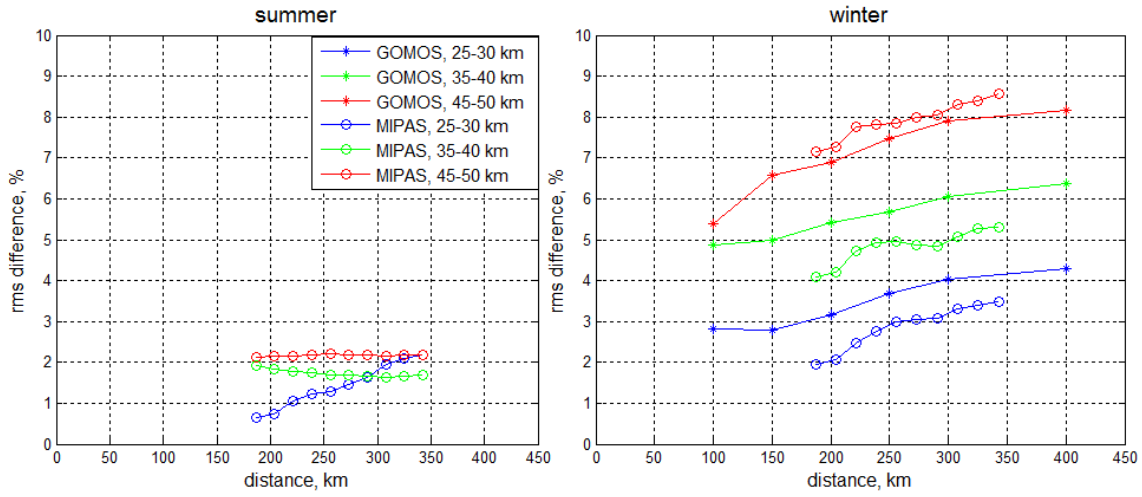


Figure 9. $\sqrt{D_0(\rho_c)}$ near the North Pole in summer (left) and winter (right).

7.3.2 Ozone variability using consecutive MIPAS measurements

Consecutive MIPAS measurements have nearly constant horizontal separation of ~400 km, mainly in the latitudinal direction. Therefore, the consecutive measurements allow estimation of natural ozone variations corresponding to a separation distance of up to ~400 km. From another point of view, a large number of such co-located consecutive observations allows detailed analysis of ozone variability at such scales.

In order to remove the mean ozone gradient, we estimated the small-scale natural ozone variability as

$$\sigma^2_{ozone} = \mathrm{var}\left(\frac{x_i - y_i}{0.5(x_i + y_i)}\right) - 2\sigma^2_{err},$$

where the first term is the variance of relative ozone differences in co-located pairs (x_i, y_i), see illustration of co-located pairs in Figure 8, while the second term is the mean relative data uncertainty. The data processing is illustrated in Figure 10.

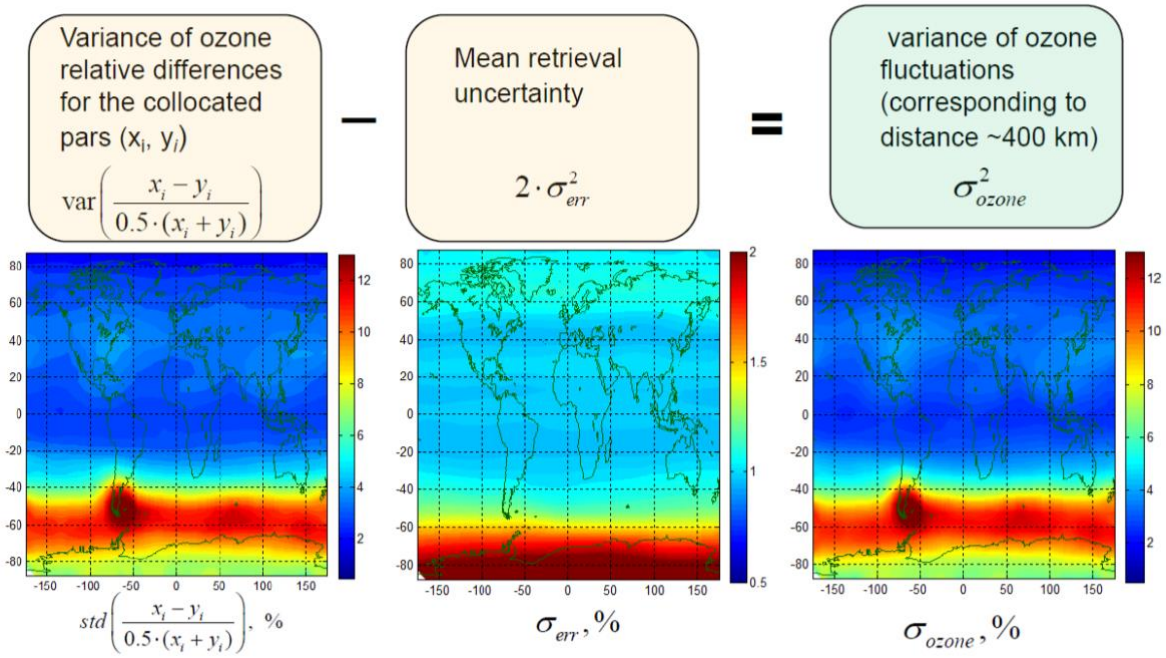


Figure 10. Data processing using consecutive MIPAS measurements. The illustration is for June-August in 2007-2011, altitudes 40-45 km.

Figure 11 shows ozone small scale variability at 40-45 km estimated from MIPAS consecutive measurements in 2007-2011 (the years fully covered with MIPAS data are selected) in two seasons: December-February (top) and June–August (bottom). As seen in Figure 11, ozone variability is small in tropics, below 5%, while it can be rather large, up to 10-15% in the polar winter stratosphere. An enhanced variability associated with the polar vortex is observed as expected. We note enhancements

of σ_{ozone} over typical regions of high gravity wave (GW) activity, in particular, over the Andes and Antarctic Peninsula in austral winter.

Similar patterns are observed in temperature variability (obtained in the same way as ozone variability using consecutive MIPAS observations), and in previous gravity wave studies (e.g., Tsuda et al., 2018; Jiang et al., 2004; Sofieva et al., 2009; Ern and Preusse, 2012, Fritts and Alexander, 2003).

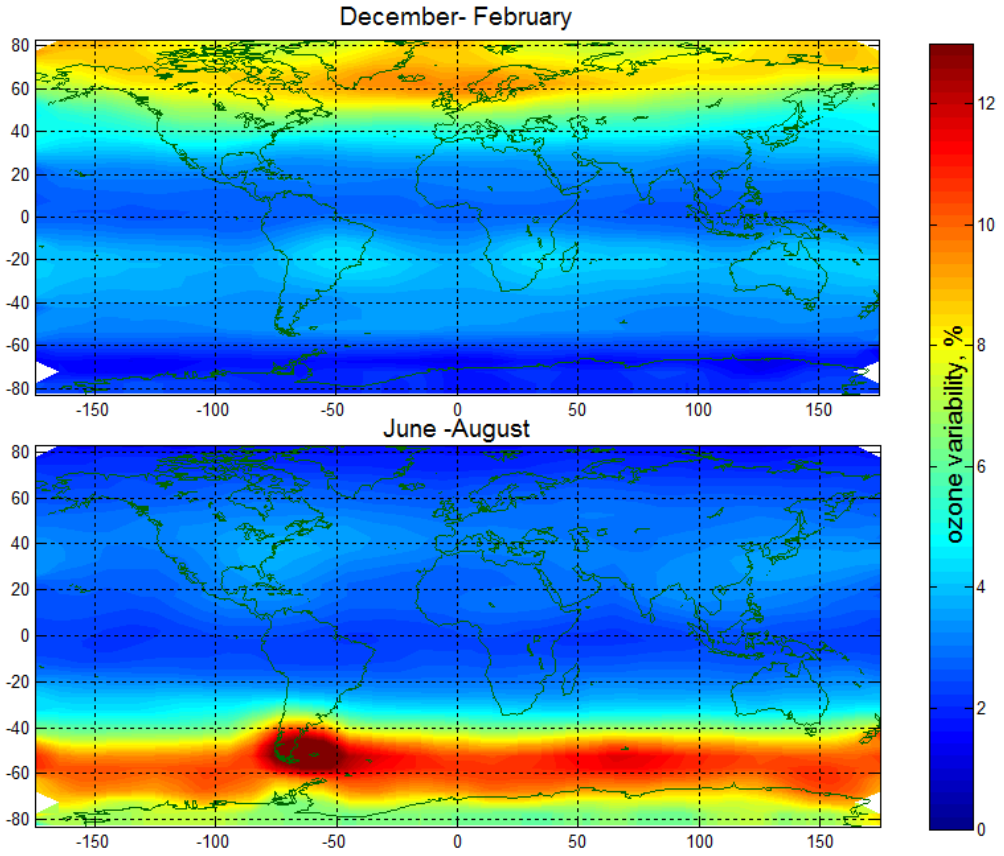


Figure 11. Ozone small-scale variability σ_{ozone} at 40-45 km estimated using consecutive MIPAS measurements in 2007-2011.

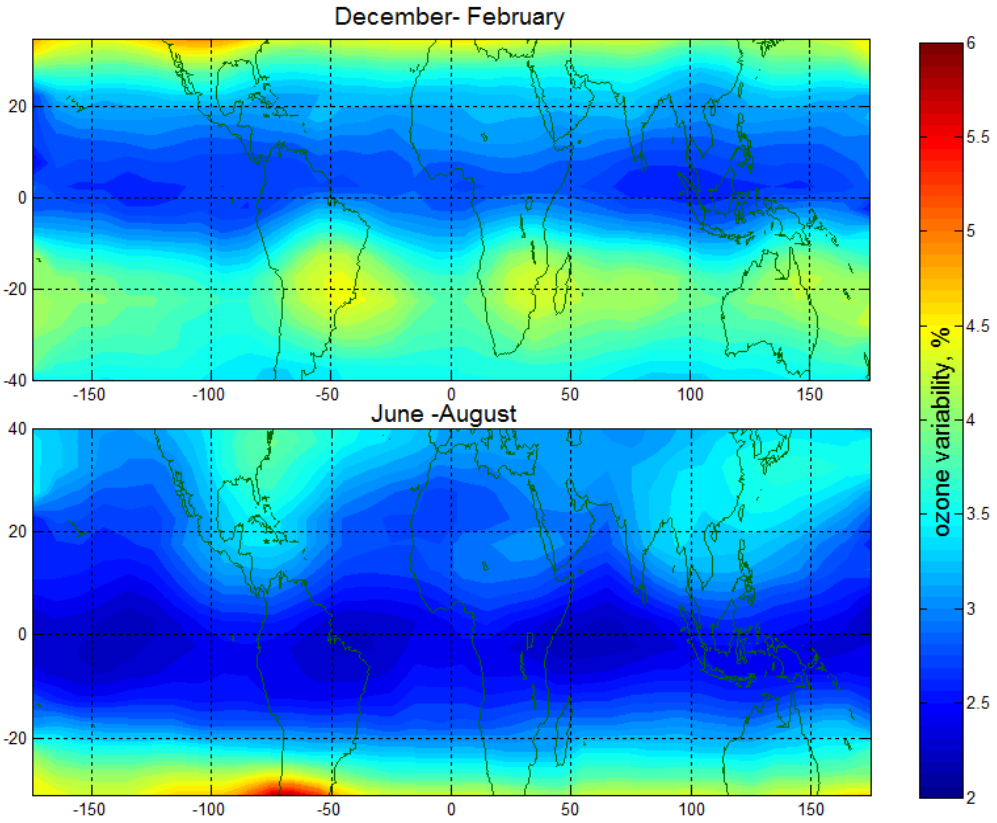


Figure 12. As Figure 11, but zoomed on the tropics.

Ozone variability is smaller in tropics than in polar regions (usually below 5 %), and it has a pronounced zonal structure, as illustrated in Figure 12. Average values of σ_{ozone} in the tropics follow the sub-solar latitude. Almost all of the local tropical enhancements of σ_{ozone} are over continents. Many of these enhancements correspond with typical regions of deep convection. Tentatively, they are therefore most likely to be induced by gravity waves generated by deep convection.

Figure 13 shows the maps of small-scale ozone variability in December-February, for several altitude levels. Gravity wave signatures in the ozone variability field become visible in the upper stratosphere, where the gravity wave amplitude is sufficiently large. This behaviour is expected.

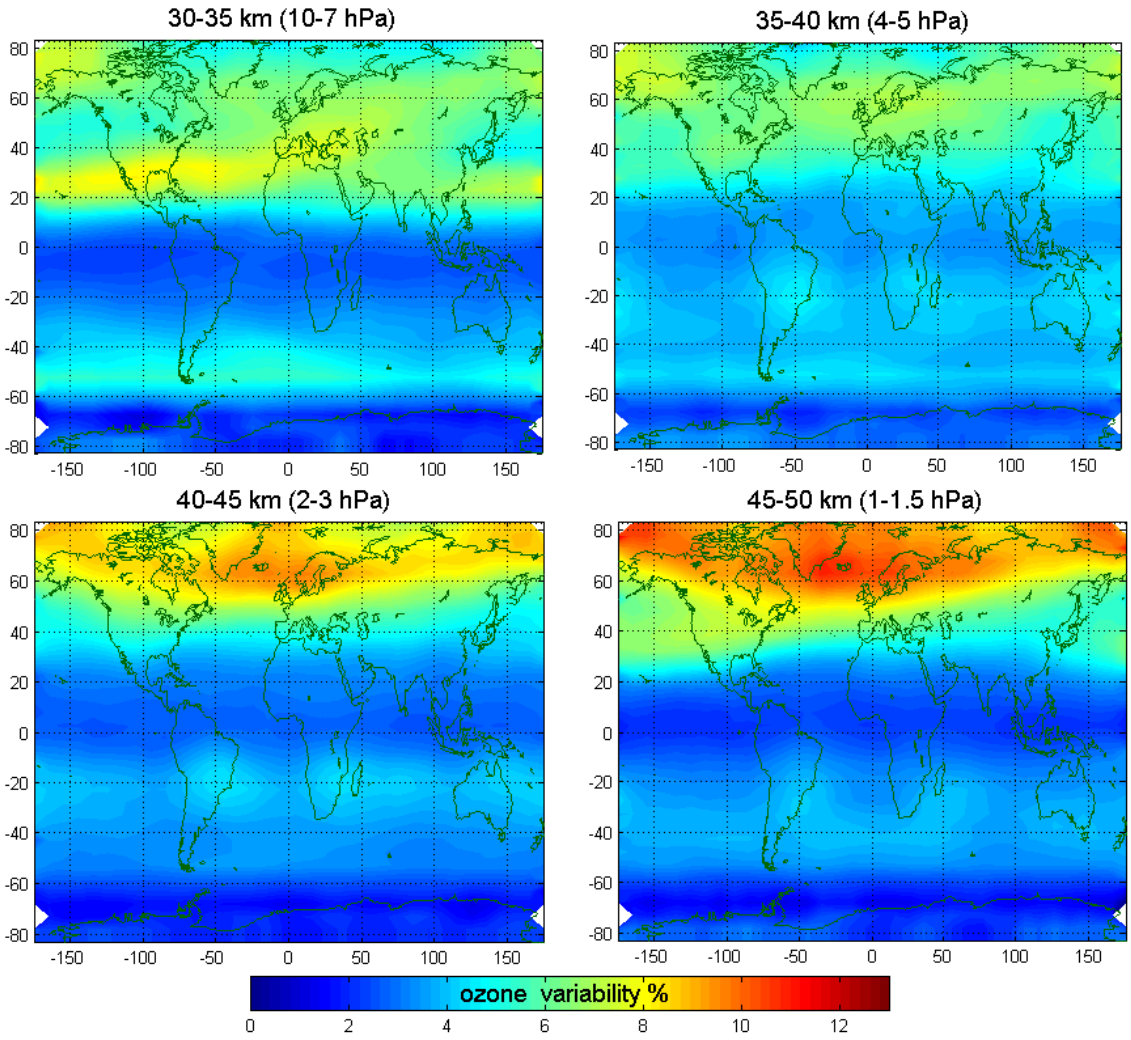


Figure 13. Small-scale ozone variability in December- February, for several altitude levels.

7.4 Conclusions for ozone

We presented analyses of variability of ozone profiles in the stratosphere using ozone profile measurements by satellite instruments GOMOS and MIPAS. The observed ozone variability is assessed for the horizontal scales up to a few hundreds of kilometres and temporal difference up to a few hours. Such variability arises from small-scale atmospheric processes.

The following main conclusions can be made from our analyses:

1. The ozone natural small-scale variability is small in tropics (< 5%) and can be rather large in polar winter stratosphere (up to 10-15 %).

2. In the upper stratosphere, larger ozone variability is observed at locations of enhanced gravity wave activity. This analysis indicated gravity wave signatures in the ozone variability field.

3. In general, ozone variability in the stratosphere increases with altitude.

These findings are in agreement with general expectations: ozone variability is larger in dynamically active atmospheric regions. In terms of gaps in ground-based measurements, our analysis indicates that ground-based measurements in dynamically active regions should be frequent for informative validation.

8 CH_4 total column

8.1 Overview of the study

The small-scale variability of the column-averaged dry air mole fraction of methane (XCH_4) is studied using CarbonTracker Europe-CH_4 model data. We applied the variogram analysis to the modelled field in order to estimate the small-scale variability in methane.

In addition to studying the small-scale variability of XCH_4 (total column burdens), we made a small case study. We used the information about the small-scale variability to identify a suitable distance from the station, inside which the measurements would best serve the validation of satellite measurements. In this case study, we used data from GOSAT satellite and the ground-based TCCON network.

8.1.1 CarbonTracker Europe-CH_4

The CarbonTracker Europe-CH_4 (CTE-CH_4) model optimizes global surface methane emissions from biospheric and anthropocentric sources [*Tsuruta et al.*, 2016]. The model is based on an ensemble Kalman filter optimization method with the TM5 chemistry transport model as an observation operator. The optimized emission estimates are constrained by global in-situ CH_4 measurements. The background of CTE-CH_4 is the global CarbonTracker model, which was originally developed to estimate CO_2 fluxes. The CH_4 model works globally, but it has a focus on Europe. In this work, we used only data from Europe. The resolution of the CTE-CH_4 model runs is 1^o x 1^o x 24h, and the studied years are 2011 and 2012.

8.1.2 GOSAT

The Greenhouse Gases Observing Satellite (GOSAT) was the first satellite dedicated to observations of greenhouse gases from space. The satellite performs near-global greenhouse gas measurements from a polar orbit [*Yokota et al.*, 2009]. The satellite was launched in January 2009 and it is a joint project of Japanese Aerospace Exploration Agency (JAXA), Japan National Institute for Environmental Studies (NIES) and the Ministry Of the Environment (MOE). GOSAT overpasses at 13:00 (in local time) every three days and the diameter of the footprint in nadir is approximately 10 km. Thermal And Near-infrared Sensor for carbon Observation Fourier Transform Spectrometer (TANSO-FTS) on-board GOSAT measures short-wavelength infrared (SWIR) radiation back scattered or emitted from the Earth's surface and atmosphere. GOSAT applies nadir mode above the land areas and sun glint mode over the ocean. Over land, in nadir mode, GOSAT uses two different gains: medium gain (Gain-M) is used over bright surfaces like deserts and high gain (Gain-H) is used elsewhere. Most of the data is Gain-H land data. In this data we used only land data that was retrieved with National Institute for Environmental Studies retrieval algorithm Version 02.21 (NIES V02.21).

8.1.3 TCCON

TCCON is a global network of ground-based Fourier Transform Spectrometers (FTS) measuring the spectrum of near-infrared radiation by looking directly to the Sun. The column-averaged abundances of methane and other atmospheric gases and molecules are retrieved from the measured spectra [*Wunch et al., 2010]*. By looking directly at the Sun, the measurements are not affected by surface albedo, which is one of the largest benefits of this measurement technique. TCCON measurements are calibrated against aircraft measurements to the World Meteorological Organization (WMO) scale and therefore they provide an ideal independent validation source for column-averaged methane measurements from space *[Wunch et al., 2010*]. The TCCON sites used in this study are shown in Figure 14. The studied sites are from the North to the South: Sodankylä, Bialystok, Bremen, Karlsruhe, Orleans and Garmisch.

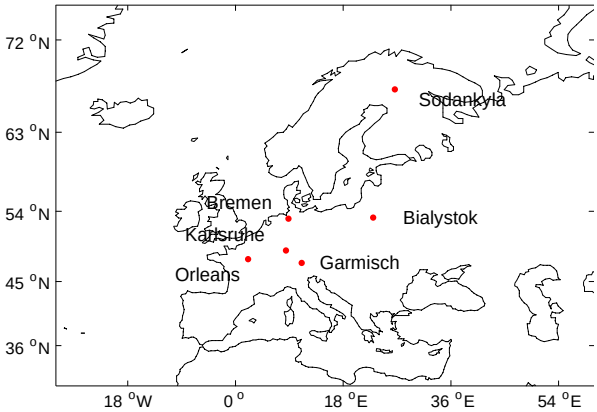


Figure 14. TCCON sites used in the case study.

8.2 Data analysis

8.2.1 The evaluation of structure functions (variogram)

In this study of CH_4 variability, we evaluate the variogram, which is also known as the structure function in the literature on atmospheric irregularities, from the spatial column-averaged dry air mole fraction of methane data produced by the CarbonTracker model. With a resolution of 1 °, we first collect the pairs of points with distances ranging from 1 to 25 °. For each pair of locations, we then computed a normalized variogram

$$D(\boldsymbol{\rho}) = D(\mathbf{r}_1 - \mathbf{r}_2) = \left\langle \frac{\left[f(\mathbf{r}_1) - f(\mathbf{r}_2) \right]^2}{\left[0.5 (f(\mathbf{r}_1) + f(\mathbf{r}_2)) \right]^2} \right\rangle$$

where $f(\mathbf{r}_1)$ and $f(\mathbf{r}_2)$ are the corresponding CH_4 mole fractions corresponding to spatial coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$. Daily variograms were computed for years 2011 and 2012, from which quarter-year averages were calculated for both years (quarter 1 is from January to March etc.). The results are visualized in Figure 15. In this figure, $\sqrt{D(\rho)}$ (in %) is shown as a function of spatial distance is $\rho = |\boldsymbol{\rho}|$ presented in °.

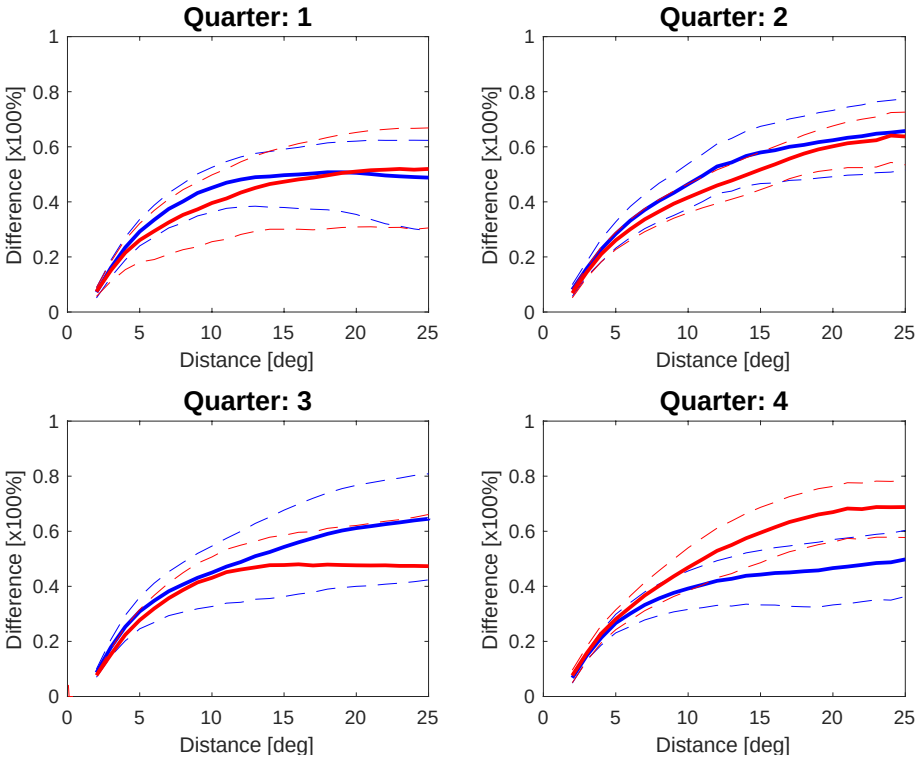


Figure 15. Quarterly mean variograms $\sqrt{D(\rho)}$ with 1 std limits for the entire are covered by the CarbonTracker Europe model. Year 2011 is seen in blue, and year 2012 in red.

As observed in Figure 15, the methane variations increase with separation distance and can be as large as 40-60 % at separations larger than 10°. A large variability, which is indicated by standard deviation in Figure 15, is observed. There is no obvious seasonal dependence, although our analysis shows somewhat larger relative variations in summer. There are also large inter-annual departures between the two years suggesting that synoptic-scale induced inter-annual variations may dominate over any seasonal signal.

8.2.2 Correlation estimation based on CarbonTracker model

For the case study, we computed north-south and east-west directed quarter-year variograms from the CarbonTracker model data over the studied TCCON sites using a standard formula for the structure function (Equation (5)). Figure 16 shows the variograms in north-south and east-west directions for quartiles in years 2011 and 2012.

As seen in Figure 16, there is no preferential direction in CH_4 variations over Europe: the variograms are similar in both latitudinal and longitudinal directions. On these variograms, we marked the points corresponding to 5ppb difference. For a few cases, this difference is rather close to the quasi-saturation of variograms, for others, it is ~1/2 of full range of variations in the explored range of distances.

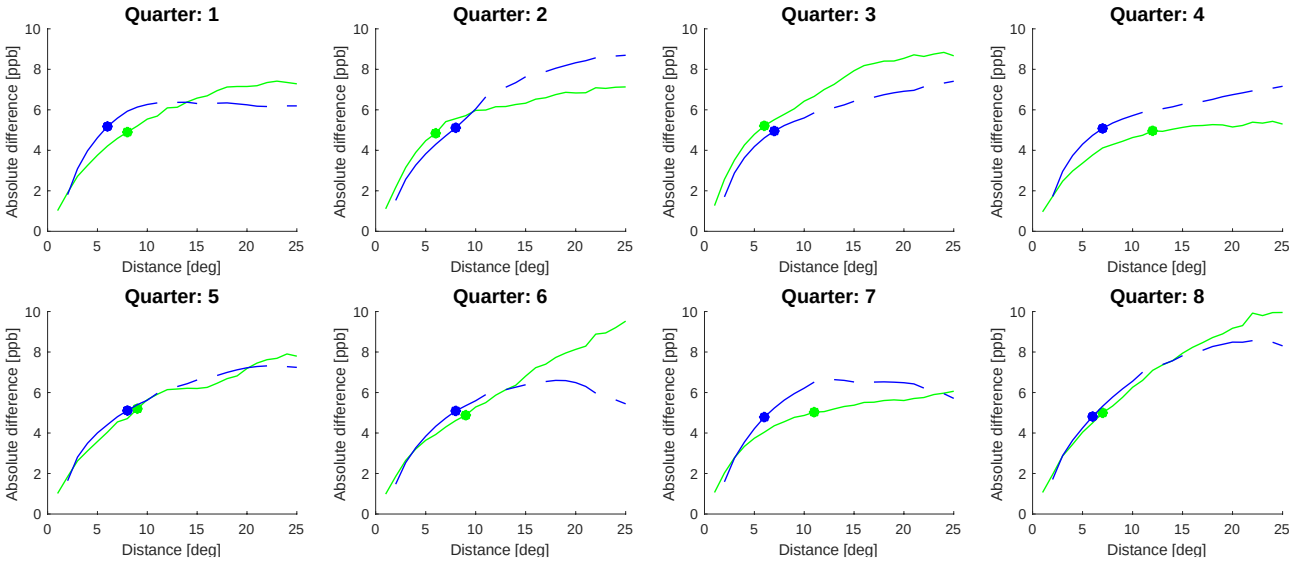


Figure 16. Quarter-yearly absolute difference variograms with quarters 1-4 corresponding year 2011 and quarters 5-8 corresponding to year 2012. The north-south direction is plotted in green and the east-west direction in blue. The dots indicate the position of the 5ppb difference.

8.3 The case study

To evaluate space-based satellite XCH_4 observation against ground-based observations the satellite observations have to be co-located with some ground-based observations. Almost in every XCH_4 satellite study, the ground-based observations are made at a TCCON site, because the measurement technique of TCCON sites corresponds closely to the measurement technique of GHG satellites. Co-location can be done in multiple ways, but all of the techniques make assumptions about the geographical region over which the space based XCH_4 observation is assumed to be measuring the same concentration as the ground-based measurement within some tolerance**.** The main problem with co-location techniques is a trade-off between having a sufficient number of measurements for reliable statistics and a sufficiently small area around the TCCON site for similarity of ground-based and satellite observations through minimising co-location effects.

Usually in validation studies the satellite observations are chosen from a small box around the TCCON site, in general the size of the box is ± 2° in latitude and in longitude. In case we want to study something that needs quite a lot of observations over the period of interest, for example the seasonal cycle, this small box co-location criterion might be too tight. To raise the number of soundings, dynamical criteria, based on modelled concentrations of XCH_4, can be used [e.g., *Lindqvist*, 2015]. Figure 17 illustrates the co-location method based on a dynamical criterion, in which the data are considered as co-located if the underlying model values are within +/-5 ppb range. In [*Lindqvist*, 2015], the threshold value of 5 ppb has been used as a criterion of "the same air mass".

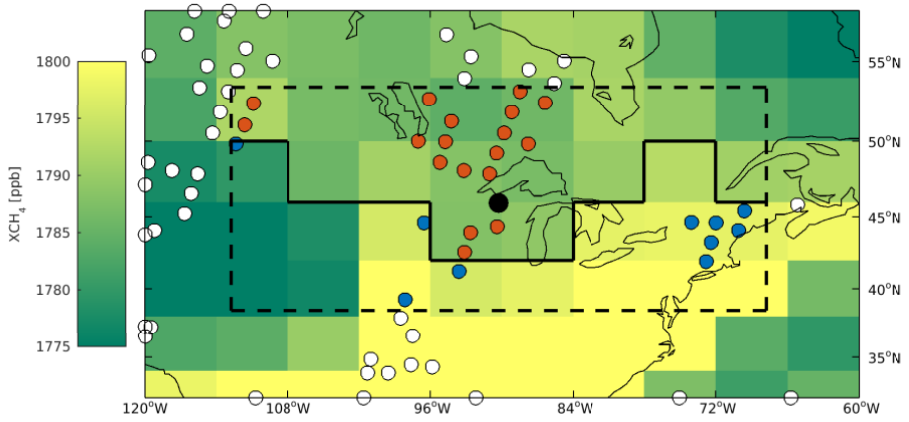


Figure 17. An example from the GOSAT-TCCON co-location method at Park Falls TCCON station (Wisconsin, USA) 2 September 2009, TCCON site location is marker with black circle. Black dashed line shows the maximum range and the black solid line shows the border of the air mass. GOSAT soundings are presented by filled circles, white circles shows the GOSAT soundings that are limited out by maximum range, blue circles shows soundings that are limited out by +/-5 ppb air mass. Red circles shows GOSAT soundings that are co-located to correspond to the TCCON soundings. The modeled concentrations in the background are from global CTE-CH_4 run, the grid of the model concentrations is in this example 6° x 4°

In addition to the traditional ± 2° criterion and the dynamical ± 5 ppb criteria, we applied also the selection of measurements based on the variogram method. In this method, we select the measurements within the distance corresponding to the 5 ppb value on the variogram. For the six stations considered herein (shown in Figure 14), the variograms are shown in Figure 16. The mean separation corresponding to the 5 ppb variogram value is 7.4° in longitude and 7.5° in latitude. In the case study, we selected the co-located measurements according to this criterion.

The correlation between GOSAT and TCCON observations is evaluated using the Pearson correlation coefficient:

$$\mathrm{C}(A,B) = \frac{\mathrm{cov}(A,B)}{\sigma_A \sigma_B} \ ,$$

where *A* refers to GOSAT and *B* to TCCON. We calculated the correlation coefficients for individual soundings and for daily averages. We show the results averaged over the studied TCCON sites in Table 7.

Table 7. The statistics of co-located data averaged over the six studied TCCON sites.

	Limitation criteria based on variogram analysis	Dynamical criteria, ±5 ppb air mass	Box: ± 2 ° longitude; ± 2 ° latitude
Number of the individual soundings, average over the sites	1054	1298	117
STD of number of the soundings	1006	946	108
$C_{individual\ soundings}$, average over the sites	0,37	0,40	0,48
STD of $C_{ndividual\ soundings}$	0,14	0,13	0,22
Number of daily averages, average over the sites	137	168	42
STD of daily averages number	101	110	30
$C_{daily\ averages}$, average over the sites	0,46	0,51	0,52
STD of $C_{daily\ averages}$	0,16	0,13	0,18

The results show that co-location criterion based on variogram analysis is almost as good as the dynamical co-location criterion, and the number of soundings is quite similar for these two techniques. This variogram method was faster than dynamical co-location method and did not require use of output from a model. Thus further developing the variogram analysis might be useful for finding the co-locations between the satellite and ground-based measurements in a more efficient and transferable manner.

8.4 Conclusions for CH_4

The CarbonTracker Europe-CH_4 model was used to produce a 1°x1° grid of daily CH_4 concentrations over Europe over 2011 to 2012. A spatial variogram analysis was then applied to find the mean separation corresponding to 5ppb difference for European TCCON measurement sites. This information was used to identify a suitable distance from the station, inside which the measurements

would best serve the validation of satellite measurements. The case study shows that the co-location criteria based on variogram analysis give almost as good results as the traditionally used dynamical limitation criteria when co-locating GOSAT and TCCON observations.

9 Overall conclusions

Task 1.4 of WP 1 in GAIA-CLIM considered methods based on geostatistics for studying geographical gaps from two different perspectives. The first one considered radiosonde temperature and humidity by means of an O-B comparison based on RAOB observations and the ERA-interim reanalysis. A geographical gap has been defined as the 15% of the global domain exhibiting the highest uncertainty. Unsurprisingly, this gap region is where few stations are available: mainly over the oceans of the Southern Hemisphere. As a consequence, this gap could be covered by a more intensive radiosonde observation from ships or sustained launch programs from remote island locations. The importance of these gaps in terms of uncertainty level increases in the lower atmosphere.

The second approach has been used to understand the small-scale uncertainty of ozone and methane. Stratospheric ozone profiles obtained by satellite instruments GOMOS and MIPAS have been considered. The results based on structure function (variogram) properties are consistent with general expectations: ozone variability is larger in dynamically active atmospheric regions. In particular, a seasonal-geographic gap with high ozone variability may be identified in winter polar zones. Moreover ozone variability is higher in the upper stratosphere and where gravitational wave activity is higher. Regarding methane (total column), the variogram analysis was applied to CarbonTracker Europe-CH_4 model outputs, and used to find the mean separation corresponding to 5ppb difference for European TCCON measurement sites. The results were comparable to the traditionally used dynamical limitation criteria when co-locating GOSAT and TCCON observations which requires access to dynamical model output.

References

Bertaux, J.-L. et al. (2010), Global ozone monitoring by occultation of stars: an overview of GOMOS measurements on ENVISAT, *Atmos. Chem. Phys.*, *10*(24), 12091–12148, doi:10.5194/acp-10-12091-2010.

Bolton, D., 1980: The Computation of Equivalent Potential Temperature. Mon. Wea. Rev., 108, 1046–1053, https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2

Calculli C, Fassò A, Finazzi F, Pollice A, Turnone A. (2015) Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy. Environmetrics. 26(6), 406–417.

von Clarmann, T. et al. (2003), Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), *J. Geophys. Res.*, *108*(D23), 4736, doi:10.1029/2003JD003602. [online] Available from: http://dx.doi.org/10.1029/2003JD003602

von Clarmann, T. et al. (2009), Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, *Atmos. Meas. Tech.*, *2*(1), 159–175, doi:10.5194/amt-2-159-2009. [online] Available from: http://www.atmos-meas-tech.net/2/159/2009/

Cressie, N. and Johannesson, G. (2008), "Fixed rank kriging for very large spatial data sets," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70, 209–226.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016), "Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets," Journal of the American Statistical Association, 111, 800–812.

Ern, M. and Preusse, P.: Gravity wave momentum flux spectra observed from satellite in the summertime subtropics: Implications for global modeling, Geophys. Res. Lett., 39(15), L15810, doi:10.1029/2012GL052659, 2012.

Fassò, A, Ignaccolo, R, Madonna, F, Demoz, B. and Franco-Villoria M. (2014) Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmos. Meas. Tech., 7, 1803–1816, doi:10.5194/amt-7-1803-2014. http://www.atmos-meas-tech.net/7/1803/2014/amt-7-1803-2014.pdf

Finazzi F, Fassò A. (2014) D-STEM: A Software for the Analysis and Mapping of Environmental Space-Time Variables. Journal of Statistical Software. Vol. 62, Issue 6, 1-29.

Fischer, H. et al. (2008), MIPAS: an instrument for atmospheric and climate research, *Atmos. Chem. Phys.*, *8*(8), 2151–2188, doi:10.5194/acp-8-2151-2008. [online] Available from: https://www.atmos-chem-phys.net/8/2151/2008/

Fortuin, J. P. F., and H. Kelder (1998), An ozone climatology based on ozonesonde and satellite

measurements, *J. Geophys. Res.*, *103*(D24), 31709–31734.

Fritts, D. C. and Alexander, M. J.: Gravity wave dynamics and effects in the middle atmosphere, Rev. Geophys., 41(1), 1003, doi:10.1029/2001RG000106, 2003.

Jiang, J. H., Wang, B., Goya, K., Hocke, K., Eckermann, S. D., Ma, J., Wu, D. L. and Read, W. J.: Geographical distribution and interseasonal variability of tropical deep convection: UARS MLS observations and analyses, J. Geophys. Res., 109(D03111), doi:10.1029/2003JD003756, 2004.

Kyrölä, E. et al. (2006), Nighttime ozone profiles in the stratosphere and mesosphere by the Global Ozone Monitoring by Occultation of Stars on Envisat, *J. Geophys. Res.*, *111*(D24306), doi:10.1029/2006JD007193.

Kyrölä, E. et al. (2010), Retrieval of atmospheric parameters from GOMOS data, *Atmos. Chem. Phys.*, *10*(23), 11881–11903, doi:10.5194/acp-10-11881-2010. [online] Available from: http://www.atmos-chem-phys.net/10/11881/2010/

Laeng, A. et al. (2015), The ozone climate change initiative: Comparison of four Level-2 processors for the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), *Remote Sens. Environ.*, *162*(0), 316–343, doi:http://dx.doi.org/10.1016/j.rse.2014.12.013. [online] Available from: http://www.sciencedirect.com/science/article/pii/S0034425714005057

Ma. C. (2016) Time varying axially symmetric vector random fields on the sphere. Random Operators and Stochastic Equations, 24(4): 255-266.

McPeters, R. D., G. J. Labow, and J. A. Logan (2007), Ozone climatological profiles for satellite retrieval algorithms, *J. Geophys. Res.*, *112*(D5), D05308, doi:10.1029/2005JD006823.

Ramsay, J. O. and Silverman, B. W. (2005) Functional Data Analysis, Springer, Dordrecht.

Sherwood, S. C. (2000a) Climate signals from station arrays with missing data, and an application to winds. J. Geophys. Res., 105, 29 489–29 500.

Sherwood, S. C. (2000b), Climate signals mapping and an application to tides. Geophys. Res. Letters, 27, 21, 3525-3528.

Sherwood, S.C., Meyer C.L., Allen R.J., Titchner H.A. (2008) Robust tropospheric warming revealed by interactively homogenised radiosonde data. J. Clim. 21, 5336 – 5352.

Sofieva, V. F., Gurvich, A. S. and Dalaudier, F.: Gravity wave spectra parameters in 2003 retrieved from stellar scintillation measurements by GOMOS, Geophys. Res. Lett., 36, L05811, doi:10.1029/2008GL036726, 2009.

Sofieva, V. F. et al. (2013), Harmonized dataset of ozone profiles from satellite limb and occultation measurements, *Earth Syst. Sci. Data*, *5*(2), 349–363, doi:10.5194/essd-5-349-2013. [online] Available from: http://www.earth-syst-sci-data.net/5/349/2013/

Sofieva, V. F. et al. (2014), Validation of GOMOS ozone precision estimates in the stratosphere, *Atmos. Meas. Tech.*, *7*(7), 2147–2158, doi:10.5194/amt-7-2147-2014. [online] Available from:

http://www.atmos-meas-tech.net/7/2147/2014/

Sofieva, V. F. et al. (2017), Improved GOMOS/Envisat ozone retrievals in the upper troposphere and the lower stratosphere, *Atmos. Meas. Tech.*, *10*(1), 231–246, doi:10.5194/amt-10-231-2017. [online] Available from: http://www.atmos-meas-tech.net/10/231/2017/

Sofieva, V. F., J. Tamminen, H. Haario, E. Kyrölä, and M. Lehtinen (2004), Ozone profile smoothness as a priori information in the inversion from limb measurements, *Ann. Geophys.*, *22*(10), 3411–3420.

Stein M.L. (2007) Spatial variation of total column ozone on a global scale. The Annals of Applied Statistics, 1(1), 191-210.

Tamminen, J. et al. (2010), GOMOS data characterisation and error estimation, *Atmos. Chem. Phys.*, *10*(19), 9505–9519, doi:10.5194/acp-10-9505-2010.

Tatarskii, V. I. (1961), *Wave Propagation in a Turbulent Medium*, edited by R. A. Silverman, McGraw-Hill, New York.

Tikhonov, A. (1963), On the solution of incorrectly stated problems and method of regularization, *Dokl. Akad. Nauk. SSSR*, *151*(3), 501–504.

Thorne, P.W., Brohan, P., Titchner, H.A., et al. (2011). A quantification of uncertainties in historical tropical tropospheric temperature trends from radiosondes. J. Geophys. Res. 116, D12116.

Tsuda, T., Nishida, M., Rocken, C. and Ware, R. H.: A global morphology of gravity wave activity in the stratosphere revealed by the GPS Occultation data (GPS/MET), J. Geophys. Res., 105, 257–273, 2000.

Venet N., Fassò A., Finazzi F. (2017) On some anisotropic covariance functions on the sphere. In preparation.

Wackernagel, H. (2003), *Multivariate Geostatistics*, Springer.

Wunch, D. et al. (2010), Calibration of the Total Carbon Column Observing Network using aircraft profile data, *Atmos. Meas. Tech.,* 3, 1351–1362, doi:10.5194/amt-3-1351-2010.

Yokota, T. et al. (2009), Global concentrations of CO2 and CH4 retrieved from GOSAT: first preliminary results, *SOLA*, 5, 160–163.