

1 Using reference radiosondes to characterise NWP model uncertainty for improved satellite
2 calibration and validation.

3

4 Fabien Carminati¹, Stefano Migliorini¹, Bruce Ingleby², William Bell², Heather Lawrence², Stuart
5 Newman¹, James Hocking¹, and Andrew Smith¹

6

7 ¹Met Office, Exeter, EX1 3PB, UK

8 ²ECMWF, Reading, RG2 9AX, UK

9

10 Abstract

11 The characterisation of errors and uncertainties in numerical weather prediction (NWP) model fields
12 is a major challenge that is addressed as part of the Horizon 2020 Gap Analysis for Integrated
13 Atmospheric ECV CLimate Monitoring (GAIA-CLIM) project. In that regard, observations from the
14 GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN) radiosondes are
15 being used at the Met Office and European Centre for Medium-range Weather Forecasts (ECMWF)
16 to assess errors and uncertainties associated with model data.

17 The software introduced in this study and referred to as the GRUAN Processor has been developed
18 to collocate GRUAN radiosonde profiles and NWP model fields, simulate top-of-atmosphere
19 brightness temperature at frequencies used by space-borne instruments, and propagate GRUAN
20 uncertainties in that simulation. A mathematical framework used to estimate and assess the
21 uncertainty budget of the comparison of simulated brightness temperature is also proposed.

22 One year of GRUAN radiosondes and matching NWP fields from the Met Office and ECMWF have
23 been processed and analysed for the purposes of demonstration of capability. We present
24 preliminary results confirming the presence of known biases in the temperature and humidity
25 profiles of both NWP centres. The night-time difference between GRUAN and Met Office (ECMWF)
26 simulated brightness temperature at microwave frequencies predominantly sensitive to
27 temperature is on average smaller than 0.1K (0.4K). Similarly, this difference is on average smaller
28 than 0.5K (0.4K) at microwave frequencies predominantly sensitive to humidity.

29 The uncertainty estimated for the Met Office – GRUAN difference ranges from 0.08 to 0.13K for
30 temperature sensitive frequencies and from 1.6 to 2.5K for humidity sensitive frequencies. From the
31 analysed sampling, 90% of the comparisons are found to be in statistical agreement.

32 This initial study has the potential to be extended to a larger collection of GRUAN profiles, covering
33 multiple sites and years, with the aim of providing a robust estimation of both errors and
34 uncertainties of NWP model fields in radiance space for a selection of key microwave and infrared
35 frequencies.

36

37 1. Introduction

38 Space-borne observational datasets are EOS key-components that have led to significant advances in
39 climate and weather applications (Joo et al., 2013; Bauer et al., 2015; Hollmann et al., 2013; Bojinski
40 et al., 2014), and therefore must be subject to high standards of calibration and validation to meet
41 user requirements. As part of an overall strategy for a harmonised and improved instrument
42 calibration, the World Meteorological Organisation (WMO), Coordination Group for Meteorological
43 Satellite (CGMS), and Global Space-based Inter-Calibration System (GSICS) have advocated the need
44 to tie the measurements to absolute references and primary standards (WMO, 2011¹; GSICS, 2015²).
45 In most cases however, commonly used validation techniques, as discussed by Zeng et al. (2015) and
46 Loew et al. (2017), do not yet provide a full metrological traceability.

47 For a full metrological traceability and uncertainty quantification, Green et al (2018) suggested
48 mirroring the measurement protocols as described by Immler et al (2010). Accordingly, consistency
49 between two independent measurements, m_1 and m_2 , is achieved when:

$$|m_1 - m_2| < k \sqrt{\sigma^2 + u_1^2 + u_2^2} \quad (1)$$

50 where u_1 and u_2 are the total uncertainties associated with m_1 and m_2 , respectively. σ represents the
51 intrinsic uncertainties of the comparison. In the case of a comparison between radiosonde and
52 satellite observations for example, this term can represent the collocation uncertainty (Calbet et al.,
53 2017). k is a coverage factor expanding the confidence interval for normally distributed error
54 probability.

55
56 In this paper, we use the terms error and uncertainty as described in the International Vocabulary of
57 Metrology (VIM) (JCGM, 2012³). The uncertainty is described in the VIM as a non-negative
58 parameter characterizing the dispersion of the quantity values being attributed to the quantity
59 intended to be measured, based on the information used. It is emphasized that all components of
60 the uncertainty contribute to this dispersion. This includes systematic effects arising from, for
61 example, corrections or reference standards. If a systematic effect is unknown it is unaccounted in
62 the uncertainty budget but contributes to the error.

63 The error is defined as the measured quantity value minus the unknown true value and may be
64 composed of a random and a systematic component.

65
66 For satellite data, pre-launch calibration characteristics are often provided by the instrument
67 manufacturer or space agency. However at launch, an uncertainty chain that may have been
68 metrologically traceable during the laboratory calibration phase can become compromised due to
69 changes in the spacecraft during the launch process itself as well as changes in the satellite

¹ https://library.wmo.int/opac/doc_num.php?explnum_id=3710

² http://www.wmo.int/pages/prog/sat/documents/GSICS-RD002_Vision.pdf

³ <https://www.bipm.org/en/publications/guides/vim.html>

70 environment in orbit compared to the laboratory testing. Furthermore, the instruments also degrade
71 over time, sometimes in quite a complex manner. These issues coupled with the current lack of true
72 on-board traceable references makes creating a metrologically traceable uncertainty chain difficult
73 for the satellite data record.

74 This aspect is being addressed by the Fidelity and Uncertainty in Climate Data Records from Space
75 (FIDUCEO) project (<http://www.fiduceo.eu/>). The project aims to develop Fundamental Climate Data
76 Records (FCDR) by reprocessing existing observations from raw satellite data to geolocated and
77 calibrated radiances with traceable uncertainties from a set of different references at the pixel level.
78 The uncertainty characterisation will account for the physical basis of the sensing process, the on-
79 board calibration system, and an estimate for the uncertainties arising from the processing.

80

81 The (re)assessment of historical, well-established, and new space-borne instruments using data
82 assimilation systems has become, over the past decade, common practice in numerical weather
83 prediction (NWP) centres (Bell et al., 2008; Zou et al., 2011; Bormann et al., 2013; Lu and Bell, 2014).
84 NWP models offer an interesting framework for the assessment of observational datasets due to a
85 physically constrained, continuous, global, and homogeneous representation of the atmosphere. An
86 optimal estimation of the state of the atmosphere is routinely performed in data assimilation
87 systems by blending information from a large volume of observations (space-borne, air-borne, and
88 ground-based) with a short-range forecast. Diagnostics are calculated in satellite observation space,
89 typically in brightness temperature, thanks to the radiative transfer models used by data assimilation
90 systems (Saunders et al., 2018). This forward approach is better posed than the inverse problem,
91 that is to say comparing model geophysical fields to retrieved satellite profiles, since multiple
92 atmospheric states can provide solutions to the retrieval, introducing further uncertainty. NWP
93 representation of atmospheric temperature and humidity fields is of sufficient quality to enable the
94 characterisation of subtle biases in satellite observations as demonstrated in the work referenced
95 herein. Loew et al. (2017) reported model fields uncertainties in the satellite observation space
96 ranging from 0.05 to 0.2K at frequencies principally sensitive to mid-tropospheric and lower
97 stratospheric temperature, and from 1 to 2K at frequencies sensitive to mid and upper tropospheric
98 humidity. However, those estimations arise from sensitivity studies and not from robust uncertainty
99 analyses. Stochastic approaches, based on ensemble forecasting techniques, have been used to
100 estimate forecast uncertainties, but with the caveat that they do not represent the systematic model
101 biases (Leutbecher et al., 2017).

102 This lack of metrologically traceable characterisation has often hampered the recognition and
103 consideration of model-based assessment outside of the NWP context, especially at space agency
104 and instrument team levels. Key climate users can also benefit from this approach, which has begun
105 to find resonance in the climate community (e.g. Massonnet et al., 2016).

106

107 It is also worth noting that bias correction schemes are generally applied to observations, especially
108 satellite radiances, used in data assimilation systems. Corrections are performed with respect to the
109 model background or analysis depending of the chosen scheme. Although this works for theoretical
110 unbiased NWP models, real world data assimilation systems also use reliable observations whose

111 role is to anchor the analysis. These anchoring observations although they may be slightly biased
112 with respect to the truth are not corrected in the data assimilation system. As a result, background
113 and analysis are weighted by the average of the non-zero biases in the model and in the anchor
114 observations. Eyre (2016) however demonstrated that a risk inherent to bias correction schemes is a
115 decrease of the weight given to anchor observations when the number of assimilated bias-corrected
116 observations increases, which results in model background and analysis to be increasingly weighted
117 toward the bias in the model. To avoid this situation, the Eyre (2016) suggests that correction should
118 be derived from areas where NWP model bias are expected to be small, along with the use of
119 numerous anchor observations.

121 ~~In this paper, we use the terms *error* and *uncertainty* as described in the International Vocabulary of~~
122 ~~Metrology (VIM) (JCGM, 2012⁴). The uncertainty is described in the VIM as a non-negative~~
123 ~~parameter characterizing the dispersion of the quantity values being attributed to the quantity~~
124 ~~intended to be measured, based on the information used. It is emphasized that all components of~~
125 ~~the uncertainty contribute to this dispersion. This includes systematic effects arising from, for~~
126 ~~example, corrections or reference standards. If a systematic effect is unknown it is unaccounted in~~
127 ~~the uncertainty budget but contributes to the error.~~

128 ~~The error is defined as the measured quantity value minus the unknown true value and may be~~
129 ~~composed of a random and a systematic component.~~

131 The Gap Analysis for Integrated Atmospheric ECV CLimate Monitoring (GAIA-CLIM) project (Thorne
132 et al., 2017) aims to address those challenges by improving the use of in-situ observations to
133 rigorously characterise a set of atmospheric Essential Climate Variables (ECVs) derived from satellite
134 observations as well as the geolocated and calibrated spectral radiances (level 1b) from which these
135 quantities are derived (<http://www.gaia-clim.eu/>). The work presented here is embedded in that
136 framework and focuses on developing NWP as a comprehensive reference by establishing
137 traceability for the model fields through comparison with traceable comparator data.

138 The NWP model error and uncertainty budget can be expressed as a function of four main
139 contributions:

- 140 a) The error and uncertainty in NWP temperature and humidity fields mapped to observation
141 space (brightness temperature).
- 142 b) The error and uncertainty in the underlying radiative transfer modelling, including biases
143 between fast radiative transfer models commonly used in NWP and reference line-by-line
144 models, fundamental spectroscopic uncertainty, and surface emissivity uncertainty.
- 145 c) The error and uncertainty due to scale mismatch. This encompasses the different scale at
146 which observation and model are resolved, and the scale of natural variability that is,
147 especially for humidity, much smaller than both observation and model scales.

⁴ <https://www.bipm.org/en/publications/guides/vim.html>

148 d) The error and uncertainty due to residual cloud. Clear-sky scenes are generally preferred
149 because simulated cloudy radiances are affected by uncertainties in model representation of
150 cloud amounts and the absorption and scattering properties of hydrometeors.

151 This study aims to address the first contribution. To that end, the Met Office and European Centre
152 for Medium-range Weather Forecasts (ECMWF) models are compared to radiosondes from the
153 Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) in a stand-alone
154 module based on the core radiative transfer modelling capability of the fast radiative transfer model
155 RTTOV and the Radiance Simulator (both available on <http://www.nwpsaf.eu/>). This software,
156 referred to as the GRUAN Processor, enables the collocation of geophysical fields and simulation of
157 top-of-atmosphere (TOA) brightness temperatures (Tb) from radiosondes and NWP models, with
158 GRUAN uncertainties propagated into the radiative transfer calculation.

159

160 Section 2 introduces the datasets used for this study, namely GRUAN radiosondes and the NWP
161 models from the Met Office and ECMWF. Sections 3 and 4 describes the GRUAN Processor
162 functionality and presents an illustrative case study. A methodology statistically assessing the
163 uncertainties is presented in section 5. Section 6 concludes the study.

164

165 2. Datasets

166 2.1. GRUAN

167 With 17 sites across the world (including two inactive sites in the Pacific), GCOS is building on
168 existing infrastructures to develop a reference network for upper-air observations
169 (<http://www.gruan.org/>). GRUAN aims to provide long-term high-quality measurements of ECVs
170 with vertically resolved uncertainty estimates. To meet the strict criteria for reference
171 measurements, GRUAN data also includes a comprehensive collection of metadata and
172 documentation of correction algorithms.

173 To date, only the Vaisala RS92 radiosonde is used to produce the GRUAN certified products (Sommer
174 et al., 2016), referred to as RS92 GRUAN Data Product Version 2 (RS92-GDP), but a new product
175 based on the Vaisala RS41 is in preparation. The RS92 GRUAN processing is documented by Dirksen
176 et al (2014). This includes the correction of the radiosonde systematic errors, due to mainly solar
177 radiation, and the derivation of the uncertainties for temperature, humidity, wind, pressure, and
178 geopotential height. The total uncertainty budget accounts for correlated and uncorrelated
179 contributions of both random sources of uncertainty and uncertainties from systematic error
180 corrections, and it is expressed as the root sum square of all contributions. The uncertainty related
181 to the short wave radiation correction (used in the temperature uncertainty budget), the correlated
182 uncertainty related to systematic error corrections, and uncorrelated uncertainty (standard
183 deviation) derived from the GRUAN processing are available in the RS92-GDP files, in addition to the
184 total uncertainty of each variables. However, not all correlated and uncorrelated components are
185 independently available (albeit used in the calculation of the total uncertainty) and some sources of
186 partially correlated uncertainty are not yet modelled in GRUAN algorithms (e.g. the pendulum

187 motion of the radiosonde under the balloon). Therefore, only the total uncertainties of temperature,
188 humidity, and pressure are considered in this study.

189 The results presented in this preliminary study focus on the profiles from Lindenberg (LIN), GRUAN
190 lead centre, Germany (52.21°N, 14.12°E) for the year 2016.

191

192 2.2. Met Office NWP

193 Met Office model data files are extracted from the Managed Archive Storage System (MASS) and
194 only $\pm 5^\circ$ latitude and longitude around the GRUAN launch site is kept to limit the data volume. For
195 LIN, the model fields cover the area 47.109-57.109°N and 9.0234-19.102°E. Each model data file
196 contains four time steps starting at T+0, the analysis, and three successive 3-hour forecasts referred
197 to as T+3, T+6, and T+9. The Met Office data assimilation system is a hybrid 4-dimensional
198 variational analysis (4D-Var) with 6-hour time window (Lorenc et al., 2000; Rawlins et al., 2007). Four
199 analyses (and their successive forecasts) are available every day at 00:00, 06:00, 12:00, and 18:00
200 Coordinated Universal Time (UTC). Assimilated satellite radiances are corrected with a variational
201 bias correction similar to the scheme described by Auligné et al. (2007). The operational forecast
202 model in 2016 had a resolution of approximately 17km at mid-latitudes for 70 levels from surface to
203 80km (N768L70). The radiative transfer calculation was performed in 2016 by the fast radiative
204 transfer model RTTOV version 9 (Saunders et al., 1999, 2007).

205 In the Met Office NWP system, the interpolation of background fields is performed twice, once for
206 all observations and later just for those observations to be assimilated. The radiosonde profiles are
207 averaged over the vertical model layers. Latitude, longitude, and time at each level are used in the
208 first interpolation of background values, but fixed coordinates are used in the latter interpolation. A
209 bias correction of radiosonde profiles is in place on a per station basis but is generally not applied
210 where RS92 are used. As noted by Ingleby and Edwards (2015), radiation corrections are now often
211 directly applied by the radiosonde manufacturer such as Vaisala, which reduces the need for
212 correction in NWP system. Bias correction and quality controls operationally applied to radiosonde
213 at the Met Office are detailed in the appendix 1 of Ingleby and Edwards (2015).

214

215 2.3. ECMWF NWP

216 ECMWF data are extracted from the Meteorological Archival and Retrieval System (MARS⁵). Data
217 come from the operational data class atmospheric model long window 4Dvar stream (see MARS
218 documentation for details). The covered area is the same as for the Met Office. Each model data file
219 contains six time steps of three hours starting from T+0 to T+15. The ECMWF analysis/forecast
220 system is documented by ECMWF⁶. A cubic octahedral reduced Gaussian grid is currently used with
221 a resolution of TCo1279 (horizontal grid spacing of about 9 km) and with 137 levels in the vertical.
222 Note that from February 2006 until June 2013, there were 91 vertical levels, and from January 2010
223 until March 2016 a linear reduced Gaussian grid was used with a horizontal spacing of around 16 km.

⁵ <https://software.ecmwf.int/wiki/display/UDOC/MARS+user+documentation>

⁶ <https://www.ecmwf.int/en/forecasts/documentation-and-support>

224 Data assimilation uses incremental 4D-Var (Courtier et al., 1994) with a 12-hour window, the
225 nominal 00:00 UTC analysis uses data from 21:00 UTC to 09:00 UTC. Forecasts and ensembles are
226 run twice daily from early-delivery 6-hour window 4D-Var analyses (Haseler, 2004). Flow-dependent
227 ensemble information from the ECMWF ensemble of data assimilations is incorporated into the
228 modelling of background-error covariances (Bonavita et al., 2016). Satellite radiative transfer
229 calculations use the fast radiative transfer model RTTOV version 11.2 (Hocking et al., 2015) has been
230 used operationally since May 2015 (Lupu and Geer, 2015). Variational bias correction of satellite
231 radiances (and, unlike the Met Office scheme, aircraft temperatures) is based on Dee (2004) and
232 Auligné et al (2007).~~Variational bias correction of satellite radiances (and aircraft temperatures) is~~
233 ~~based on Dee (2004) and Auligné et al (2007).~~

234 The treatment of radiosondes in the ECMWF system differs from that of the Met Office in that there
235 is no average on model levels and each level is treated as a point value. In addition, the balloon drift
236 in space and time was not accounted for in 2016 (i.e. the ascension was assumed instantaneous and
237 vertical). The treatment of the radiosonde drift (from BUFR reports) has been introduced in the
238 operational system in 2018 (Ingleby et al., 2018). Also in contrast to the Met Office, radiosondes at
239 ECMWF are bias corrected for temperature and humidity. The correction, described by Agusti-
240 Panareda et al (2009), uses monthly statistics of background departure based on night-time RS92
241 and is applied as a function of radiosonde type, pressure, and solar elevation angle.

242

243 3. Processor design

244 The GRUAN Processor, a software based on the NWP Satellite Application Facility (SAF) Radiance
245 Simulator (Smith, 2017), is designed to collocate NWP model fields from the Met Office or ECMWF
246 with radiosondes from the GRUAN network and simulate TOA Tb from those collocated profiles. The
247 simulations are performed at frequencies used by meteorological space-borne instruments and
248 supported by RTTOV. Figure 1 illustrates the Processor top-level design with its main processing
249 steps.

250

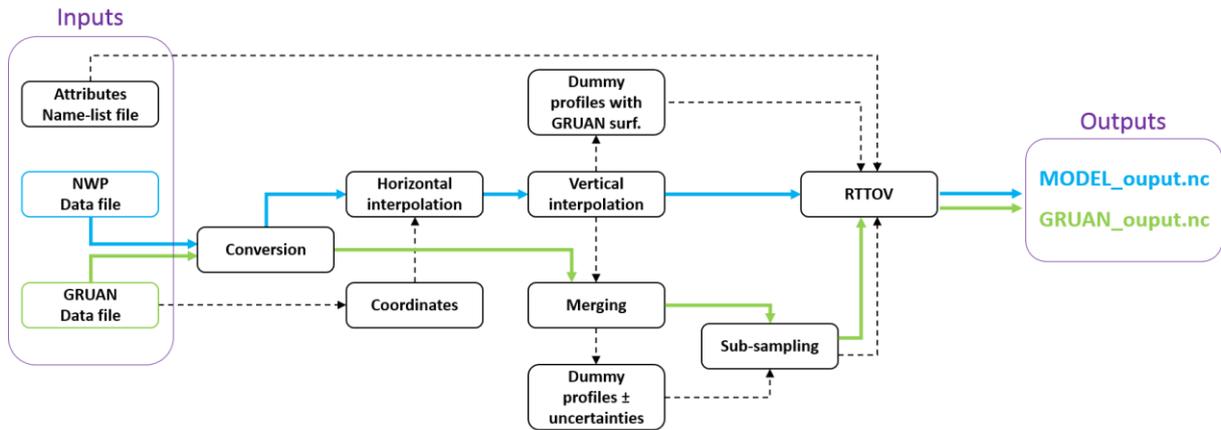
251 3.1. Inputs

252 The Processor requires as input a GRUAN and a model data file. Supported products are GRUAN
253 RS92-GDP, Met Office Unified Model (UM) Fieldfiles (or PP files, see Smith (2017)), and ECMWF GRIB
254 files. Both model file types must contain the minimum set of required variables as described by
255 Smith (2017) for the Radiance Simulator. Processing options and RTTOV attributes are provided via a
256 text file read by the Processor. This file includes the instrument characteristics (e.g. channels) to be
257 simulated and output options (output in unit of radiances or Tb for example). Optionally, RTTOV bias
258 and root mean square error (rmse) estimated from comparisons between RTTOV and line-by-line
259 model calculations, as provided by NWP SAF⁷, can be written to the output files. Finally, an option
260 allows to opt for a model-radiosonde collocation following the balloon drift (in space and time, see

⁷ <https://www.nwpsaf.eu/site/software/rttov/download/coefficients/comparison-with-lbl-simulations/>

261 section 3.3) or assuming no drift. Note that all collocations presented in this paper account for the
 262 radiosonde drift.

263



264

265 Figure1: GRUAN Processor top-level design. Solid arrows show the main processing steps from input
 266 (blue for NWP model data and green for GRUAN data) to output. Dashed arrows represent the
 267 internal processing.

268

269 3.2. Conversion

270 The conversion step ensures that both model and GRUAN variables (e.g. temperature or humidity)
 271 are expressed in the same units and that those units are compatible with RTTOV (see section 3.5).
 272 Two main types of conversion are supported: temperature from potential temperature and specific
 273 humidity from relative humidity.

274

275 Model data files may sometimes contain potential temperature instead of temperature profiles, as is
 276 the case for the Met Office. Temperature (T) is therefore calculated as a function of potential
 277 temperature (θ) and pressure (P) as follows:

$$278 \quad T = \theta \left(\frac{P}{P_0} \right)^\kappa \quad (2)$$

278

279 where P_0 is the standard reference pressure equal to 1000hPa and κ the ratio of the gas constant of
 280 air to the specific heat capacity at constant pressure.

281 Similarly, it is worth noting that model data files may not directly contain pressure profiles (e.g. in
 282 ECMWF files) or the pressure may be expressed on a different set of levels with respect to other
 283 variables (e.g. Met Office files). In both cases however, the pressure on model levels can be
 284 calculated from coefficients provided in the model data files.

285

286 The expression of humidity also needs to be harmonised. GRUAN provides profiles of relative
 287 humidity (RH), whereas the humidity from both NWP models is expressed in specific humidity (q), in
 288 units $\text{kg}\cdot\text{kg}^{-1}$. GRUAN RH is converted to q as follows:

$$q = \frac{\varepsilon RH e_s}{(P - (1 - \varepsilon) RH e_s)} \quad (3)$$

289

290 where ε is the ratio of the molecular weight of water vapour to the molecular weight of dry air and
 291 e_s the saturation vapour pressure over liquid. For consistency with GRUAN and Vaisala processing, e_s
 292 is expressed as defined by Hyland and Wexler (1983), such that:

$$\ln(e_s) = \frac{C_1}{T} + C_2 + C_3 T + C_4 T^2 + C_5 T^3 + C_6 \ln(T) \quad (4)$$

293

294 with:

295 $C_1 = -5.8002206 \times 10^3$

296 $C_2 = 1.3914993 \times 10^0$

297 $C_3 = -4.8640239 \times 10^{-2}$

298 $C_4 = 4.1764768 \times 10^{-5}$

299 $C_5 = -1.4452093 \times 10^{-8}$

300 $C_6 = 6.5459673 \times 10^0$

301 for e_s in Pa and T in K.

302

303 3.3. Interpolations

304 The GRUAN Processor generates a set of model profiles (i.e. one profile per variable), on model
 305 levels, interpolated in space and time along the radiosonde path, which are then vertically
 306 interpolated on a fixed set of 278 levels as follows.

307 First, model fields are linearly interpolated at the radiosonde coordinates (latitude-longitude-time),
 308 weighted by the distance to the eight closest grid points. Therefore, for an observation at the
 309 coordinate $p=[x_p, y_p, z_p]$, as illustrated on figure 2, in a cube of vertices $[(x,y,z), (x+dx,y,z), (x,y+dy,z),$
 310 $(x,y,z+dz), (x+dx,y+dy,z), (x+dx,y,z+dz), (x,y+dy,z+dz), (x+dx,y+dy,z+dz)]$, where dx and dy represent
 311 the grid point interval in latitude and longitude, respectively, and dz the interval between the time
 312 $T+n$ and $T+(n+3)$, with associated field values F_p and $[F_{000}, F_{100}, F_{010}, F_{001}, F_{110}, F_{101}, F_{011}, F_{111}]$,
 313 respectively, F_p is calculated as follows:

$$\begin{aligned}
F_p = & F_{000}(1 - w_x)(1 - w_y)(1 - w_z) \\
& + F_{100}w_x(1 - w_y)(1 - w_z) \\
& + F_{010}(1 - w_x)w_y(1 - w_z) \\
& + F_{001}(1 - w_x)(1 - w_y)w_z \\
& + F_{101}w_x(1 - w_y)w_z \\
& + F_{011}(1 - w_x)w_yw_z \\
& + F_{110}w_xw_y(1 - w_z) \\
& + F_{111}w_xw_yw_z
\end{aligned} \tag{5}$$

314

315 where w_x , w_y , and w_z are the weights defined as:

$$w_x = (x_p - x)/dx \tag{6}$$

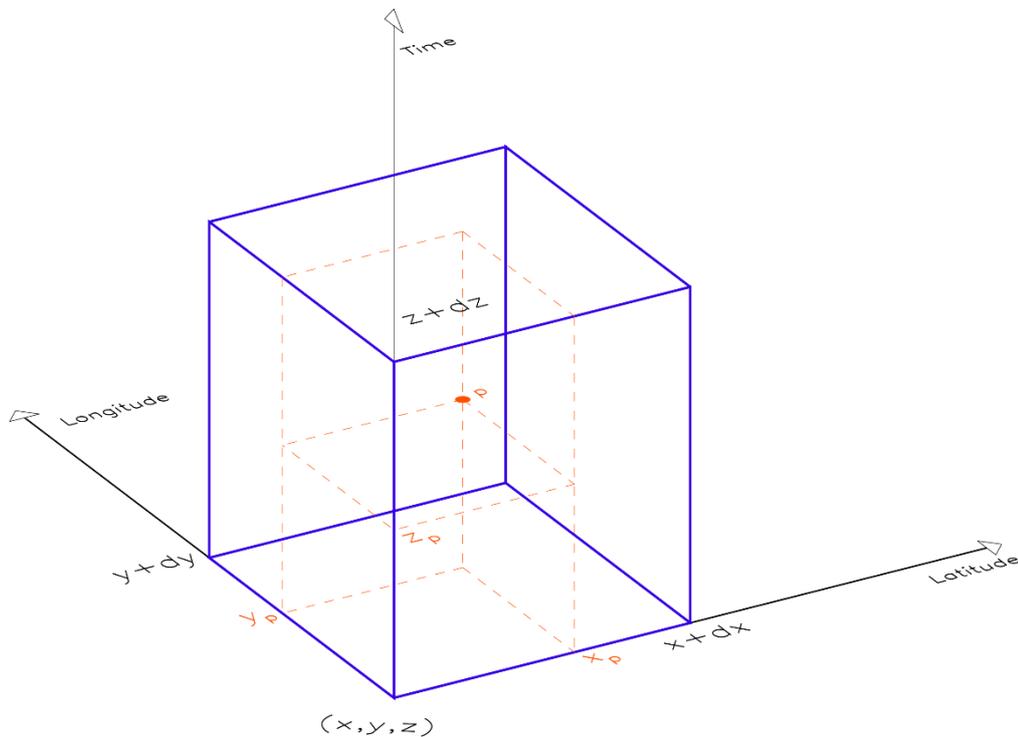
$$w_y = (y_p - y)/dy \tag{7}$$

$$w_z = (z_p - z)/dz \tag{8}$$

316

317 This operation is repeated along the radiosonde path with a time-step of 15 seconds based on the
318 radiosonde time profile. A unique model profile (one for each variable) is reconstructed by
319 combining the model fields from the pressure levels crossed by the radiosonde between two
320 consecutive interpolated model profiles.

321



322

323 Figure 2: illustration of an observation of coordinate (x_p, y_p, z_p) in a cube that vertices represent the
324 model latitude (x axis), longitude (y axis), and forecast time (z axis).

325

326 The reconstructed set of profiles is then interpolated on a fixed vertical grid of 278 pressure levels.
 327 The fixed grid, referred to as Processor grid (Pg), has been designed to have at least one Pg level
 328 between any two levels of the coarser model (Met Office or ECMWF) grid, referred to as Coarse grid
 329 (Cg). Therefore, for Pg of dimension n equal to 278 and Cg of dimension m (equal to 70 for the Met
 330 Office, 91 or 137 for ECMWF), the interpolation is calculated by weighting the fields with respect to
 331 the pressure via the interpolation matrix W of dimension $n \times m$, such as:

$$Pg = W Cg \quad (9)$$

332

333 where for the j^{th} pressure (P) level of Pg located between the i^{th} and $i+1^{\text{th}}$ levels of Cg :

$$Pg_j = W_{j1} Cg_1 + W_{j2} Cg_2 + \dots + W_{jm} Cg_m \quad (10)$$

$$W_{ji} = \frac{P_{i+1} - P_j}{P_{i+1} - P_i} \quad (11)$$

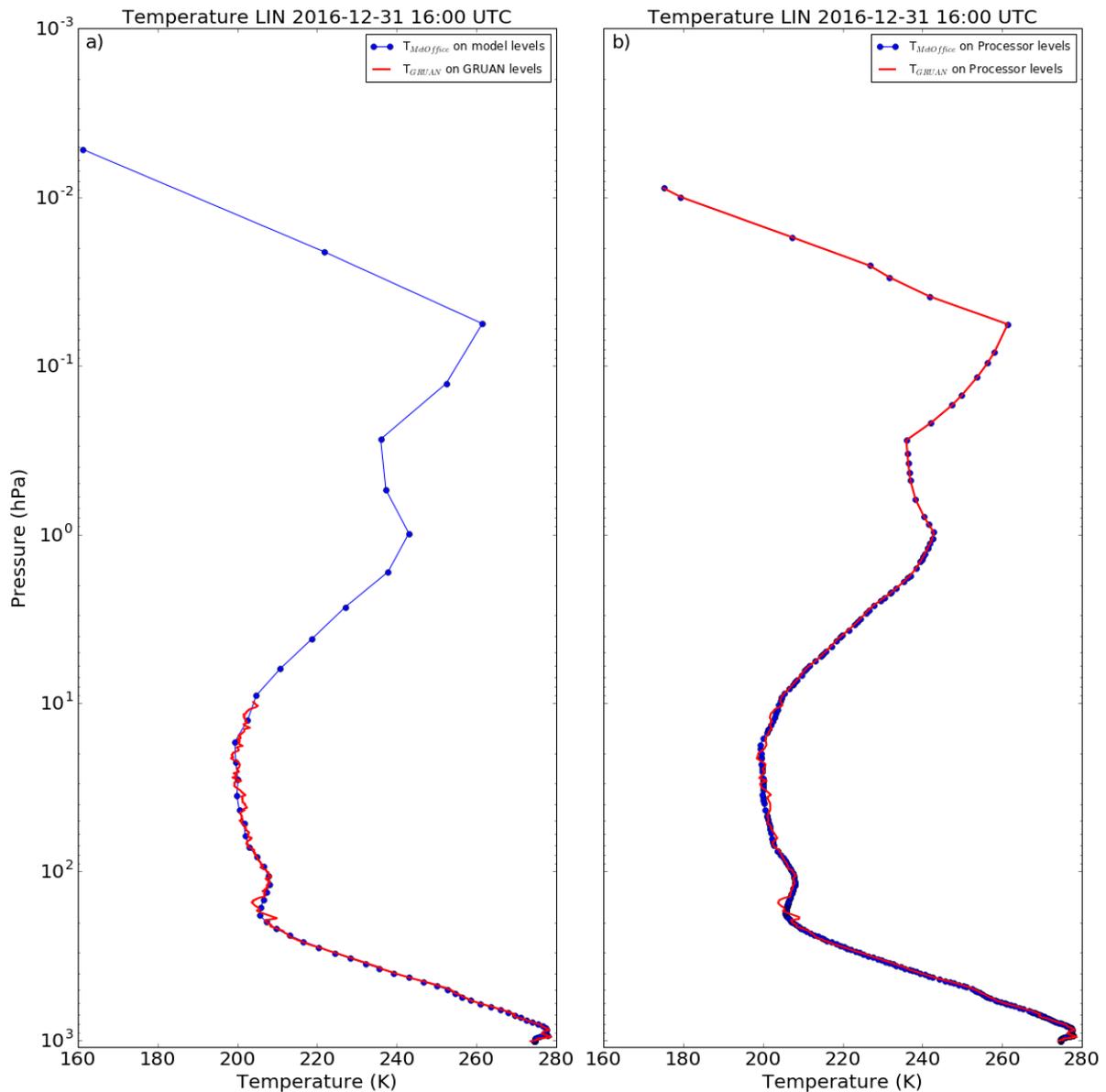
$$W_{ji+1} = 1 - W_{ji} \quad (12)$$

$$W_{jk} = 0 \text{ where } k \neq i, i + 1 \quad (13)$$

334

335 The vertical interpolation of model profiles as well as the subsampling of the radiosonde profiles
 336 (see section 3.4) to the Processor grid aims to provide a homogenised number of vertical levels on
 337 which the radiative transfer equation is calculated. Although the coarse model grid and the fine
 338 radiosonde grid could be directly used as input in RTTOV, it was observed that the lack of
 339 homogenisation between model and radiosonde profiles causes a bias in radiance space. It was
 340 therefore decided to interpolate the model profiles and provide a way to estimate the uncertainty
 341 associated to this interpolation (see section 5).

342 Figure 3 illustrates the change from a collocated Met Office temperature profile (LIN 31 December
 343 2016, 16:00 UTC) on model levels (70 levels) (a) to a collocated Met Office profile interpolated on
 344 the Processor grid (278 levels) (b).



345

346 Figure 3: (a) GRUAN temperature profile (red line) from Lindenberg on 31 December 2016, 16:00
 347 UTC as provided in the RS92-GDP data file with full GRUAN vertical resolution and collocated Met
 348 Office temperature profile (blue dotted line) on the model vertical levels. (b) GRUAN temperature
 349 profile subsampled at the Processor 278 pressure levels and merged with the Met Office profile
 350 above 9.8 hPa (red line) and collocated Met Office temperature profile interpolated on the Processor
 351 vertical levels (blue dotted line).

352

353 3.4. Merging and subsampling

354 A caveat of processing radiosonde profiles in RTTOV is the lack of information between the top of a
 355 profile (bursting point of the balloon) and the TOA. This is addressed by merging the radiosonde
 356 profiles with the model profiles above the last available point of the radiosonde. Note that this step
 357 occurs after the interpolation of the model profiles so that the upper merged part of the radiosonde
 358 and model profiles are strictly identical.

359 Similarly, RTTOV requires surfaces information: 2m temperature and humidity, surface pressure and
360 altitude, 10m wind (u and v components, used for microwave simulations over ocean), and skin
361 temperature. While GRUAN provides the surface pressure, temperature, relative humidity, and
362 altitude at launch site in all the data files, the skin temperature (T_{skin}^G) has to be derived from the
363 difference between the model skin (T_{skin}^M) and the 2m temperature (T_{2m}^M) applied to the GRUAN
364 surface temperature (T_{2m}^G) such as:

$$T_{skin}^G = T_{2m}^G + (T_{skin}^M - T_{2m}^M) \quad (14)$$

365

366 Although the 10m wind could be provided by the Vaisala wind profiles (available in GRUAN data
367 files) or calculated from GRUAN profiles of wind speed and direction, the chaotic rotation of the
368 radiosonde just after launch results in unreliable wind information near the surface. Therefore, the
369 model 10m wind (u and v components) is also merged with the GRUAN data. Note that 10m wind is
370 used to calculate the sea surface emissivity (for microwave simulations) and therefore only concerns
371 GRUAN sites on small island sites (i.e. La Reunion, Nauru, Manau, Ny-Ålesund, Graciosa, and
372 Tenerife).

373

374 In the raw RS92 data and GRUAN data, the samplings are provided every second but filtering reduces
375 the effective resolution of temperature to approximately 10s at low levels; the effective resolution
376 of humidity is similar but it is reduced to 40-50s at upper levels (Dirksen et al., 2014). As a result,
377 GRUAN profiles count several thousand levels in the vertical that need to be reduced to the number
378 of levels on the Processor grid. This is achieved with a subsampling of the radiosonde profiles to the
379 nearest levels for each of the 278 Processor pressure levels, at levels where data are available, with
380 the imposed constraint that the ratio radiosonde pressure by Processor pressure must be less than
381 0.1%.

382 The subsampling of GRUAN profiles has been preferred over layer-averaging or convolution
383 techniques for several reasons. First, we aimed to avoid all unnecessary modification of the GRUAN
384 profiles, used as reference in this study. Second, GRUAN uncertainties are vertically resolved and
385 their processing would have resulted in an information loss. Third, the aim of the Processor is to
386 evaluate uncertainties in radiance space. During the testing phase, we observed that neither the
387 choice of averaged layers nor sub-sampled levels significantly affects the calculation of radiative
388 transfer and the resulting brightness temperatures.

389 Fig. 3 shows the changes from a GRUAN temperature profile (LIN 31 December 2016, 16:00 UTC) as
390 provided in the RS92-GDP data file (5821 levels, from the surface to 9.88 hPa) (a) to a Processor
391 merged and subsampled profile (278 levels, from the surface to 0.008 hPa) (b).

392

393 3.5. RTTOV and uncertainties

394 The radiosonde and model profiles, both on the Processor vertical grid, and their respective surface
395 parameters are passed to RTTOV for the calculation of the TOA Tb. RTTOV version 11.3, currently
396 used by the GRUAN Processor, is documented by Hocking et al. (2015).

397 The surface emissivity depends on the surface type. For land and sea ice, the Processor uses a fixed
398 value, 0.95 and 0.92, respectively. Those estimates are potentially far from the truth, but any bias
399 introduced by fixed emissivity terms is expected to cancel out when the difference in simulated Tb is
400 calculated. Note that RTTOV allows the use of the emissivity atlases over land and sea ice, but this
401 option has not yet been investigated. Over sea, the surface emissivity is calculated by the RTTOV
402 FAST Emissivity Model (FASTEM) version 5 (Kazumori and English, 2015). Although the version 5 is
403 the default version, this can be changed in the input attribute file. It is worth noting here that
404 although the radiosonde may drift from above land to above sea (ice) (or the opposite), the surface
405 type can only be of one kind. The land surface type is typically used as most radiosonde launch sites
406 are well inside land masses. However, for the small island sites of La Reunion, Nauru, Manau, Ny-
407 Ålesund, Graciosa, and Tenerife, the radiosonde is expected to rapidly drift over sea and therefore
408 the sea surface type is used instead. The difference between sea and sea ice is controlled by the sea-
409 ice mask used by the NWP model.

410 The viewing angle is set by default to nadir (0°) for all simulations. However, different angles could
411 potentially be used for the purpose of better comparisons with real satellite data, for example.

412 All simulations assume clear sky scenes and uses RTTOV direct mode (ignoring the scattering) with
413 the cloud liquid water option off (data not available from GRUAN data file). It is acknowledged that
414 this may introduce discrepancies in the comparison between model and radiosonde in situations
415 where the radiosonde encounters one or several cloud layers. The brightness temperatures
416 calculated from the radiosonde data perturbed by the presence of clouds (e.g. peaks in the humidity
417 profile and to a lesser extent in the temperature profile) will differ from those calculated from the
418 model data that assume clear sky conditions. Because the RS92-GDP does not provide a cloud flag,
419 indirect screening may be required for fine comparisons. To that end, one can use the precipitable
420 water column from the RS92-GDP metadata as a proxy for cloud and or assume the presence of
421 cloud when the relative humidity exceeds a threshold value.

422 Finally, note that RTTOV interpolation mode (used to interpolate the input levels to the coefficient
423 levels for the calculation of the atmospheric optical depth, and then back from the coefficient levels
424 to the input levels for the calculation of the radiative transfer equation) uses the log-linear on
425 weighting function mode as described by Hocking et al. (2015). This is aimed to avoid a known issue
426 causing the oscillation of the temperature Jacobians.

427

428 It was observed that the interpolation of the model fields at the GRUAN launch site coordinates
429 results in large discrepancies, especially affecting surface parameters (surface pressure and
430 elevation) and the lower part of the profiles, when the local orography presents large variations at
431 scales of the same order as the model grid resolution. The interpolation, using the weighted average
432 of the four neighbouring grid points at a given forecast time may result in the model surface being
433 below or above the actual GRUAN launch site surface. A typical example is the site at La Réunion
434 where the radiosondes are launched from the Maïdo observatory at an altitude of 2200m, compared
435 to which the interpolation of the ECMWF model gives an altitude of 980m and the interpolation of
436 the Met Office model 0m. In Lindenberg by comparison, the radiosondes are launched from the
437 altitude of 103m while both models estimates the altitude to be 57m. To estimate the associated
438 error, a set of dummy model profiles are generated with the surface pressure forced to that

439 provided in the GRUAN metadata. If the model has a surface below that of the observations, the
440 model profiles are linearly interpolated and cut at the observed surface pressure, and the surface
441 parameters become those of the lowest level. If the model has a surface above that of the
442 observations, the model profiles are linearly extrapolated to the observed surface pressure, and the
443 model surface parameters become those of the new lowest level. The difference between the T_b
444 calculated from those modified profiles and the T_b calculated from the original profiles provides an
445 estimation of the associated error. This is referred to as u_surf_bt in the Processor output.

446

447 Finally, the GRUAN uncertainties are propagated into radiance space. As described by Calbet et al.
448 (2017), this can be achieved by multiplying the GRUAN profiles of uncertainty by the Jacobians
449 derived by RTTOV from the GRUAN atmospheric profiles, or by applying the radiative transfer to the
450 input atmospheric GRUAN profiles perturbed with their associated uncertainties. The GRUAN
451 Processor is designed to follow the second method although the first one will be further discussed in
452 section 5. In the Processor, two sets of perturbed profiles are created, one containing the GRUAN
453 profiles of temperature, pressure, and humidity, incremented by their respective total uncertainty
454 ($T+u_temp$, $P+u_press$, and $q+u_q$), and one containing the GRUAN profiles decremented by their
455 total uncertainty ($T-u_temp$, $P-u_press$, and $q-u_q$). The resulting brightness temperatures
456 calculated by RTTOV based on those two sets of perturbed profiles, referred to as Tb^+ and Tb^- ,
457 respectively, are compared to T_b , calculated with the unperturbed profiles, to estimate the
458 associated uncertainty in radiance space. The greatest difference between $|T_b - Tb^+|$ and $|T_b - Tb^-|$
459 is given in output as u_gruan_bt . Note that the eight combinations of sign that this approach can
460 allow have been tried during the test phase. The resulting uncertainty was not found significantly
461 different from that obtained with Tb^+ and Tb^- , but the processing time significantly increased. Tb^+
462 and Tb^- were therefore retained as the best compromise.

463 It should be noted that the simplified nature of this approach, which applies a perturbation of
464 constant sign in the vertical, ignores the complicated fluctuations that the propagation of
465 uncertainty via a multiplication by the Jacobians would induce (see section 5). Here, we assume that
466 the GRUAN profiles of uncertainty used to perturb the atmospheric profiles are fully correlated at all
467 levels. This assumption differs from the truth in that GRUAN total uncertainty consist of a root sum
468 square of correlated and uncorrelated components (Dirksen et al., 2014). Nevertheless, assuming a
469 fully correlated perturbation enables the estimation of the total GRUAN uncertainty upper bound in
470 radiance space allowed by this approach. The lower bound, not addressed in the GRUAN Processor,
471 can be obtained by assuming the uncertainty profiles completely uncorrelated, and lies close to zero
472 as demonstrated by Calbet et al. (2017).

473 Ideally, the correlated and uncorrelated components of GRUAN uncertainty should be treated
474 separately with, for example, the Monte Carlo method described in the Guide to the expression of
475 Uncertainty in Measurement (GUM) (JCGM, 2008⁸). However, those components are not all
476 independently available and it is currently not possible to differentiate them in the RS92-GDP. Note
477 that the radiosonde (random and/or systematic) errors are not provided. Instead, GRUAN algorithm
478 corrects the systematic errors in the radiosonde measurements, acknowledging that the correction

⁸ <https://www.bipm.org/en/publications/guides/gum.html>

479 is not perfect and introduces an associated residual uncertainty (accounted for in the total
480 uncertainty).

481 For completeness, perturbations to the surface parameters could be added to the total uncertainty
482 budget in radiance space, but GRUAN does not provide uncertainties associated with these
483 measurements. An alternative is discussed in section 5.

484

485 3.6. Outputs

486 For each pair of collocated radiosonde and NWP model fields, the GRUAN Processor generates two
487 outputs files in netcdf format. The first file contains the model-related fields including, but not
488 limited to, the profiles of temperature, humidity, and pressure on the Processor vertical grid, the
489 interpolation matrix W , the simulated brightness temperature, the temperature, humidity, and
490 pressure Jacobians, and a quality control flag (qcflags). Note that for successful simulations, qcflags is
491 equal to zero. The second file contains the GRUAN-related fields, including e.g. GRUAN atmospheric
492 profiles and associated uncertainties on the Processor vertical grid, the Jacobians, and the T_b and T_b
493 uncertainties estimated from the perturbed GRUAN profiles (u_gruan_bt).

494 Both files also contain metadata documenting the GRUAN Processor version number (here 6.2); the
495 NWP model, model validity time, and model version number; the simulated satellite name, platform,
496 and channel; the RTTOV version, RTTOV coefficients creation date, and bias and root mean square
497 error (when available); and the metadata available from the original RS92-GDP.

498

499 Note that some GRUAN Processor simulated brightness temperatures have been ingested into the
500 GAIA-CLIM Virtual Observatory (<http://gaia-clim.vo.eumetsat.int/>) for the purposes of visualisation,
501 manipulation, and extraction of collocated GRUAN-NWP-Satellite data.

502

503 4. Data analysis illustration

504 For illustration purposes, one year of collocated profiles and simulated T_b is presented. The dataset
505 corresponds to 1160 radiosondes launched from Lindenberg, Germany, in 2016, compared to the
506 Met Office and ECMWF models. T_b values have been simulated at the Advanced Technology
507 Microwave Sounder (ATMS) 22 channel frequencies, a microwave radiometer with sounding
508 capability in the oxygen band (53-57GHz), sensitive to tropospheric and lower stratospheric
509 temperature, and in the water vapour band (around 183GHz), sensitive to mid-to-upper
510 tropospheric humidity (Bormann et al., 2013).

511 The dataset is divided into two samples composed of day and night-time profiles, respectively. This
512 is aimed at discriminating the GRUAN profiles affected by solar radiation, the dominant source of
513 uncertainty according to Dirksen et al. (2014). All profiles with a solar zenith angle (calculated as a
514 function of latitude, longitude, and UTC) smaller (greater) than 90° at launch time is considered as
515 day (night) time. Note that for a refined analysis, the whole profile (not just launch time) should be
516 checked and only profiles with the sun below (or above) the horizon throughout should be used.

517 Note that for simplicity, no cloud screening is applied in this case study. This caveat may, as
518 suggested in the previous section, exacerbate the biases observed when comparing brightness
519 temperature simulated from radiosonde profiles and from model fields. Future work dedicated to
520 the in-depth analysis of model errors and uncertainties based on the Processor outputs will address
521 the impact of clouds on the simulations.~~Note that no cloud screening is applied in this study.~~

522 After screening, 573 pairs of GRUAN Processor outputs are available in daytime and 587 in night-
523 time for each model. The mean difference *NWP* – *GRUAN* in temperature, humidity, and simulated
524 *T_b* is shown in figures 4 (daytime) and 5 (night-time) together with the number of available
525 comparisons as a function of the pressure. Note that at pressures less than 10hPa, the data sampling
526 decreases rapidly as less balloons reach those levels. An arithmetic mean is used to average the
527 uncertainty over the sampling according to Immler et al. (2010) Eq. (4). For temperature and
528 humidity, the GRUAN total uncertainty as provided in the RS92-GDP is used (the relative humidity
529 uncertainty is converted into specific humidity uncertainty in the GRUAN Processor), while the
530 uncertainty in *T_b* shows the GRUAN uncertainties propagated in radiance space via the perturbation
531 of the atmospheric profiles. Note that the model uncertainty and the uncertainty associated with
532 the vertical interpolation are ignored in this section, but addressed in section 5.

533 It is important to note that both Met Office and ECMWF are operationally assimilating the
534 radiosonde profiles from the GCOS Upper Air Network (GUAN), which, in Lindenberg, are the same
535 as the GRUAN profiles but without the specific GRUAN processing (and without uncertainty
536 characterisation). Therefore, unlike the forecasts, the model analyses (T+0) are not completely
537 independent from the observations. However, this is not expected to affect significantly the mean
538 comparison as only about 5% of the profiles fall in the first time window (i.e. interpolation between
539 T+0 and T+3).

540

541 In Fig.s 4 and 5, the main feature for ECMWF is a 0.5K cold bias in the stratosphere (100-10hPa),
542 observed both day and night. This bias has also been detected by Shepherd et al. (2018) in the ERA5
543 reanalysis that are based on IFS cycle 41r2, the operational model in 2016. It is attributed to an
544 excess of moisture transported into the lower stratosphere, which lead to a cold bias by radiative
545 cooling. The model also presents a 50-75% wet bias peaking between 200 and 100hPa, slightly more
546 pronounced during the day. This is consistent with the results from Ingleby (2017) who showed a
547 similar behaviour for several kinds of radiosonde.

548 The Met Office model presents a persistent 0.2 to 0.5K cold bias at pressure greater than 300hPa
549 and a 0.25K warm bias between 200 and 100hPa seen at night-time only. This is consistent with
550 Ingleby and Edwards (2015) who showed similar features in the comparison between radiosondes
551 and the Met Office regional model covering the United Kingdom. The Met Office tropospheric
552 humidity fits generally the radiosonde profiles well but presents a 50-60% wet bias with a peculiar
553 double peak at 200 and 100hPa. A wet bias peaking at 300hPa was already observed by Ingleby et al.
554 (2013), the coarser vertical resolution used by the authors potentially explaining the different
555 pressure level at which the bias is observed. However, the second maximum (at 100hPa) seems to
556 be a new feature that appears in 2015 and persists in 2017 (not shown). This remains unexplained to
557 date.

558 In radiance space, it is important to distinguish between frequencies representative of the difference
559 *NWP – GRUAN* and those significantly affected by the surface and the mid to upper stratosphere
560 where the GRUAN profiles are merged with the model. Hence, ATMS frequencies sensitive to the
561 surface (23.8-54.4 and 88.2-165.5GHz, channel 1-7 and 16-17, respectively) and to the upper
562 stratosphere ($57.29 \pm 0.3222 \pm 0.022$ - $57.29 \pm 0.3222 \pm 0.0045$ GHz, channel 13-15, respectively) should be
563 considered with caution and not used for scientific applications. On the contrary, frequencies
564 sensitive to the upper tropospheric-lower stratospheric temperature (peaking between 300 and
565 20hPa) and to the mid tropospheric humidity (peaking between 650 and 350hPa) cover the same
566 vertical domain as the information provided by GRUAN. For those frequencies, ATMS channel
567 characteristics and mean Tb difference are provided in Table 1.

568

569 Table 1: Mean difference *NWP – GRUAN* in simulated Tb for ECMWF ($\Delta T_{b_{ECMWF}}$) and Met Office
570 ($\Delta T_{b_{MetOffice}}$) and 1σ standard deviation for ATMS channels 8-12 and 18-22 at day and night-time.

Channel	Frequency (GHz)	$\Delta T_{b_{ECMWF}} (1\sigma) (K)$		$\Delta T_{b_{MetOffice}} (1\sigma) (K)$	
		night	day	night	day
8	54.94	-0.08 (0.09)	-0.16 (0.10)	-0.00 (0.11)	-0.04 (0.12)
9	55.5	-0.15 (0.12)	-0.24 (0.13)	0.04 (0.13)	-0.02 (0.14)
10	57.29	-0.32 (0.18)	-0.45 (0.18)	0.01 (0.16)	-0.07 (0.20)
11	57.29 ± 0.217	-0.39 (0.21)	-0.54 (0.22)	-0.04 (0.20)	-0.16 (0.25)
12	$57.29 \pm 0.3222 \pm 0.048$	-0.34 (0.25)	-0.53 (0.27)	-0.09 (0.28)	-0.26 (0.31)
18	183.31 ± 7.0	0.35 (0.91)	0.25 (1.09)	0.02 (0.83)	-0.36 (1.02)
19	183.31 ± 7.0	0.37 (1.13)	0.15 (1.24)	-0.09 (1.03)	-0.48 (1.14)
20	183.31 ± 3.0	0.34 (1.31)	-0.01 (1.36)	-0.18 (1.22)	-0.61 (1.27)
21	183.31 ± 1.8	0.22 (1.48)	-0.29 (1.50)	-0.31 (1.42)	-0.81 (1.45)
22	183.31 ± 1.0	0.04 (1.61)	-0.61 (1.64)	-0.46 (1.57)	-1.01 (1.60)

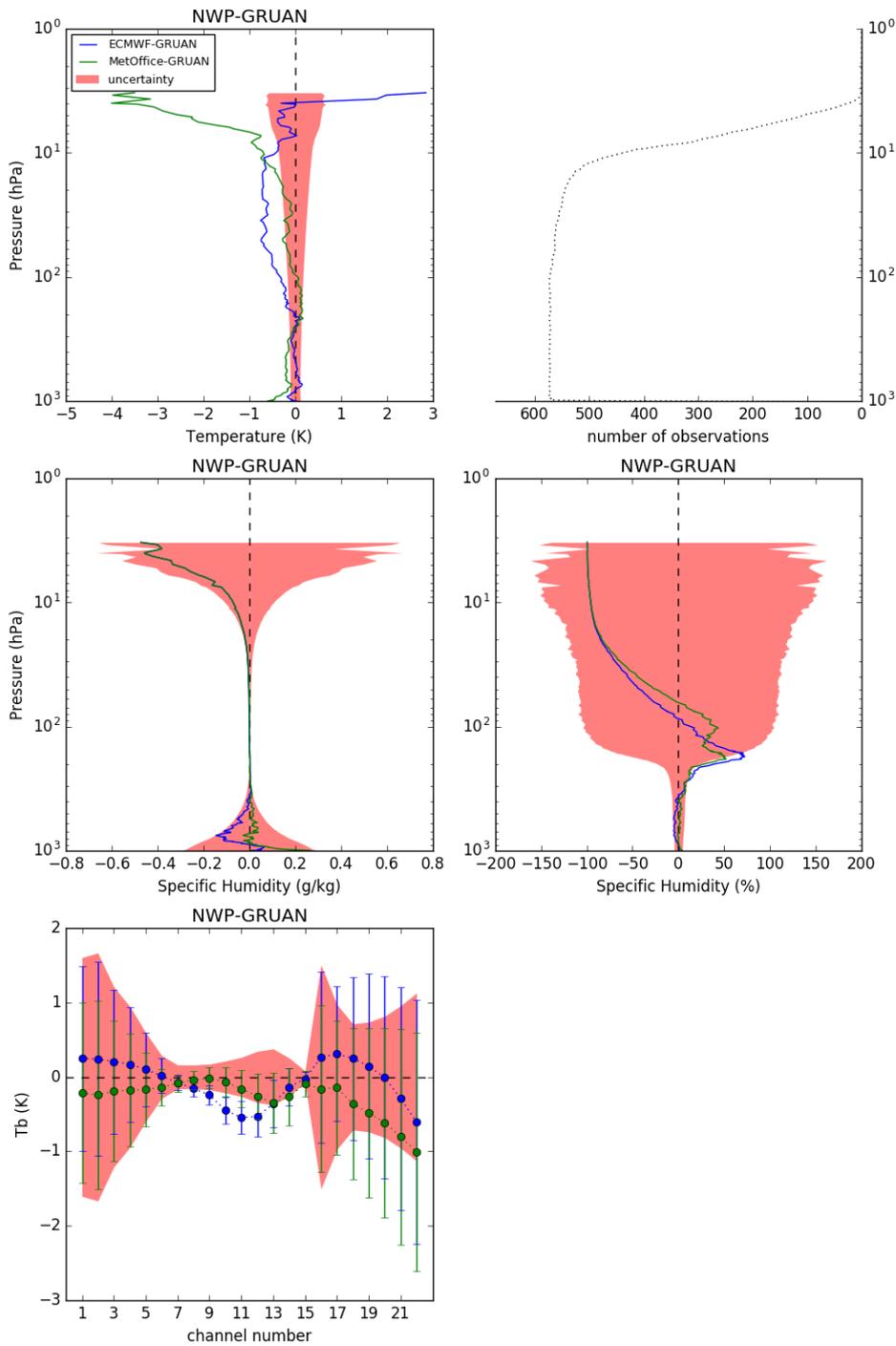
571

572 At frequencies sensitive to temperature (54-57GHz, channels 8-12), hereafter referred to as
573 temperature channels, the mean difference for ECMWF varies from -0.08 to -0.39K at night, mostly
574 outside GRUAN uncertainty (red shading, Fig. 5), reflecting the cold bias observed in the
575 stratosphere. Note that a difference greater than GRUAN uncertainty does not mean a statistical
576 disagreement since the uncertainty related to the model is unaccounted for (i.e. the total
577 uncertainty of the comparison as expressed in Eq. (1) is larger than the GRUAN uncertainty alone).
578 The difference is slightly larger in daytime (-0.16 to -0.54K). Similarly, the difference at frequencies
579 sensitive to humidity (around 183GHz, channels 18-22), hereafter referred to as humidity channels,
580 varies from 0.04 to 0.37K at night (-0.01 to -0.61K during the day), within GRUAN uncertainty.

581 The mean difference in Tb for the Met Office is always found within GRUAN uncertainty and varies
582 from -0.09 to 0.04K during the night (-0.02 to -0.26K in daytime) for the temperature channels and
583 from -0.46 to 0.02K during the night (-0.36 to -1.01K in daytime) for the humidity channels.

584 The standard deviation of the differences is similar for both centres and does not vary much from
 585 day to night.

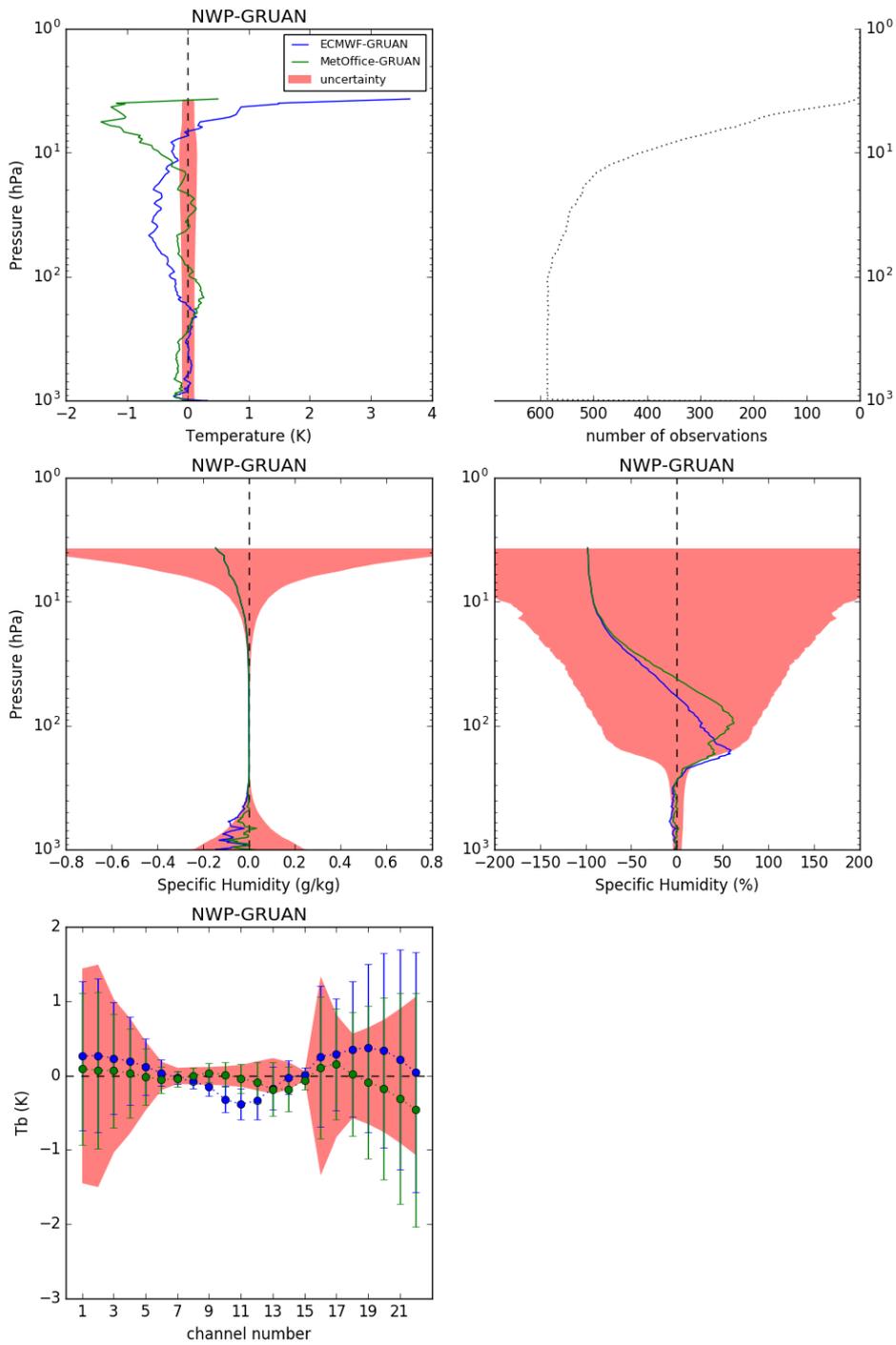
586



587

588 Figure 4: Mean difference ECMWF – GRUAN (blue) and Met Office – GRUAN (green) calculated from
 589 573 daytime collocation from Lindenberg in 2016. The temperature difference (top left) is expressed
 590 in K, the humidity difference is expressed in g.kg⁻¹ (middle left) and in percentage ($\frac{\text{NWP} - \text{GRUAN}}{\text{GRUAN}}$) (middle right), and the difference in simulated brightness temperatures for the 22 ATMS
 591

592 channels is expressed in K (bottom) with the 1σ standard deviation (vertical bars). The red shading
 593 shows GRUAN uncertainty. The number of observations is shown as a function of the pressure (top
 594 right).



595
 596 Figure 5: Same as figure 4 but for the 587 night-time collocations.

597

598 5. Comparison assessment

599 The previous section gives insights into the GRUAN uncertainty propagated in radiance space by the
600 GRUAN Processor. The approach offers a rapid but incomplete evaluation of the *NWP – GRUAN*
601 comparison, but several aspects are overlooked in the final budget, that for various reasons are not
602 part of the internal Processor processing. This includes: a) the uncertainty associated with surface
603 parameters, not provided in RS92-GDP and likely to change from station to station, b) the NWP
604 model uncertainty, often expressed as a covariance matrix and used in the data assimilation process
605 by the NWP centres, but not available in the input data files, and c) the uncertainty associated with
606 the vertical interpolation operated by the Processor for which estimation requires information on
607 the last two points.

608 In this section, a mathematical framework is elaborated to estimate a robust uncertainty budget for
609 the comparison between NWP fields and GRUAN observations, in radiance space, and statistically
610 assess this comparison. This includes uncertainties in the GRUAN observations, in the vertical
611 interpolation of the GRUAN Processor, and in the model fields. Note that, as previously mentioned,
612 any comparison to satellite radiances should also include other sources of uncertainty such as in the
613 underlying radiative transfer models and cloud detection. For this study, we focus on the
614 comparison to the Met Office model fields, but the same method could be applied to the
615 comparison with ECMWF fields.

616

617 We define \mathbf{x}_{rs} as the radiosonde profiles and \mathbf{x}_m as the model profiles (temperature, humidity, and
618 pressure, with a pressure coordinate). Note that \mathbf{x}_{rs} and \mathbf{x}_m are on different vertical grids. \mathbf{x}_{rs} is on
619 the GRUAN Processor vertical grid, composed of 278 levels, hereafter referred to as the fine grid (f),
620 subsampled from the original GRUAN profiles (noting that with a ratio radiosonde pressure by
621 Processor pressure less than 0.1%, the subsampling uncertainty is assumed negligible). \mathbf{x}_m is on the
622 model vertical grid, hereafter referred to as the coarse grid (c), as given in input.

623 Given H , the observation operator, we can express the simulated Tb as follows:

$$\mathbf{y}_{rs} \equiv H(\mathbf{x}_{rs}) \quad (15)$$

$$\mathbf{y}_m \equiv H(\mathbf{W}\mathbf{x}_m) \quad (16)$$

624 where \mathbf{W} is the interpolation matrix.

625 Eq.s (15) and (16) can be further expanded as a function of the profiles true value on the fine and
626 coarse grid, hereafter \mathbf{x}_f^t and \mathbf{x}_c^t , respectively, and the errors associated with the radiosonde and the
627 model fields, hereafter $\boldsymbol{\varepsilon}_{rs}$ and $\boldsymbol{\varepsilon}_m$, as follows:

$$\mathbf{y}_{rs} = H(\mathbf{x}_f^t + \boldsymbol{\varepsilon}_{rs}) \quad (17)$$

$$\mathbf{y}_m = H(\mathbf{W}\mathbf{x}_c^t + \mathbf{W}\boldsymbol{\varepsilon}_m) \quad (18)$$

628 with \mathbf{x}_c^t defined as $\mathbf{x}_c^t \equiv \mathbf{W}^* \mathbf{x}_f^t$ where an expression for \mathbf{W}^* , the pseudo-inverse of \mathbf{W} , is given in
629 Appendix B.

630 The comparison carried out in this study is in radiance space and the observation operator used to
631 simulate the brightness temperatures is identical for both radiosonde and model fields simulations.

632 For this reasons, we consider the radiance space as our reference and ignore any errors associated
 633 with observation operator, that would cancel out in the difference anyway since mainly systematic.
 634 Note that those errors need however to be taken into account if a simulated product is compared to
 635 real satellite observations.

636 Defining the vertical interpolation error $\boldsymbol{\varepsilon}_{int}$ as:

$$\boldsymbol{\varepsilon}_{int} \equiv \mathbf{W}\mathbf{x}_c^t - \mathbf{x}_f^t \quad (19)$$

637 Eq. (18) can be written as follows:

$$\begin{aligned} \mathbf{y}_m &= H(\mathbf{W}\mathbf{x}_c^t - \mathbf{x}_f^t + \mathbf{W}\boldsymbol{\varepsilon}_m + \mathbf{x}_f^t) \\ &= H(\mathbf{W}\boldsymbol{\varepsilon}_m + \boldsymbol{\varepsilon}_{int} + \mathbf{x}_f^t) \end{aligned} \quad (20)$$

638 Given \mathbf{H} , the Jacobian matrix provided by RTTOV and containing the partial derivatives of $\partial\mathbf{y}/\partial\mathbf{x}$
 639 (i.e. the change in radiance, $\partial\mathbf{y}$, for a change in the state vector, $\partial\mathbf{x}$), Eq.s (17) and (20) can be
 640 approximated, assuming small errors, as follows:

$$\mathbf{y}_{rs} \cong H(\mathbf{x}_f^t) + \mathbf{H}_{\mathbf{x}_f^t} \boldsymbol{\varepsilon}_{rs} \quad (21)$$

$$\mathbf{y}_m \cong H(\mathbf{x}_f^t) + \mathbf{H}_{\mathbf{x}_f^t}(\mathbf{W}\boldsymbol{\varepsilon}_m + \boldsymbol{\varepsilon}_{int}) \quad (22)$$

641 Therefore, the *NWP – GRUAN* comparison in radiance space is expressed as follows:

$$\begin{aligned} \delta\mathbf{y} &\equiv \mathbf{y}_m - \mathbf{y}_{rs} \\ &\cong \mathbf{H}_{\mathbf{x}_f^t}(\mathbf{W}\boldsymbol{\varepsilon}_m + \boldsymbol{\varepsilon}_{int} - \boldsymbol{\varepsilon}_{rs}) \end{aligned} \quad (23)$$

642 Assuming a complete uncorrelation between the interpolation error and those of the radiosonde
 643 and the model, the covariance of the difference is expressed as follows:

$$\mathbf{S}_{\delta\mathbf{y}} \equiv E\{(\delta\mathbf{y} - E\{\delta\mathbf{y}\})^T(\delta\mathbf{y} - E\{\delta\mathbf{y}\})\} \quad (24)$$

644 where E is the expectation operator. We can approximate Eq. (24) as:

$$\mathbf{S}_{\delta\mathbf{y}} \cong \mathbf{H}\mathbf{R}_f^{rs}\mathbf{H}^T + \mathbf{H}\mathbf{W}\mathbf{B}_c^m\mathbf{W}^T\mathbf{H}^T + \mathbf{H}\mathbf{S}_f^{int}\mathbf{H}^T \quad (25)$$

645 where \mathbf{R}_f^{rs} , \mathbf{B}_c^m , and \mathbf{S}_f^{int} are the error covariance matrices of GRUAN measurements (on the fine
 646 grid), the forecast (on the coarse grid), and the vertical interpolation (on the fine grid), respectively,
 647 as described below.

648

649 We first define the GRUAN covariance matrix. GRUAN does not provide a full covariance matrix with
 650 the measurements, therefore \mathbf{R}_f^{rs} is built as a diagonal matrix accounting for the different sources of
 651 uncertainty such as:

$$\begin{aligned}
\mathbf{H}\mathbf{R}_f^{rs}\mathbf{H}^T &= \mathbf{H}_T\mathbf{R}_T\mathbf{H}_T^T + \mathbf{H}_q\mathbf{R}_q\mathbf{H}_q^T + \mathbf{H}_P\mathbf{R}_P\mathbf{H}_P^T \\
&+ \mathbf{h}_{skinT}u_{skinT}^2\mathbf{h}_{skinT}^T + \mathbf{h}_{T2m}u_{T2m}^2\mathbf{h}_{T2m}^T \\
&+ \mathbf{h}_{q2m}u_{q2m}^2\mathbf{h}_{q2m}^T + \mathbf{h}_{P2m}u_{P2m}^2\mathbf{h}_{P2m}^T
\end{aligned} \tag{26}$$

652 where \mathbf{R}_T , \mathbf{R}_q , and \mathbf{R}_P are diagonal matrices whose diagonals are the square of GRUAN profiles of
653 total uncertainty for T , q (converted from RH), and P , respectively, on the Processor vertical grid;
654 u_{skinT} , u_{T2m} , u_{q2m} , and u_{P2m} the uncertainties associated with the surface parameters (i.e. skin
655 temperature, 2m temperature, 2m humidity, and 2m pressure) set to 0.3K, 0.3K, 0.04 RH, and
656 0.1hPa, respectively (Dr. S. Brickmann, DWD, private communication), estimated for the Lindenberg
657 site. \mathbf{H}_T , \mathbf{H}_q , and \mathbf{H}_P are the Jacobians of the temperature, humidity and pressure profiles,
658 respectively, and \mathbf{h}_{skinT} , \mathbf{h}_{T2m} , \mathbf{h}_{q2m} , and \mathbf{h}_{P2m} the Jacobians of the surface parameters.

659 \mathbf{R}_T , \mathbf{R}_q , and \mathbf{R}_P are diagonal which precludes a proper propagation of the correlation in radiance
660 space. In this suboptimal case, \mathbf{R}_f^{rs} , and by extension, $\mathbf{S}_{\delta y}$, the covariance of the comparison, will not
661 capture the most accurate representation of the uncertainty budget.

662

663 Then, we define the forecast error covariance matrix. For the purposes of this study, the forecast
664 covariance matrix from the operational Met Office Observation Processing System, a one-
665 dimensional variational analysis (1D-Var) performed ahead of the main variational process, is used
666 for \mathbf{B}_c^m . Alternatively, the forecast error covariance matrix can be estimated from an ensemble of
667 NWP profiles as described in Appendix A.

668

669 Finally, we define vertical interpolation covariance matrix. To estimate \mathbf{S}_f^{int} , the interpolation error
670 must be quantified.

671 From Eq. (19) we have:

$$\begin{aligned}
\boldsymbol{\varepsilon}_{int} &= \mathbf{W}\mathbf{W}^*\mathbf{x}_f^t - \mathbf{x}_f^t \\
&= (\mathbf{W}\mathbf{W}^* - \mathbf{I})\mathbf{x}_f^t
\end{aligned} \tag{27}$$

672 where the random vector \mathbf{x}_f^t , representing the true state on the fine grid, is assumed to have
673 mean $E\{\mathbf{x}_f^t\}$, the (unknown) mean model forecast profile on the fine grid, and covariance
674 $E\{(\mathbf{x}_f^t - E\{\mathbf{x}_f^t\})^T(\mathbf{x}_f^t - E\{\mathbf{x}_f^t\})\} \equiv \mathbf{B}_f^m$, the covariance of \mathbf{x}_f^t in model space on the fine grid. It
675 follows that we can express the covariance of the interpolation uncertainty as:

$$\begin{aligned}
\mathbf{S}_f^{int} &\equiv E\{(\boldsymbol{\varepsilon}_{int} - E\{\boldsymbol{\varepsilon}_{int}\})^T(\boldsymbol{\varepsilon}_{int} - E\{\boldsymbol{\varepsilon}_{int}\})\} \\
&= (\mathbf{W}\mathbf{W}^* - \mathbf{I})\mathbf{B}_f^m(\mathbf{W}\mathbf{W}^* - \mathbf{I})^T
\end{aligned} \tag{28}$$

676 Note that when the model grid coincides with the fine grid we have $\mathbf{W}^* = \mathbf{W}^{-1}$ and $\mathbf{S}_{int} = \mathbf{0}$ as
677 expected. Replacing \mathbf{W}^* by its form expressed in Appendix B we obtain:

$$\mathbf{S}_f^{int} = \mathbf{B}_f^m (\mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{B}_f^{m-1} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{B}_f^{m-1}) \quad (29)$$

678 Note that in practice (i.e. for numerical calculations) it is more convenient to use the form expressed
 679 in Eq. (28) to get \mathbf{S}_f^{int} as a symmetric and positive definite matrix.

680

681 This methodology has been applied to the 587 profiles of the night-time dataset described in the
 682 previous section. The covariances $\mathbf{S}_{\delta y}$ of each comparison as approximated in Eq. (25) have been
 683 averaged (arithmetic mean, hereafter $\overline{\mathbf{S}_{\delta y}}$) and the square root of the diagonal (i.e. the 1σ standard
 684 deviation of the comparison total uncertainty distribution) is shown in figure 6. In practice, we
 685 calculate $\mathbf{S}_{\delta y}$ as the sum of the covariance matrices of each variable: the surface measurements
 686 covariance (\mathbf{S}_{surf_rs}); the model surface covariance (\mathbf{S}_{surf_m}); the total humidity covariance
 687 (\mathbf{S}_{q_total}); the total temperature covariance (\mathbf{S}_{T_total}); and the GRUAN pressure covariance (\mathbf{S}_{P_rs}).
 688 The square root of their diagonal is also shown in figure 6. In addition, \mathbf{S}_{q_total} and \mathbf{S}_{T_total} can be
 689 further decomposed into the sum of the covariance matrices of each of their components: the
 690 GRUAN humidity and temperature covariance (\mathbf{S}_{q_rs} and \mathbf{S}_{T_rs}); the model humidity and
 691 temperature covariance (\mathbf{S}_{q_m} and \mathbf{S}_{T_m}); and the covariance of the vertical interpolation of the
 692 model humidity and temperature profiles ($\mathbf{S}_{q_m_int}$ and $\mathbf{S}_{T_m_int}$). The square root of their diagonal
 693 is also shown in figures 7 and 8.

694 Note that on some occasions, the Processor fine grid does not capture the lowermost or upper most
 695 model levels, which caused missing values in \mathbf{W} . The calculation has consequently been done, for
 696 those cases, on the remaining levels of \mathbf{W} . It is planned to refine the Processor grid in the future
 697 version in order to avoid such missing data in the interpolation matrix.

698

699 As expected, the surface components of the total uncertainty are dominant at frequencies where
 700 the radiance is sensitive to the surface (ATMS channels 1-7 and 16-17). Amongst them, the surface
 701 component from the model is the largest due to the low confidence in surface emission and
 702 properties. Channels with frequencies sensitive to temperature and humidity are dominated by the
 703 temperature and humidity total components, respectively.

704 The decomposition of the temperature and humidity total uncertainties in the temperature channels
 705 (fig. 7) and in the humidity channels (fig. 8), respectively, shows that, again, the model components
 706 are largely dominant. Note that for the highest peaking temperature channel (channel 12) the
 707 second largest uncertainty is the GRUAN pressure component. Also, the lowest peaking humidity
 708 channels (channels 18-19) are significantly affected by the surface uncertainty, although this may
 709 vary with the location and the water vapour burden making those channels peak more or less high in
 710 the atmosphere and therefore more or less sensitive to surface.

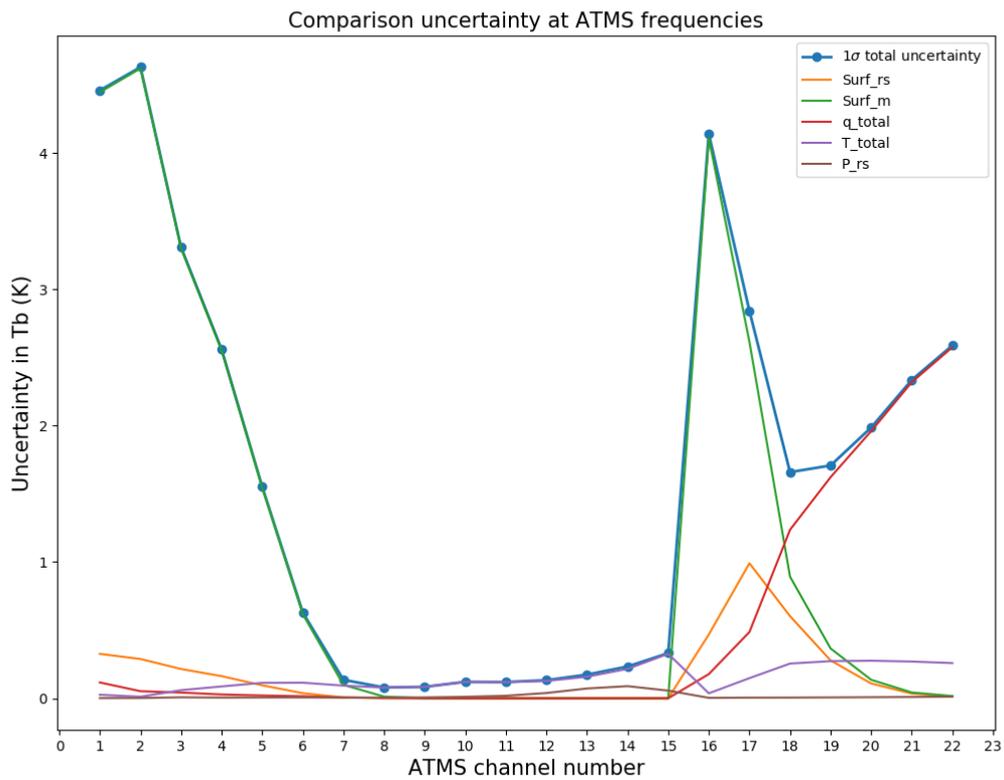
711 The total uncertainty ranges from 0.08 to 0.13K for the temperature channels in figure 7, and from
 712 1.6 to 2.5K for the humidity channels in figure 8. Compared to the mean difference $\Delta T_{b_{MetOffice}}$
 713 documented in Table 1, the night-time sampling satisfies the consistency requirement of Eq. (1) with
 714 $k=1$, noting that the σ term in Eq. (1) that should represent the uncertainty associated with the tri-
 715 linear horizontal interpolation, is currently unknown, although assumed small, and therefore

716 ignored. Future work will be dedicated to the estimation of this σ term using high resolution regional
717 model.

718 These preliminary results are in line with the uncertainty range provided by Loew et al. (2017). This
719 should however be confirmed with the careful evaluation of multiple GRUAN sites over longer time
720 periods, beyond the scope of this paper but planned to be addressed in the near future.

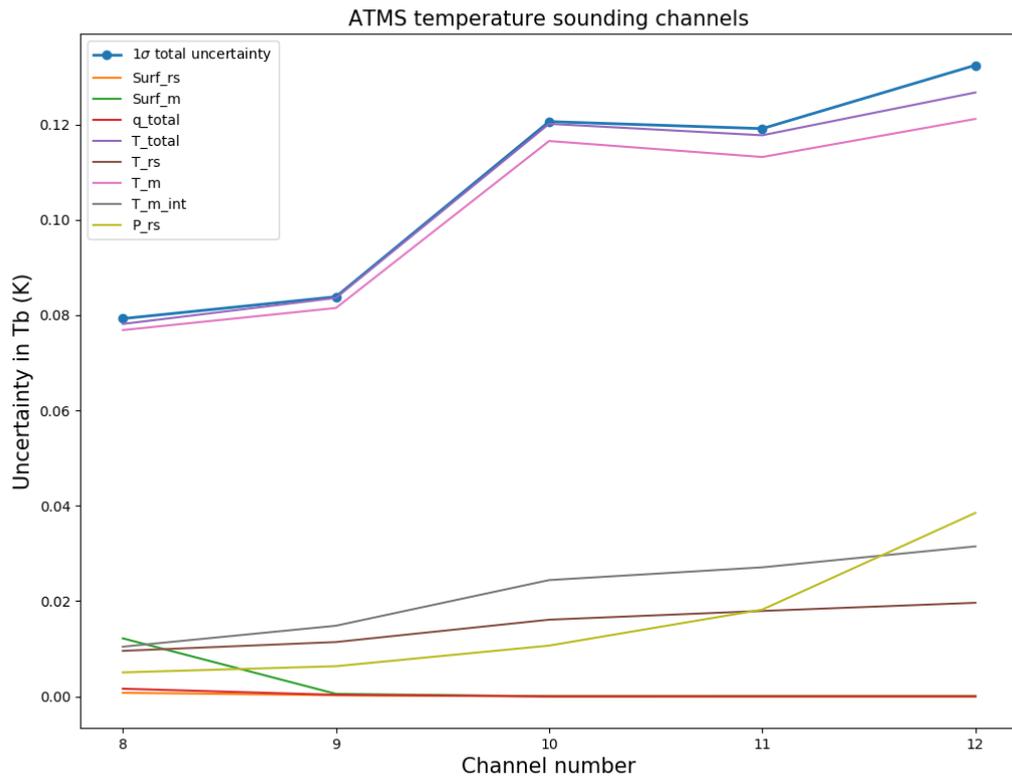
721

722



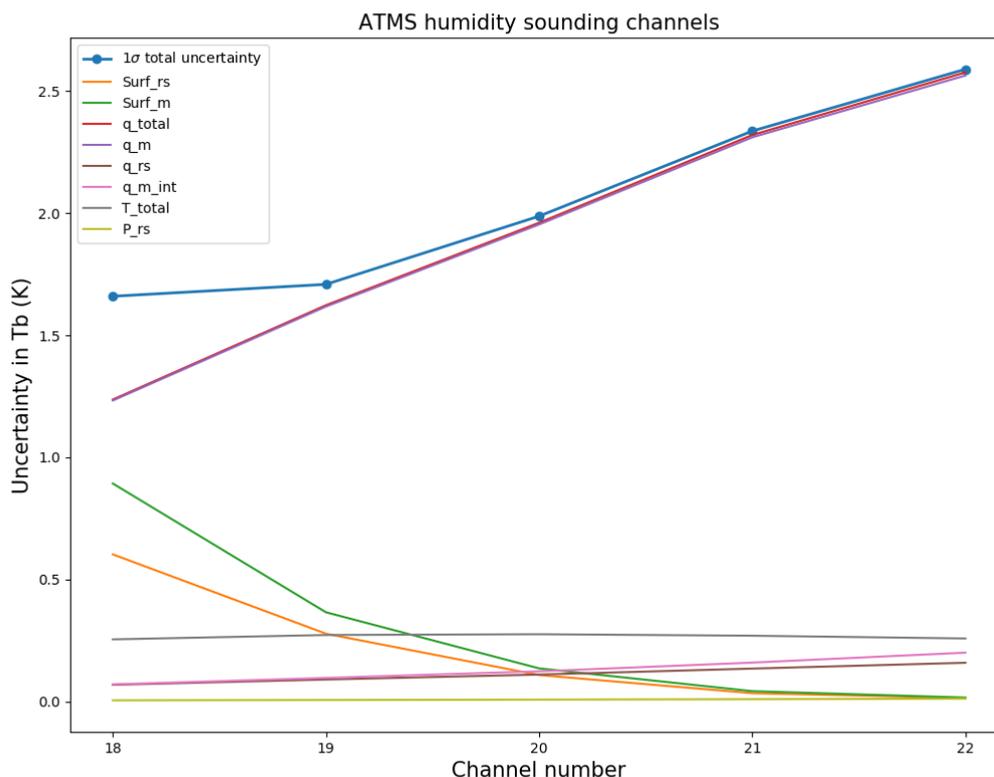
723

724 Figure 6: 1σ standard deviation of the total uncertainty distribution expressed as the square root of
725 the diagonal of the mean comparison covariance $\overline{\mathbf{S}}_{\delta y}$ (blue dots), and the square root of the
726 diagonal of the components forming $\overline{\mathbf{S}}_{\delta y}$, namely, the GRUAN surface uncertainty (Surf_rs, orange),
727 the model surface uncertainty (Surf_m, green), the humidity total uncertainty (q_total, red), the
728 temperature total uncertainty (T_total, purple), and the GRUAN pressure uncertainty (P_rs, brown).



729

730 Figure 7: Same as figure 6 but only for ATMS temperature upper tropospheric-lower stratospheric
 731 channels 8-12, with in addition the square root of the diagonal of the components forming $S_{T_{total}}$,
 732 namely, the GRUAN temperature uncertainty (T_{rs}, olive), the model temperature uncertainty (T_m,
 733 pink), the model vertical interpolation uncertainty (T_m_int, grey).



734

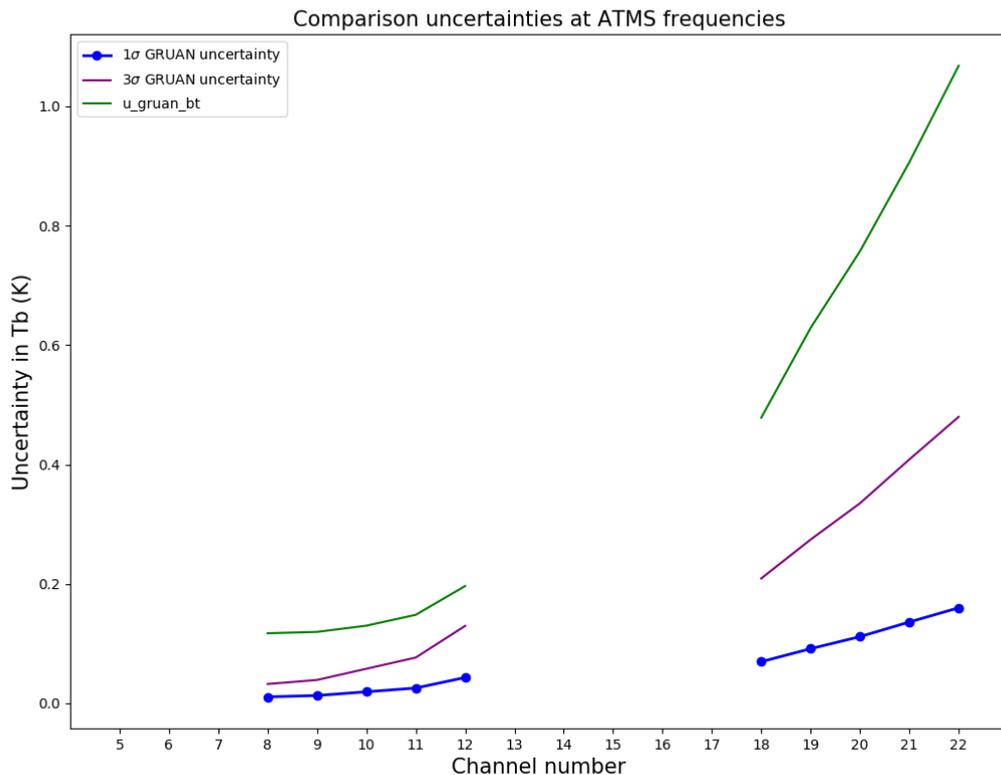
735 Figure 8: Same as figure 6 but only for ATMS humidity tropospheric channels 18-22, with in addition
 736 the square root of the diagonal of the components forming \mathbf{S}_{q_total} , namely, the GRUAN humidity
 737 uncertainty (q_rs, olive), the model humidity uncertainty (q_m, pink), the model vertical
 738 interpolation uncertainty (q_m_int, grey).

739

740 It is interesting to compare the GRUAN processor upper bound uncertainty, calculated assuming a
 741 complete correlation, i.e. u_gruan_bt , with the GRUAN contribution to $\overline{\mathbf{S}}_{\delta y}$. Ignoring the
 742 uncertainties associated with the surface parameters, the GRUAN contribution to $\overline{\mathbf{S}}_{\delta y}$ can be
 743 calculated as the square root of the first three term of Eq. (26). Figure 9 shows that u_gruan_bt is
 744 consistently four times larger than the 3σ standard deviation of the GRUAN contribution to $\overline{\mathbf{S}}_{\delta y}$ at
 745 the frequencies of interest. It may indicate that the assumption of complete correlation in the
 746 uncertainty (i.e. the use of GRUAN total uncertainty as if correlated at all levels), associated with the
 747 calculation of the maximal total uncertainty in Tb results in a large overestimation of the uncertainty
 748 in radiance space. In addition, it should be remembered that the use of diagonal matrices in Eq. (26)
 749 is suboptimal and may not capture the full extent of the uncertainty. The lack of explicit systematic
 750 and random errors associated with the radiosonde profiles and the lack of discretisation between
 751 correlated and uncorrelated uncertainty components in GRUAN products is also suboptimal. This
 752 stresses the need for the GRUAN community to provide proper covariance matrices, better defined
 753 error profiles, and better discretisation of correlated and uncorrelated uncertainties. Finally, it is
 754 possible, although not likely, that a violation of the assumption of ‘small’ uncertainties in Eq.s (21-

755 22) could result in non-linear perturbations potentially causing the GRUAN contribution to $\overline{\mathbf{S}_{\delta\mathbf{y}}}$ to be
 756 underestimated.

757



758

759 Figure 9: 1σ standard deviation of the uncertainty distribution from GRUAN contribution to $\overline{\mathbf{S}_{\delta\mathbf{y}}}$ is
 760 shown in blue (dotted line). It is calculated as the square root of the first three term of Eq. (26), i.e.
 761 $\sqrt{\text{diag}(\mathbf{S}_{q_{rs}} + \mathbf{S}_{T_{rs}} + \mathbf{S}_{P_{rs}})}$. The 3σ standard deviation of the uncertainty distribution is shown in
 762 purple (solid line). $u_{\text{gruan_bt}}$, the GRUAN uncertainty propagated into radiance space by the
 763 GRUAN Processor and averaged over the night-time sample is shown in green (solid line).

764

765 Next, the overall agreement between the Met Office model and GRUAN, in radiance space, is
 766 assessed via a χ^2 test. Here, a reduced χ^2 , hereafter $\tilde{\chi}^2$, is estimated for each profile as follows:

$$\tilde{\chi}^2 = \frac{1}{c} (\delta\mathbf{y}_i - \overline{\delta\mathbf{y}})^T \mathbf{S}_{\delta\mathbf{y}}^{-1} (\delta\mathbf{y}_i - \overline{\delta\mathbf{y}}) \quad (30)$$

767

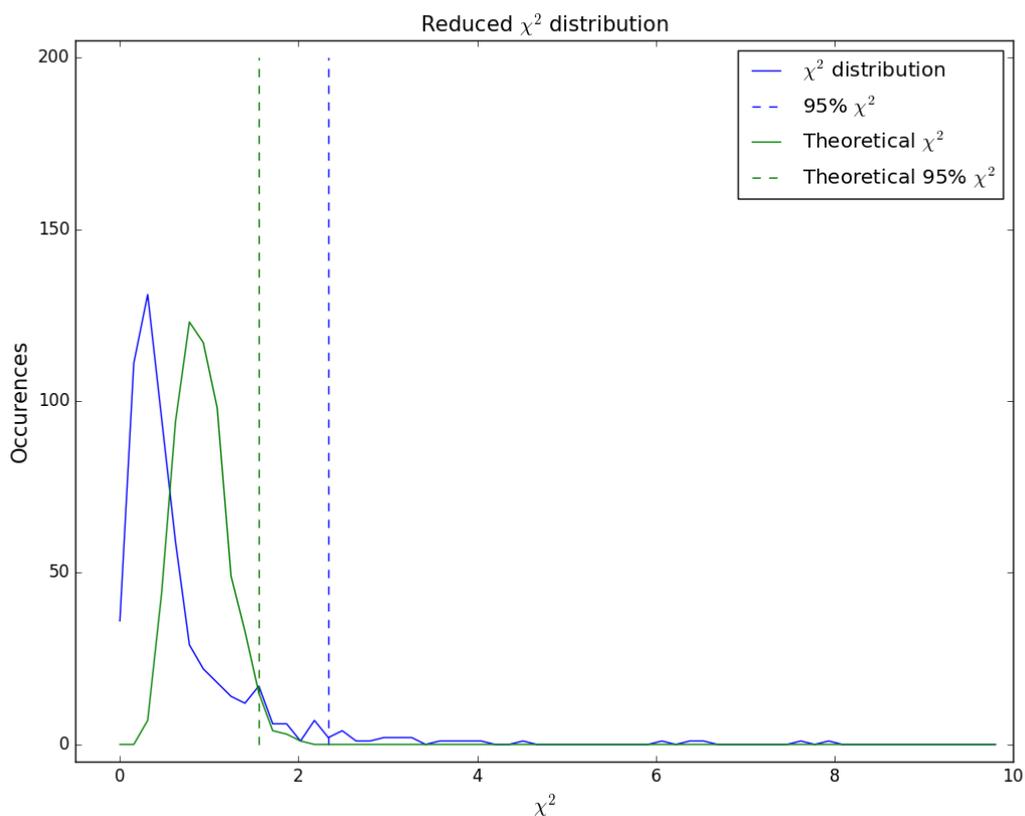
768 where $\delta\mathbf{y}_i$ is the $NWP - GRUAN$ difference in Tb for the i^{th} comparison, $\overline{\delta\mathbf{y}}$ the mean comparison
 769 over the sample. The number of degrees of freedom c , in this context, is the number of channels
 770 regardless any constraints as defined in Rodgers, 2000 (section 12.2).

771

772 Comparing calculated and theoretical \tilde{X}^2 will allow, in theory, the assessment and eventually
773 revision of the uncertainty estimates used for the NWP model and GRUAN. Figure 10 shows the
774 distribution of \tilde{X}^2 calculated for the night-time sampling (blue line) and how it compares to the
775 theoretical \tilde{X}^2 estimated from random data of similar sampling size (green line). Dashed lines show
776 the 95-percentile of each distribution. \tilde{X}^2 values beyond the theoretical 95-percentile line reflect the
777 comparisons where the model and GRUAN are significantly different. For this example, the 95-
778 percentile of the calculated \tilde{X}^2 (blue dashed line) is 5% larger than the theoretical one (green dashed
779 line): i.e. about 10% of the calculated \tilde{X}^2 are greater than the theoretical 95-percentile threshold.
780 This relatively good match between calculated and theoretical \tilde{X}^2 rules out the hypothesis of the
781 violation of small uncertainties in Eq.s (21-22). However, it might be that one (or more) component
782 of $S_{\delta y}$ have been underestimated and could be revised until both 95-percentiles match. It is also
783 possible that unforeseen sources of uncertainty have been unaccounted for in Eq. (25). In both
784 cases, the increased total uncertainty will reduce the number of comparisons failing the test and
785 reduce the difference between the calculated and theoretical 95-percentile threshold.

786 A refined assessment using a larger sample spanning several years and several GRUAN sites will be
787 addressed as part of future work, but is out of scope of this study.

788



789

790 Figure 10: Reduced X^2 distribution from the NWP – GRUAN night-time sampling (blue) and

791 theoretical reduced X^2 estimated from a random sampling of equal size and equal degrees of
792 freedom (blue). Dashed lines show the 95-percentile of each distribution.

793

794 6. Conclusion

795 Numerical weather prediction models have demonstrated ability to act as suitable reference
796 comparators for the calibration and validation of satellite instruments. Model analysis and short-
797 range forecast uncertainties are incrementally reduced by progressive improvements in data
798 assimilation techniques and the ingestion of a large and growing number of observations from
799 multiple sources. From the state-of-the-art of NWP output fields, biases as small as a tenth of a
800 Kelvin can be highlighted in some satellite datasets. In addition, NWP models provide global fields,
801 which allow for the evaluation of satellite data across the full dynamic range of the instrument. Yet
802 model uncertainty estimates do not meet international metrological traceability standards as
803 provided by other reference datasets, such as the GRUAN radiosondes.

804

805 In order to address the missing links in the traceability chain of model uncertainty, a collocation and
806 radiance simulation tool (the GRUAN Processor) has been developed in the framework of the GAIA-
807 CLIM project. This allows us to quantify differences between GRUAN radiosonde profiles of well-
808 defined uncertainties and NWP fields, in both observation and radiance space.

809 Based on the radiative transfer core capability of the radiance simulator developed and maintained
810 by NWP SAF, the Processor collocates model fields to GRUAN radiosonde profiles in space and time,
811 then simulates top-of-atmosphere brightness temperatures for both datasets at frequencies used by
812 satellite instruments, and propagates GRUAN uncertainties in radiance space. The details of the
813 GRUAN Processor have been described in this paper and a mathematical methodology aimed at
814 assessing *NWP – GRUAN* comparisons in radiance space has been expounded.

815

816 For this study, a small sampling of 573 daytime and 587 night-time GRUAN radiosonde profiles from
817 Lindenberg, Germany, in 2016, and matching NWP fields from the Met Office and ECMWF global
818 models have been processed and analysed to demonstrate the GRUAN Processor capability.

819 In the geophysical space of the radiosonde observations, the *NWP – GRUAN* comparison has
820 highlighted 0.5K cold biases located in the stratosphere of the ECMWF model and in the lower
821 troposphere of the Met Office model. A wet bias ranging from 50 to 75% of the local specific
822 humidity is visible in both models at pressure between 200 and 100hPa.

823 In radiance space, the Met Office and ECMWF T_b are found to be within $\pm 0.09K$ and $\pm 0.39K$,
824 respectively, to GRUAN night-time profiles (when GRUAN biases are minimal), at frequencies
825 predominantly sensitive to temperature (54-57GHz) in the vertical domain where GRUAN
826 radiosonde observations are available. Similarly, the Met Office and ECMWF T_b are found to be
827 within $\pm 0.46K$ and $\pm 0.37K$, respectively, to GRUAN night-time profiles at frequencies predominantly
828 sensitive to humidity (around 183GHz).

829

830 The propagation of GRUAN uncertainties in radiance space is performed in the GRUAN Processor via
831 perturbation of the temperature, humidity and pressure profiles by plus and minus their total
832 uncertainty as provided in the RS92-GDP data files. This process assumes a complete correlation of
833 the uncertainties at all levels. This is a pessimistic assumption and the resulting uncertainty obtained
834 in radiance space is therefore representative of a maximum uncertainty of the GRUAN component
835 (the model uncertainty is not accounted for). The true GRUAN uncertainty in radiance space is
836 smaller than that calculated as only a fraction of GRUAN total uncertainty (in observation space) is
837 really correlated over the entire profile.

838 Independently from that maximum GRUAN uncertainty estimate, a rigorous estimation of the
839 uncertainties in radiance space associated with the *NWP – GRUAN* difference is proposed in this
840 study as a post-processing application based on the GRUAN Processor outputs. The covariance of
841 this difference, $\mathcal{S}_{\delta y}$, is calculated as the sum of the GRUAN, model, and interpolation uncertainties
842 propagated in radiance space.

843 Tested with the Met Office background error covariance, the NWP component of $\mathcal{S}_{\delta y}$ is found to be
844 the dominant source of uncertainty. The total uncertainty of the difference ranges from 0.08 to
845 0.13K at frequencies sensitive to temperature and from 1.6 to 2.5K at frequencies sensitive to
846 humidity, satisfying, on average, the consistency check (Eq. 1) for night-time profiles.

847 The GRUAN component of $\mathcal{S}_{\delta y}$ is found to be four times smaller (at 3σ) than the maximum GRUAN
848 uncertainty estimated in the Processor, demonstrating the large overestimation of the complete
849 correlation assumption. However, it is worth stressing that in absence of covariance information,
850 error (random and systematic) characterisation, and discretisation between correlated and
851 uncorrelated uncertainty components in GRUAN data files, the estimation of $\mathcal{S}_{\delta y}$ remains
852 suboptimal.

853 The X^2 distribution calculated for the comparisons between model-based (Met Office) and GRUAN-
854 based simulated Tb revealed that the number of significantly different comparisons is close although
855 slightly larger than that of the corresponding theoretical X^2 distribution. Implications are that either
856 one or several components of $\mathcal{S}_{\delta y}$ are underestimated, or that a source of uncertainty has been
857 overlooked.

858

859 The next step will be to process and analyse collocated profiles spanning several years and multiple
860 GRUAN sites. This will provide a better, although incomplete, geographical distribution of model
861 biases as well as their evolution in time. Away from the surface, NWP model biases are to first order
862 a function of latitude and height, and can usefully be studied for polar, mid-latitude and tropical
863 bands. For northern latitude bands, the NWP uncertainties can be studied by comparison with
864 GRUAN observations, but for the tropics and southern latitudes, where there are few or no GRUAN
865 data, these could be supplemented with other high quality radiosonde reports. The aim will be to
866 provide a refined set of model uncertainty for selected frequencies spanning both microwave and
867 infrared domains. Ultimately, the contribution from this work will help draw the full model
868 uncertainty budget (composed of uncertainties in radiance space, radiative transfer modelling, scale

869 mismatch, and cloud residual) for more robust assessment of satellite observations. Finally, the
 870 larger sampling will also ensure a more robust X^2 analysis and, if deemed necessary, help revise the
 871 model covariance matrices used in operation at the Met Office and ECMWF.

872 The quantitative estimate of errors and uncertainties in NWP models, both temperature, humidity,
 873 and radiance space, could aid in the interpretation of observation minus short-range forecast
 874 statistics for satellite instruments, for example by helping to identify whether biases in observation-
 875 minus-model background values could be due to systematic errors in the NWP model short-range
 876 forecasts. In future work, it is planned to use the GRUAN processor output to evaluate biases in
 877 observation-minus-model background statistics of satellite data.

878

879 GRUAN Processor-based studies also have the potential to refine and improve bias correction
 880 schemes used in NWP centres by helping identify regions where NWP model biases are small as
 881 suggested by Eyre (2016). Similarly, the processing and inter comparison of multiple radiosonde
 882 types can help determine which sets of observations could be use as anchors.

883

884 Finally, the GRUAN processor will also evolve with the evolution of RTTOV. For example, a parallel
 885 version of the Processor is currently being tested with the fast radiative transfer model RTTOV
 886 Ground-based (RTTOV-gb). RTTOV-gb is a modified version of RTTOV that allow for simulations of
 887 ground-based upward-looking microwave sensors (De Angelis et al., 2016). Model and GRUAN
 888 simulated Tb and propagated uncertainties are expected to help estimate the uncertainties in the
 889 microwave radiometer observations for which RTTOV-gb has been developed. It is also planned to
 890 upgrade the Processor in order to support RTTOV 12 (Hocking et al., 2017). This upgrade will allow
 891 the better handling of surface emissivity and give the option to output principal components (PC)
 892 used for the new generation of hyperspectral infrared sounders. Note that other fast radiative
 893 transfer models, such as the Community Radiative Transfer Model (CRTM), could potentially be
 894 tested with the GRUAN Processor, although there is no immediate plan to do so.

895

896 Appendix A: Forecast error covariance matrix estimation

897 If the forecast error covariance matrix from the NWP forecast model used as input to the Processor
 898 is not available, it can be determined from an ensemble of K NWP profiles, with $K > N$ where N is the
 899 number of vertical levels, such that:

$$B_c^m = \frac{X' X'^T}{K - 1} \quad (A1)$$

900 where $K - 1$ gives the best estimate of the covariance of the population from which the sample K is
 901 drawn, and with X' such as:

$$X' = (x_c^{m_1} - \overline{x_c^m}, \dots, x_c^{m_j} - \overline{x_c^m}, \dots, x_c^{m_K} - \overline{x_c^m}) \quad (A2)$$

902 where $\mathbf{x}_c^{m_j}$ is the j^{th} model profile of the K ensemble, and $\overline{\mathbf{x}_c^m}$ is the mean of the K profiles, both on
 903 the coarse model vertical grid.

904

905 Appendix B: Interpolation matrix pseudo inverse

906 The interpolation matrix \mathbf{W} is not square and therefore its inverse cannot be calculated. Instead, a
 907 pseudo inverse, \mathbf{W}^* , can be to express using, for example, the weighted least square estimate of \mathbf{x}_c^t
 908 (Rodgers, 2000). For that, we need to minimize:

$$\mathbf{r} = \frac{1}{2} (\mathbf{x}_f^t - \mathbf{W}\mathbf{x}_c^t)^T \mathbf{B}_f^{m-1} (\mathbf{x}_f^t - \mathbf{W}\mathbf{x}_c^t) \quad (\text{B1})$$

909 where, for the weight, we use the forecast error covariance matrix expressed on the fine grid, \mathbf{B}_f^m ,
 910 since we interpolate the model profiles on that grid.

911 By taking the derivative with respect to \mathbf{x}_c^t and setting it to zero, we find:

$$\mathbf{x}_c^t = (\mathbf{W}^T \mathbf{B}_f^{m-1} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{B}_f^{m-1} \mathbf{x}_f^t \quad (\text{B2})$$

912 where.

$$\mathbf{W}^* = (\mathbf{W}^T \mathbf{B}_f^{m-1} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{B}_f^{m-1} \quad (\text{B3})$$

913 In order to find an expression for \mathbf{B}_f^m , we refer to \mathbf{B}_c^m , the forecast covariance matrix on the coarse
 914 model grid, to calculate the forecast error correlation matrix \mathbf{C}_c^m , on the coarse model grid. The
 915 correlation matrix is then reconditioned on the fine Processor grid, and referred to as \mathbf{C}_f^{rec} , as
 916 explained below.

917 Defining $\mathbf{\Sigma}$, a diagonal matrix representing the square root of \mathbf{B}_c^m variance, such as:

$$\mathbf{\Sigma} = \sqrt{\text{diag}(\mathbf{B}_c^m)} \quad (\text{B4})$$

918 \mathbf{C}_m can be expressed as:

$$\mathbf{C}_m = \mathbf{\Sigma}^{-1} \mathbf{B}_c^m \mathbf{\Sigma}^{-1} \quad (\text{B5})$$

919 We can then define \mathbf{C}_f^m as:

$$\mathbf{C}_f^m = \mathbf{W} \mathbf{C}_c^m \mathbf{W}^T \quad (\text{B6})$$

920 However, Eq. (B6) does not guarantee that \mathbf{C}_f^m diagonal elements are equal to one. This constraint
 921 needs to be imposed such as:

$$\mathbf{C}_f^{rec} = \mathbf{W} \mathbf{C}_c^m \mathbf{W}^T - \text{diag}(\mathbf{W} \mathbf{C}_c^m \mathbf{W}^T) + \mathbf{I} \quad (\text{B7})$$

922 Given σ_m , a vector composed of the square root of the variance of $\boldsymbol{\varepsilon}_m$ variance, \mathbf{B}_f^m is expressed as
 923 follows:

$$\mathbf{B}_f^m = \text{diag}(\mathbf{W}\sigma_m)\mathbf{C}_f^{\text{rec}}\text{diag}(\mathbf{W}\sigma_m) \quad (\text{B8})$$

924

925 Data availability

926 For further information on the GRUAN Processor source code and/or outputs availability, please
927 contact the lead author (fabien.carminati@metoffice.gov.uk).

928

929 Author contribution

930 FC developed the GRUAN Processor code, analysed the data, and prepared the manuscript with
931 contributions from all co-authors. SM developed the mathematical framework presented in section
932 5. BI and HL provided ECMWF datasets and helped with code design and data analyses. WB and SN
933 helped with code design and data analyses. JH develops and manages RTTOV. AS developed the
934 Radiance Simulator.

935

936 Competing interests

937 The authors declare that they have no conflict of interest.

938

939 Acknowledgment

940 This work and its contributors (Fabien Carminati, Stefano Migliorini, Bruce Ingleby, Bill Bell, Heather
941 Lawrence, and Stu Newman) were supported by the European Union's Horizon 2020 research and
942 innovation programme under the GAIA-CLIM grant agreement No. 640276. We are grateful to John
943 Eyre, Emma Turner, and the referees for their comments and suggestions that helped us improve
944 the quality of our study.

945

946 References

947 Agusti-Panareda, A., Vasiljevic, D., Beljaars, A., Bock, O., Guichard, F., Nuret, M., Garcia Mendez, A.,
948 Andersson, E., Bechtold, P., Fink, A., Hersbach, H., Lafore, J.-P., Ngamini, J.-B., Parker, D.J.,
949 Redelsperger, J.-L., Tompkins, A.M.: Radiosonde humidity bias correction over the West African
950 region for the special AMMA reanalysis at ECMWF, Q.J.R. Meteorol. Soc., 135 : 595-617. doi:
951 10.1002/qj.396, 2009.

952 Auligné, T. , McNally, A. P. and Dee, D. P.: Adaptive bias correction for satellite data in a numerical
953 weather prediction system, Q.J.R. Meteorol. Soc., 133: 631-642, doi:10.1002/qj.56, 2007.

954 Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature,
955 525, 47, doi:10.1038/nature14956 , 2015.

956 Bell, W., Candy, B., Atkinson, N., Hilton, F., Baker, N., Bormann, N., Kelly, G., Kazumori, M., Campbell,
957 W. F., and Swadley, S. D.: The assimilation of SSMIS radiances in numerical weather prediction
958 models, *IEEE Transactions on Geoscience and Remote Sensing*, 46, 884–900, doi:
959 10.1109/TGRS.2008.917335, 2008.

960 Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of
961 essential climate variables in support of climate research, applications, and policy, *Bulletin of the*
962 *American Meteorological Society*, 95, 1431–1443, <https://doi.org/10.1175/BAMS-D-13-00047.1>,
963 2014.

964 Bonavita, M., Hólm, E.V., Isaksen, L., Fisher, M.: The evolution of the ECMWF hybrid data
965 assimilation system, *Q. J. R. Meteorol. Soc.*, 142, 287–303, doi:10.1002/qj.2652, 2016.

966 Bormann, N., Fouilloux, A., and Bell, W.: Evaluation and assimilation of ATMS data in the ECMWF
967 system, *Journal of Geophysical Research: Atmospheres*, 118, 12.970–12.980,
968 doi:10.1002/2013JD020325, 2013.

969 Calbet, X., Peinado-Galan, N., Rípodas, P., Trent, T., Dirksen, R., and Sommer, M.: Consistency
970 between GRUAN sondes, LBLRTM and IASI, *Atmos. Meas. Tech.*, 10, 2323–2335,
971 <https://doi.org/10.5194/amt-10-2323-2017>, 2017.

972 Courtier, P. , Thépaut, J. and Hollingsworth, A.: A strategy for operational implementation of 4D-Var,
973 using an incremental approach. *Q.J.R. Meteorol. Soc.*, 120: 1367–1387,
974 doi:10.1002/qj.49712051912, 1994.

975 De Angelis, F., Cimini, D., Hocking, J., Martinet, P., and Kneifel, S.: RTTOV-gb – adapting the fast
976 radiative transfer model RTTOV for the assimilation of ground-based microwave radiometer
977 observations, *Geosci. Model Dev.*, 9, 2721–2739, <https://doi.org/10.5194/gmd-9-2721-2016>,
978 2016.

979 Dee, D.: Variational bias correction of radiance data in the ECMWF system, in *Proc. of the ECMWF*
980 *Workshop on Assimilation of High Spectral Resolution Sounders in NWP*, Vol. 28, pp. 97–112,
981 <https://www.ecmwf.int/search/elibrary>, 2004.

982 Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-
983 air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, *Atmos. Meas. Tech.*,
984 7, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>, 2014.

985 [Eyre, J. R., 2016: Observation bias correction schemes in data assimilation systems: a theoretical](#)
986 [study of some of their properties, *Q. J. R. Meteorol. Soc.*, 142\(699\), pp.2284–2291,](#)
987 [DOI:10.1002/qj.2819.](#)

988 Eyre, J.: A fast radiative transfer model for satellite sounding systems, *ECMWF Tech. Memo 176*,
989 <https://www.ecmwf.int/search/elibrary>, 1991.

990 Green, P., Gardiner, T., Medland, D., Cimini, D.: Guide to Uncertainty in Measurement & its
991 Nomenclature, GAIA-CLIM Deliverable D2.6 <http://www.gaia-clim.eu/>, 2018.

992 Gobiet, A., Foelsche, U., Steiner, A. K., Borsche, M., Kirchengast, G. and Wickert J.: Climatological
993 validation of stratospheric temperatures in ECMWF operational analyses with CHAMP radio
994 occultation data, *Geophys. Res. Lett.*, 32, L12806, doi: 10.1029/2005GL022617, 2005.

995 Han, W. and Bormann, N.: Constrained adaptive bias correction for satellite radiance assimilation in
996 the ECMWF 4D-Var system. ECMWF Technical Memorandum 783,
997 <https://www.ecmwf.int/search/elibrary>, 2016.

998 Haseler, J.: Early-delivery suite. ECMWF Technical Memorandum No. 454,
999 <https://www.ecmwf.int/search/elibrary>, 2004.

1000 Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P., and Vidot, J.:
1001 RTTOV v12 Users Guide, Tech. Rep. NWPSAF-MO-UD-037, EUMETSAT Satellite Application Facility
1002 on Numerical Weather Prediction (NWPSAF),
1003 <https://www.nwpsaf.eu/site/software/rttov/documentation/>, 2017.

1004 Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P., and Vidot, J.:
1005 RTTOV v11 Users Guide, Tech. Rep. NWPSAF-MO-UD-028, EUMETSAT Satellite Application Facility
1006 on Numerical Weather Prediction (NWPSAF), [https://www.nwpsaf.eu/site/software/rttov/rttov-](https://www.nwpsaf.eu/site/software/rttov/rttov-v11/)
1007 [v11/](https://www.nwpsaf.eu/site/software/rttov/rttov-v11/), 2015.

1008 Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E.,
1009 Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath,
1010 S., van Roozendaal, M. and Wagner, W.: The ESA climate change initiative: Satellite data records
1011 for essential climate variables, *Bulletin of the American Meteorological Society*, 94, 1541–1552,
1012 <https://doi.org/10.1175/BAMS-D-11-00254.1>, 2013.

1013 Hyland, R. W., Wexler, A. Formulation for the thermodynamic properties of the saturated phases of
1014 H₂O from 173.15K to 473.15K. *ASHRAE Trans.*, 89(2A), 500–519, 1983.

1015 Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference
1016 Quality Upper-Air Measurements: guidance for developing GRUAN data products, *Atmos. Meas.*
1017 *Tech.*, 3, 1217-1231, <https://doi.org/10.5194/amt-3-1217-2010>, 2010.

1018 Ingleby, B., Isaksen, L., Kral, T., Haiden, T., Dahoui, M.: Improved use of atmospheric in situ data,
1019 ECMWF Newsletter, number 155, Page 20-25, <https://www.ecmwf.int/node/18208>,
1020 doi:10.21957/cf724bi05s, 2018.

1021 Ingleby, B.: An assessment of different radiosonde types 2015/2016, European Centre for Medium
1022 Range Weather Forecasts, <https://www.ecmwf.int/search/elibrary>, 2017.

1023 Ingleby, B. and Edwards, D.: Changes to radiosonde reports and their processing for numerical
1024 weather prediction, *Atmospheric Science Letters*, 16: 44-49, doi:10.1002/asl2.518, 2015.

1025 Ingleby, N., Lorenc, A., Ngan, K., Rawlins, F., and Jackson, D.: Improved variational analyses using a
1026 nonlinear humidity control variable, *Quarterly Journal of the Royal Meteorological Society*, 139:
1027 1875-1887, doi:10.1002/qj.2073, 2013.

- 1028 Joo, S., Eyre, J., and Marriott, R.: The impact of Metop and other satellite data within the Met Office
 1029 global NWP system using an adjoint-based sensitivity method, *Monthly Weather Review*, 141,
 1030 3331–3342, <https://doi.org/10.1175/MWR-D-12-00232.1>, 2013.
- 1031 Kazumori, M. and English, S. J.: Use of the ocean surface wind direction signal in microwave radiance
 1032 assimilation, *Quarterly Journal of the Royal Meteorological Society*, 141: 1354-1375,
 1033 doi:10.1002/qj.2445, 2015.
- 1034 Lesht, B., Miloshevich, L., and Liljegren, J.: Comparison of ECMWF model analyses with the observed
 1035 upperair temperature and relative humidity climatology at the ARM NSA, SGP, and TWP Climate
 1036 Research Facility Sites, in Proc. 14th Atmospheric Radiation Measurement (ARM) Science Team
 1037 Meeting, Albuquerque, New Mexico, March 22-26, 2004.
- 1038 Leutbecher, M. , Lock, S. , Ollinaho, P. , Lang, S. T., Balsamo, G. , Bechtold, P. , Bonavita, M. ,
 1039 Christensen, H. M., Diamantakis, M. , Dutra, E. , English, S. , Fisher, M. , Forbes, R. M., Goddard, J.
 1040 , Haiden, T. , Hogan, R. J., Juricke, S. , Lawrence, H. , MacLeod, D. , Magnusson, L. , Malardel, S. ,
 1041 Massart, S. , Sandu, I. , Smolarkiewicz, P. K., Subramanian, A. , Vitart, F. , Wedi, N. and
 1042 Weisheimer, A.: Stochastic representations of model uncertainties at ECMWF: state of the art
 1043 and future vision. *Q.J.R. Meteorol. Soc.*, 143: 2315-2339. doi:10.1002/qj.3094, 2017.
- 1044 Loew, A., Bell, W., Brocca, L., Bulgin, C.E., Burdanowitz, J., Calbet, X., Donner, R.V., Ghent, D., Gruber,
 1045 A., Kaminski, T. and Kinzel, J., Klepp, C., Lambert, J. C., Schaepman-Strub, G., Schröder, M.,
 1046 Verhoelst, T.: Validation practices for satellite based earth observation data across communities,
 1047 *Reviews of Geophysics*, 55, 779–817, doi:10.1002/2017RG000562, 2017.
- 1048 Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L., Barker, D. M., Bray, J. R., Clayton,
 1049 A. M., Dalby, T. , Li, D. , Payne, T. J. and Saunders, F. W.: The Met. Office global three-dimensional
 1050 variational data assimilation scheme, *Quarterly Journal of the Royal Meteorological Society*, 126:
 1051 2991-3012, doi:10.1002/qj.49712657002, 2000.
- 1052 Lu, Q. and Bell, W.: Characterizing channel center frequencies in AMSU-A and MSU microwave
 1053 sounding instruments, *Journal of Atmospheric and Oceanic Technology*, 31, 1713–1732,
 1054 <https://doi.org/10.1175/JTECH-D-13-00136.1>, 2014.
- 1055 Lupu, C., Geer, A. J.: Operational Implementation of RTTOV-11 in the IFS. ECMWF Tech. Memo. 748,
 1056 <https://www.ecmwf.int/search/elibrary>, 2015
- 1057 Massonnet, F., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J.: Using climate models to estimate
 1058 the quality of global observational data sets, *Science*, doi:10.1126/science.aaf6369, 2016.
- 1059 Paufler, P.: Landolt-Börnstein: Numerical data and functional relationships in science and
 1060 technology, Ed. by O. Madelung Springer-Verlag Berlin, 23: 1360-1360,
 1061 doi:10.1002/crat.2170231029, 1988.
- 1062 Rawlins, F., Ballard, S., Bovis, K., Clayton, A., Li, D., Inverarity, G., Lorenc, A., and Payne, T.: The Met
 1063 Office global four-dimensional variational data assimilation scheme, *Quarterly Journal of the*
 1064 *Royal Meteorological Society*, 133: 347-362, doi:10.1002/qj.32, 2007.
- 1065 Rodgers CD. Inverse methods for atmospheric sounding: theory and practice. World scientific; 2000.

1066 Saunders, R., Matricardi, M., and Brunel, P.: An improved fast radiative transfer model for
1067 assimilation of satellite radiance observations, *Quarterly Journal of the Royal Meteorological*
1068 *Society*, 125: 1407-1425, doi:10.1002/qj.1999.49712555615, 1999.

1069 Saunders, R., Rayer, P., Blackmore, T., Matricardi, M., Bauer, P., and Salmond, D.: A new fast
1070 radiative transfer model-RTTOV-9, in *Proc. Joint 2007 EUMETSAT Meteorological Satellite*
1071 *Conference and the 15th Satellite Meteorology and Oceanography Conference of the American*
1072 *Meteorological Society*, Amsterdam, The Netherlands, 2007.

1073 Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Rocquet, P.,
1074 Matricardi, M., Geer, A., Bormann, N., and Lupu, C.: An update on the RTTOV fast radiative
1075 transfer model (currently at version 12), *Geosci. Model Dev. Discuss.*,
1076 <https://doi.org/10.5194/gmd-2018-64>, in review, 2018.

1077 Shepherd, T.G., Polichtchouk, I., Hogan, R.J., Simmons, A.J.: Report on Stratosphere Task Force,
1078 ECMWF Technical Memorandum, 824, [https://www.ecmwf.int/en/elibrary/18259-report-](https://www.ecmwf.int/en/elibrary/18259-report-stratosphere-task-force)
1079 [stratosphere-task-force](https://www.ecmwf.int/en/elibrary/18259-report-stratosphere-task-force), 2018.

1080
1081 Simmons, A., Untch, A., Jakob, C., Källberg, P., and Uden, P.: Stratospheric water vapour and
1082 tropical tropopause temperatures in ECMWF analyses and multi-year simulations, *Quarterly*
1083 *Journal of the Royal Meteorological Society*, 125: 353-386, doi:10.1002/qj.49712555318, 1999.

1084 Smith, A.: Radiance simulator v2.0 user guide, EUMETSAT Satellite Application Facility on Numerical
1085 Weather Prediction (NWPSAF), Tech. Rep. NWPSAF-MO-UD-040,
1086 <https://www.nwpsaf.eu/site/software/radiance-simulator/documentation/>, 2014.

1087 Sommer, M., Dirksen, R., and Rohden, C.: Brief Description of the RS92 GRUAN Data Product (RS92-
1088 GDP), Technical Document GRUAN-TD-4, <https://www.gruan.org/documentation/gruan/td/>,
1089 2011.

1090 Thorne, P. W., Madonna, F., Schulz, J., Oakley, T., Ingleby, B., Rosoldi, M., Tramutola, E., Arola, A.,
1091 Buschmann, M., Mikalsen, A. C., Davy, R., Voces, C., Kreher, K., De Maziere, M., and Pappalardo,
1092 G.: Making better sense of the mosaic of environmental measurement networks: a system-of-
1093 systems approach and quantitative assessment, *Geosci. Instrum. Method. Data Syst.*, 6, 453-472,
1094 <https://doi.org/10.5194/gi-6-453-2017>, 2017.

1095 Zeng, Y., Su, Z., Calvet, J. C., Manninen, T., Swinnen, E., Schulz, J., Roebeling, R., Poli, P., Tan, D.,
1096 Riihelä, A., Tanis, C. M., Arslan, A. N., Obregon, A., Kaiser-Weiss, A., John, V. O., Timmermans, W.,
1097 Timmermans, J., Kaspar, F., Gregow, H., Barbu, A. L., Fairbairn, D., Gelati, E., Meurey, C.: Analysis
1098 of current validation practices in Europe for space-based climate data records of essential climate
1099 variables, *International Journal of Applied Earth Observation and Geoinformation*, Volume 42,
1100 Pages 150-161, ISSN 0303-2434, <https://doi.org/10.1016/j.jag.2015.06.006>, 2015.

1101 Zou, X., Wang, X., Weng, F., and Li, G.: Assessments of Chinese FengYun Microwave Temperature
1102 Sounder (MWTS) measure- ments for weather and climate applications, *Journal of Atmospheric*
1103 *and Oceanic Technology*, 28, 1206–1227, <https://doi.org/10.1175/JTECH-D-11-00023.1>, 2011.

1104

1105 Tables

1106

1107 Table 1: Mean difference *NWP – GRUAN* in simulated Tb for ECMWF ($\Delta T_{b_{ECMWF}}$) and Met Office
1108 ($\Delta T_{b_{MetOffice}}$) and 1σ standard deviation for ATMS channels 8-12 and 18-22 at day and night-time.

Channel	Frequency (GHz)	$\Delta T_{b_{ECMWF}} (1\sigma)$ (K)		$\Delta T_{b_{MetOffice}} (1\sigma)$ (K)	
		night	day	night	day
8	54.94	-0.08 (0.09)	-0.16 (0.10)	-0.00 (0.11)	-0.04 (0.12)
9	55.5	-0.15 (0.12)	-0.24 (0.13)	0.04 (0.13)	-0.02 (0.14)
10	57.29	-0.32 (0.18)	-0.45 (0.18)	0.01 (0.16)	-0.07 (0.20)
11	57.29±0.217	-0.39 (0.21)	-0.54 (0.22)	-0.04 (0.20)	-0.16 (0.25)
12	57.29±0.3222±0.048	-0.34 (0.25)	-0.53 (0.27)	-0.09 (0.28)	-0.26 (0.31)
18	183.31±7.0	0.35 (0.91)	0.25 (1.09)	0.02 (0.83)	-0.36 (1.02)
19	183.31±7.0	0.37 (1.13)	0.15 (1.24)	-0.09 (1.03)	-0.48 (1.14)
20	183.31±3.0	0.34 (1.31)	-0.01 (1.36)	-0.18 (1.22)	-0.61 (1.27)
21	183.31±1.8	0.22 (1.48)	-0.29 (1.50)	-0.31 (1.42)	-0.81 (1.45)
22	183.31±1.0	0.04 (1.61)	-0.61 (1.64)	-0.46 (1.57)	-1.01 (1.60)

1109

1110