Evaluation of the boundary layer dynamics of the TM5 model over Europe

E. N. Koffi[1], P. Bergamaschi[1], U. Karstens[2,3], M. Krol[4,5,6], A. Segers[7], M. Schmidt[8,11], I. Levin[8], A. T. Vermeulen[9,3], R. E. Fisher[10], V. Kazan[11], H. Klein Baltink[12], D. Lowry[10], G. Manca[1], H. A. J. Meijer[13], J. Moncrieff[14], S. Pal[15], M. Ramonet[11], H.A. Scheeren[1,13], A. G. Williams[16]

1 European Commission Joint Research Centre, Institute for Environment and Sustainability, Ispra (Va), Italy
2 Max-Planck-Institute for Biogeochemistry, Jena, Germany
3 ICOS Carbon Portal, ICOS ERIC at Lund University, Sweden
4 SRON Netherlands Institute for Space Research, Utrecht, Netherlands,
5 Institute for Marine and Atmospheric Research Utrecht, Utrecht, University, Utrecht, Netherlands
6 MAQ, Wageningen University and Research Centre, Wageningen, Netherlands,
7 Netherlands Organisation for Applied Scientific Research (TNO), Utrecht, Netherlands
8 Institut für Umweltphysik, Heidelberg University, Germany
9 Energy research Center Netherlands (ECN), Petten, Netherlands
10 Royal Holloway, University of London (RHUL), Egham, UK
11 Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
12 Royal Netherlands Meteorological Institute (KNMI), Netherlands
13 Centrum voor Isotopen Onderzoek (CIO), Rijksuniversiteit Groningen, Netherlands
14 Atmospheric Chemistry Research Group, University of Bristol, UK
15 Department of Meteorology, Pennsylvania State University, State College, PA, USA
16 Australian Nuclear Science and Technology Organisation (ANSTO) Environment Research Theme, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia

Abstract

We evaluate the capability of the global atmospheric transport model TM5 to simulate the boundary layer dynamics and associated variability of trace gases close to the surface, using radon (^{222}Rn). Focusing on the European scale, we compare the boundary layer height (BLH) in the TM5 model with observations from the NOAA Integrated Global Radiosonde Archive and also with ceilometer and lidar BLH retrievals at two stations. Furthermore, we compare TM5 simulations of ^{222}Rn activity concentrations, using a novel, process-based ^{222}Rn flux map over Europe (Karstens et al., 2015),with harmonized ^{222}Rn measurements at 10 stations.

The TM5 model reproduces relatively well the daytime BLH (within 10-20% for most of the stations), except for coastal sites, for which differences are usually larger due to model representation errors. During night, however, TM5 overestimates the shallow nocturnal BLHs, especially for the very low observed BLHs (< 100 m) during summer.

The ^{222}Rn activity concentration simulations based on the new ^{222}Rn flux map show significant improvements especially regarding the average seasonal variability, compared to simulations using constant ^{222}Rn fluxes. Nevertheless, the (relative) differences between simulated and observed daytime minimum ^{222}Rn activity concentrations are larger for several stations (on the order of 50%) than the (relative) differences between simulated and observed BLH at noon. Although the nocturnal BLH is often higher in the model than observed, simulated ^{222}Rn nighttime maxima are actually larger at several continental stations. This counterintuitive behaviour points to potential deficiencies of TM5 to correctly simulate the vertical gradients within the nocturnal boundary layer, limitations of the ^{222}Rn flux map, or issues related to the definition of the nocturnal BLH.

At several stations the simulated decrease of ^{222}Rn activity concentrations in the morning is faster than observed. In addition, simulated vertical ^{222}Rn activity concentration gradients at Cabauw decrease faster than observations during the morning transition period, and are in general lower than observed gradients during daytime. Although these effects may be partially due to the slow response time of the radon detectors, they clearly point to too fast vertical mixing in the TM5 boundary layer during daytime. Furthermore, the capability of the TM5 model to simulate the diurnal BLH cycle is limited by the current coarse temporal resolution (3hr/6hr) of the TM5 input meteorology.

1. Introduction

The boundary layer, being the lowest portion of the atmosphere, is largely affected by the Earth's surface forcing. This layer is usually separated from the free troposphere (where the surface effects are weak) by a thin and strongly stable layer (capping inversion) that traps turbulence, moisture, and trace gases below. The thickness of the boundary layer is variable in space and time and can range from tens of meters to 4 km, depending on both the synoptic and local meteorological conditions (Stull, 1988). The height of the boundary layer is a critical parameter in atmospheric transport models, since it controls the extent of the vertical mixing of trace gases emitted near the surface. Previous studies that evaluated the ability of atmospheric transport models to reproduce boundary layer dynamics demonstrated the importance of temporal resolution of meteorological data, horizontal and vertical model resolutions, and parameterizations of vertical mixing (e.g., Denning et al., 1999; Dentener et al., 1999; Krol et al., 2005; Locatelli et al., 2015). The realistic simulation of boundary layer height (BLH) is crucial, especially for regional flux inversions which make use of networks of surface and tower-based trace gas concentration measurements to capture the signals of regional sources (and sinks). Regional inversions of greenhouse gases (GHG) (CO_2, CH_4, N_2O, halocarbons) were reported especially for Europe and North America, making use of the increasing number of regional monitoring stations in these areas (e.g., Gerbig et al., 2003; Carouge et al., 2008;;Bergamaschi et al., 2010; Corazza et al., 2011; Manning et al., 2011; Broquet et al., 2013; Bergamaschi et al., 2015; Ganesan et al., 2015) as well as aircraft observations (e.g., Kort et al., 2008; Miller et al. 2013).

In order to evaluate the quality of such flux inversions, a thorough validation of the applied atmospheric transport models is essential. In this study, we present a detailed evaluation of the boundary layer dynamics of the TM5 model (Krol et al., 2005), which is the global transport model used in the TM5-4DVAR inverse modelling system (Meirink et al., 2008), applied in several of the European inversions mentioned above (Corazza et al., 2011; Bergamaschi et al., 2010; 2015). As a first step, we compare the model BLH with the sounding-derived BLH of the NOAA Integrated Global Radiosonde Archive (IGRA) (Seidel et al., 2012) at European scale. Radiosonde data have been considered to give the most accurate BLHs (Collaud Coen et al., 2014). The model BLHs are also compared to those derived from the ceilometer and lidar measurements at two European stations (Cabauw and Trainou). As a second step, we compare TM5 simulations of ^{222}Rn activity concentrations with measurements at 10 European stations. ^{222}Rn is an excellent tracer for boundary layer mixing due to its short lifetime (half-life) of 3.82 days and has been widely used for model validation (e.g., Jacob and Prather, 1990; Jacob et al., 1997; Dentener et al., 1999; Chevillard et al., 2002; Taguchi et al., 2011) and mixing studies (e.g., see reviews in Zahorowski et al., 2004; Chambers et al., 2011; Williams et al., 2011, 2013). However, the use of ^{222}Rn for this purpose has been limited by the simplified assumption of constant ^{222}Rn fluxes over land used in most ^{222}Rn validation studies published so far. It has also been limited by the fact that the observed ^{222}Rn activity concentrations from different stations were not harmonized.
Here, we make use of a novel detailed ^{222}Rn flux map over Europe (Karstens et al., 2015) based on a parameterization of ^{222}Rn production and transport in the soil as well as improved observed ^{222}Rn activity concentrations obtained through a detailed comparison study (Schmithüsen et al., 2016). The development of this ^{222}Rn flux map has been performed within the European project

InGOS ('Integrated non-CO_2 Greenhouse gas Observing System'), including also a comparison of different transport models (including TM5). While this model comparison will be published elsewhere (Karstens et al., 2016, manuscript in preparation), we present here the analysis for the TM5 model aiming at the identification and quantification of potential systematic errors in the simulation of the BLH dynamics, which could directly translate into systematic errors in the derived surface fluxes. Our study also includes the evaluation of a new parameterization of convection in TM5, based on ECMWF (re)analysis, compared to the default convection scheme used so far, based on the parameterization of Tiedtke (1989).

2. Observations

2.1. Boundary layer height

Vertical mixing in the atmospheric boundary layer is mostly turbulent. The BLH is confined by a thin layer where steep vertical gradients of meteorological variables, trace gases, and aerosols occur. Consequently, all the observational devices built for the retrieval of BLH are based on the search of the height at which the strongest gradients occur. These gradients can be based either on the atmospheric potential temperature profile, the wind profile, or the aerosol backscatter profile. For meteorological data sets and atmospheric transport models, the bulk Richardson number, a dimensionless parameter defined as the ratio of turbulence due to buoyancy and the mechanic generation of turbulence by wind shear, has been widely used to determine BLHs (e.g., Vogelezang and Holtslag, 1986; Seibert et al., 2000; Seidel et al., 2012). Thus, the BLH is the vertical level at which the bulk Richardson number (R_{ib}) reaches a critical value R_{ic} characterizing the passage of turbulent fluid flow to laminar one. The general expression of Vogelezang and Holtslag (1986) used to compute R_{ib} is given as follows:

$$R_{ib} = \left(\frac{g}{\theta_{vs}}\right)\frac{(\theta_{vh}-\theta_{vs})(z_h-z_s)}{(u_h-u_s)^2+(v_h-v_s)^2+bu_*^2} \qquad (1)$$

where g is the gravitational acceleration (9.81m s^{-2}), θ_v the virtual potential temperature, z the geopotential height, u the zonal wind speed, and v the meridional wind speed. The indices h and s denote the vertical layer, and the surface, respectively. bu_*^2 depicts the turbulence production due to the surface friction, a term which also prevents an undetermined R_{ib} in case of uniform high wind speeds relevant for neutral boundary layers. b is a coefficient estimated to be 100 (Vogelezang and Holtslag, 1986) and u_* is the surface friction velocity. The geopotential height z is expressed in m. The virtual potential temperature θ_v is in K, and the velocities are in ms^{-1}.
The vertical profile of R_{ib} is linearly interpolated between consecutive vertical layers. The BLH is defined as the height, where R_{ib} reaches the critical value R_{ic}. Commonly, a R_{ic} value of 0.25 has been used (e.g., Vogelezang and Holtslag, 1986; Seibert et al., 2000; Seidel et al., 2012). The boundary layer height is defined with reference to surface elevation, and not to sea level (Seidel et al., 2012).

2.1.1. IGRA data

We use BLHs of the NOAA IGRA database, which covers the 1990-2010 period (Seidel et al., 2012). The IGRA data is based on radiosonde measurements that are usually released at 00 and 12 UTC. The IGRA radiosonde network over Europe is shown in **Figure 1**. The dynamic (wind speed and direction) and thermal (temperature and humidity) profiles from the radiosondes are utilized to compute BLHs using the bulk Richardson number method [Eq.1; Section 2.1]. In these BLH calculations both the surface wind (i.e., u_s and v_s in Eq.1) and the surface friction velocity (u^*) are unknown and set to zero. The critical value of the bulk Richardson number (R_{ic}) is set to 0.25 (instead of 0.3 as used in TM5; see Section 3.2). Further details on the choice of the settings as well as the vertical profiles of the dynamic, thermodynamic, and bulk Richardson number quantities are described in Seidel et al. (2012). These settings for the IGRA database were also adopted in the InGOS protocol for the evaluation of the transport models involved in InGOS inverse modelling analyses (Karstens et al., 2016, manuscript in preparation). The methodological uncertainties in the IGRA BLH data were evaluated based on paired soundings released at the same site (Seidel et al., 2012). Results show that the choice of R_{ic} does not introduce large uncertainty, but other methodological choices (including surface wind speed estimates and vertical interpolation of the bulk Richardson number profile) as well as the vertical resolution of the sounding data are larger sources of uncertainty in the derived BLHs (Seidel et al., 2012). The authors reported relative uncertainties in the IGRA BLHs that can be large (>50%) for shallow BLHs (< 1 km; mainly observed during night or early in the morning), but much smaller (usually <20%) for deep BLHs (> 1 km) during daytime.

2.1.2. Lidar and ceilometer data

The principle of LIDAR (LIght Detection And Ranging; hereafter lidar) is based on a pulsed laser light emitted into the atmosphere which is back-scattered by aerosol particles and molecules. The lidar algorithms derive the BLHs by searching the location of the strongest aerosol gradient in the vertical dimension (e.g., Haeffelin et al., 2011; Pal et al., 2012; Griffiths et al., 2013; Pal et al., 2015). A ceilometer is a 'low-cost lidar' which was initially used for the detection of cloud base heights. However, since the backscatter signal of aerosols is lower than that of clouds, the sensitivity of ceilometers in retrieving the boundary layer height is much less than that of lidar instruments (Pal, 2014). In contrast to IGRA data (i.e., radiosonde based BLH), the ceilometer and lidar allow measurements of the diurnal BLH cycle. However, the algorithms of both lidar and ceilometer have some difficulties to assign the BLH during night and tend to wrongly attribute the height of the residual layer of aerosol (often with larger signal) as height of the real mixed layer (e.g., Angevine et al., 1998; Eresmaa et al., 2006; Haij et al., 2006). Lidar/ceilometer nocturnal BLHs are also higher due to the fact that their overlap height can be above the nocturnal shallow BLH (Pal et al., 2015). Uncertainties in lidar retrieved BLHs were assessed based on a comparison between radiosonde based BLHs and wavelet derived BLH estimates from lidar and found to be about 60 m (Pal et al., 2013).

We use the BLHs retrieved from lidar and ceilometer measurements at Trainou and Cabauw, respectively (see **Figure 1** for their locations). The lidar (ALS-300) measurements at Trainou are described by Pal et al. (2012). The ceilometer at Cabauw is part of the network of the Vaisala LD-40 ceilometer in the Netherlands operated by the Royal Netherlands Meteorological Institute (KNMI; Haij et al., 2006). We analyze the ceilometer measurements at Cabauw for 2010 and the lidar data at Trainou for 2011. For Cabauw we compare the ceilometer based BLH for 2010 with

the BLH data from the closest IGRA station (De Bilt), with results shown in the Supplement (**Figure S1**).

2.2. Observed ^{222}Rn activity concentrations

The observed ^{222}Rn activity concentrations are obtained from 2 different measurement methods:
 (1) The 'two-filter' method developed by the Australian Nuclear Science and Technology Organisation (ANSTO) (Whittlestone and Zahorowski, 1998; Chambers et al., 2011). After drawing the sampled air continuously through a delay volume to let all short-lived ^{220}Rn (thoron) gas in the sampled air decay, it passes through a first filter that removes all ambient ^{222}Rn and ^{220}Rn decay products. Filtered air then enters in a delay chamber in which new ^{222}Rn progeny (^{218}Po and ^{214}Po) are produced. An internal flow loop within the delay chamber passes the air through a second filter, which collects the new ^{222}Rn progeny formed under controlled conditions. Hence, in the ANSTO system ^{222}Rn activity concentration in the sampled air is measured directly through its newly formed progeny within the controlled environment of the delay chamber (Whittlestone and Zahorowski, 1998; Zahorowski et al., 2004; Chambers et al., 2011). In routine operation, ANSTO monitors are calibrated monthly by injecting ^{222}Rn from a well characterized (to about ±4%) ^{226}Radium source. For ambient air measurements at 1 Bq m^{-3} activity concentration, the total uncertainty of hourly measurements is of order 10%, which includes uncertainty in flow rate as well as counting statistics. ANSTO two-filter detectors have a response time of around 45min, and are quite bulky (~3m^2) which can hinder their deployment in constricted locations.

(2) The one-filter methods used at the European stations are all based on the direct collection and counting of the short-lived ambient ^{222}Rn and ^{220}Rn (^{212}Pb) decay products which are attached to aerosols in the sampled air. These decay products are accumulated on either static or moving aerosol filters and measured by α or β spectroscopy (see references given in **Table 1**). In order to derive the atmospheric ^{222}Rn activity concentration, this method requires corrections for the atmospheric radioactive disequilibrium between the measured ^{222}Rn daughters (^{214}Po and/or ^{218}Po) and ^{222}Rn (e.g., Levin et al., 2002).

We use ^{222}Rn activity concentration measurements from 10 European stations over the 2006-2011 period (**Figure 1** and **Table 1**). The data from the different stations have been harmonized based on an extensive comparison study performed within the InGOS project (Schmithuesen et al., 2016). Based on the tall tower measurements at Cabauw and Lutjewad conducted at different heights above ground level as well as on an earlier comparison at Schauinsland station (Xia et al., 2010) and new comparison measurements in Heidelberg with an ANSTO system, correction factors for disequilibrium have also been estimated (Schmithuesen et al., 2016). All data used in the present study have been corrected accordingly and brought to a common ANSTO scale. A typical uncertainty of ^{222}Rn data from the different one-filter systems, including the uncertainty of the disequilibrium is estimated to 10-15%.

At the monitoring station Ispra, ^{222}Rn activity concentration has been measured using an ANSTO instrument, sampling air at an inlet positioned at 3.5m above the ground, close to the GHG-sampling mast with a height of 15m. Recent additional ^{222}Rn measurements using the 15m inlet

of the GHG mast (employing an Alphaguard PQ2000 (Genitron) instrument, calibrated against the ANSTO monitor) revealed significant differences of the [222]Rn activity at the two sampling heights during periods with low wind speeds. These differences showed that there are significant vertical [222]Rn gradients close to the ground. Based on the comparison of the two sampling heights during a 3-month period, we derive a wind-speed dependent correction, in order to 'normalize' the entire time series of the ANSTO measurements (at 3.5m above ground) to the 15m inlet, which is considered to be more representative. The uncertainty of this wind-speed dependent correction (based on the 1σ standard deviation during the 3-month comparison) is included in the time series shown in the Supplement (**Figure S24**).

3. Model simulations

3.1. TM5 Model

TM5 is a global chemistry transport model, which allows two-way nested zooming (Krol et al., 2005). In this study we apply the zooming with $1^{o} \times 1^{o}$ resolution over Europe, while the global domain is simulated at a horizontal resolution of 6° (longitude) $\times$ 4° (latitude). TM5 is an offline transport model, driven by meteorological fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) ERA-Interim reanalysis (Dee et al., 2011). The spatial resolution of this data set is approximately 80 km (T255 spectral) on 60 vertical levels from the surface up to 0.1 hPa. We employ the standard TM5 version with 25 vertical levels, defined as a subset of the 60 layers of the ERA-Interim reanalysis. The extraction of the meteorological fields is performed through a pre-processing software, which supplies fully consistent meteorology data with those of ECMWF at the different spatial resolutions of TM5 (Krol et al., 2005). The boundary layer, the free troposphere, and the stratosphere are represented by 5 (up to 1 km), 10, and 10 layers, respectively. The temporal resolution of the data is 3-hourly for near surface data (e.g., BLHs) and 6-hourly for 3D fields (e.g., temperature, wind, humidity, and convection).

Tracers in TM5 are transported by advection (in both horizontal and vertical directions), cumulus convection, and vertical diffusion. Tracer advection is based on the so-called "slopes scheme" which considers a tracer mass within a grid cell as a mean concentration and the spatial gradient of the concentration within the grid box (Russel and Lerner, 1981), which is caused by the motion of the tracer into and out of the grid box. Non-resolved transport by shallow cumulus and deep convection in TM5 is parameterized by a bulk mass flux approach originally described in Tiedtke (1989). Such convective clouds are described by single pairs of entraining/detraining plumes representing the updraft/downdraft motion. The parameterization of the vertical turbulent diffusion in the boundary layer is based on the scheme of Holtslag and Moeng (1991), while the formulation of Louis (1979) is considered in the free troposphere. The BLH is computed by using the expression of Vogelezang and Holtslag (1986), as described in Section 2.1. The exchange coefficients from the vertical diffusion are combined with the vertical convective mass fluxes to calculate the sub-grid scale vertical tracer transport. After redistributing the tracer mass by convection and diffusion, the slopes are updated.
Recently, van der Veen (2013) proposed a revised scheme to update the slopes. This "revised slopes scheme" results in enhanced horizontal transport in TM5 by increasing the horizontal

diffusivity of the numerical scheme of the convection routine. van der Veen (2013) found an improvement of the inter-hemispheric mixing gradient in TM5, which was initially underestimated as reported in e.g., Patra et al. (2011). This "revised slopes scheme" has been used for the sensitivity tests described below. Furthermore, we performed sensitivity tests using directly the convection fields from the ECMWF IFS model, instead of the default convection scheme based on Tiedtke (1989). The ECMWF convection scheme includes several improvements of the parameterizations of deep convection, radiation, clouds and orography, introduced operationally since ECMWF ERA-15 analyses (e.g., Gregory et al., 2000; Jakob and Klein, 2000; Morcrette et al., 2001). Finally, we evaluate the combination of the "revised slopes scheme" and the use of ECMWF convection fields.

3.2. TM5 Boundary layer height scheme

In the TM5 model, the full expression of Vogelezang and Holtslag (1986) is used to compute R_{ib}, (Eq. 1). First, R_{ib} is computed at each model level by using the equation (1). The vertical profile of R_{ib} is then linearly interpolated between consecutive levels of the model. The BLH is defined as the height, where R_{ib} reaches the critical value R_{ic}. In TM5, R_{ic} is set to 0.3. As discussed in Seidel et al. (2012), the choice of R_{ic} value close to 0.25 does not introduce large uncertainty. Moreover, the differences between BLHs determined by using R_{ic} of 0.25 and R_{ic} of 0.3 are very small as shown in this study (see e.g., **Figures S2-S11** in the Supplement). In the TM5 model, the minimum BLH is set to 100 m.

In addition, we calculate the BLH in TM5 based on the definition of Seidel et al. (2012) as used in the InGOS model validation exercise (i.e., R_{ic} = 0.25 and both surface wind and friction velocity are set to zero in Eq.1; see Section 2.1). Furthermore, because InGOS and IGRA sites are not co-located, we extract the BLH in the model both at the location of the InGOS station and at the location of the nearest IGRA station, resulting two sets of modeled BLHs labelled by the following acronyms:
- 'TM5_INGOS': BLHs extracted at InGOS station
- 'TM5_INGOS_IGRA': BLHs extracted at IGRA station, which is closest to the selected InGOS station.

In both cases, we use a 2-dimensional interpolation (longitude / latitude) to the location of the (InGOS or IGRA) station.
Furthermore, we extract also the default TM5 BLH (both at the InGOS and IGRA station) and the BLHs from ECMWF reanalyses. In general, the difference between the BLH based on Seidel et al. (2012) and the TM5 default and ECMWF BLHs are very small. Therefore, the latter are only shown in the Supplement (**Figures S2-S1**1).

3.3 InGOS ^{222}Rn flux map

We use the new ^{222}Rn flux map developed by Karstens et al. (2015) within the InGOS project (called hereafter 'InGOS ^{222}Rn flux map'). This map is based on a parameterization of ^{222}Rn production and transport in the soil, using a deterministic model based on the equations of

continuity and diffusion (Fick's 1st law) to compute the transport of the ^{222}Rn flux from the soil to the atmosphere. The modelled radon flux is dependent on soil porosity and moisture, with the latter obtained from two different soil moisture data sets, i.e., from the Land Surface Model Noah (driven by NCEP-GDAS meteorological reanalysis and part of the Global Land Data Assimilation System (GLDAS); Rodell et al., 2004), and from the ERA-Interim/Land reanalysis, respectively. Karstens et al. (2015) found that the flux estimates based on the GLDAS Noah soil moisture model on average better represent observed fluxes. Therefore, we apply in this study the ^{222}Rn flux map version based on the Noah soil moisture data set. Furthermore, the ^{222}Rn flux map considers the water table (from a hydrological model simulation), the distribution of the ^{226}Ra content in the soil, and the soil texture. For comparison, we apply also the commonly used constant emission maps with uniform continental ^{222}Rn exhalation of 21.98mBqm^{-2}s^{-1} between 60°S and 60°N; uniform continental ^{222}Rn emissions of 11.48 mBqm^{-2}s^{-1} between 60°N and 70°N (excluding Greenland); and zero flux elsewhere (Jacob et al., 1997). The InGOS ^{222}Rn flux map provides monthly ^{222}Rn fluxes over the 2006-2011 period, aggregated to 0.5°×0.5° grid for Europe and complemented by the constant emissions for the regions outside Europe. **Figures 2a and 2b** illustrate the spatial and mean seasonal variations of the ^{222}Rn fluxes from the InGOS ^{222}Rn flux map over Europe. The modelled ^{222}Rn flux is found to be larger in the areas where the ^{226}Ra activity concentration in the upper soil is very high, such as the Iberian Peninsula, areas in Central Italy and the Massif Central in Southern France (**Figure 2a**). The mean seasonal variations of the ^{222}Rn fluxes are mainly driven by the soil moisture. On average, the InGOS ^{222}Rn emissions over Europe are smaller than the constant emission (except July - September; **Figure 2b**).

3.4. Simulated ^{222}Rn activity concentrations

We simulate ^{222}Rn activity concentrations using either the InGOS ^{222}Rn flux map based on Noah soil moisture data, or constant ^{222}Rn fluxes (see Section 3.3). Furthermore, we apply also the "revised slopes scheme" and the updated convection scheme based on ECMWF reanalyses (see section 3.1) for the InGOS ^{222}Rn flux map based simulations only. These different simulations are labelled by the following acronyms:

- FC_CT: constant ^{222}Rn fluxes, and default convection scheme in TM5 based on Tiedtke (1989)
- FI_CT: InGOS ^{222}Rn flux map, and default convection
- FI_CU: InGOS ^{222}Rn flux map by using both the "revised slopes scheme" and the convection scheme based on ECMWF reanalyses

We also analyzed the use of "revised slopes scheme" and the updated convection scheme independently (see Supplement; **Figures S14-S24**)

The model simulations are 3-D linearly interpolated (i.e. horizontally and vertically) to the location of the station, and averaged over 1 hour.

4. Results

4.1. Simulated boundary layer heights versus observations

We focus the analysis on the InGOS stations (measuring CH_4 and N_2O, and / or ^{222}Rn activity concentrations; **Figure 1**) at low altitudes (i.e., excluding mountain stations) and compare the modelled BLHs with observations at the closest IGRA stations. **Figures 3** and **4** show the mean seasonal variation for the nocturnal (00 UTC) and daytime (12 UTC) BLH, respectively (2006-2010 average). The nocturnal BLHs show a clear seasonal cycle at most stations, with typically higher nocturnal BLHs during winter (but also larger range between 25% and 75% percentile) compared to summer. This seasonal pattern is very consistent between measurements and model simulations. However, at some continental stations (e.g. Heidelberg, Gif-sur-Yvette) the IGRA data show very low nocturnal BLHs (median value below 100m) during summer, which are not reproduced by the models. In general, the Whisker plots (**Figure 3**) show a skewed (non-normal) distribution for most monthly data (observations and model simulations) with the median value being usually significantly lower than the mean. The daytime BLHs show a very pronounced seasonal cycle at most continental stations (opposite in phase with the seasonal cycle of the nocturnal BLH), with typical values around 500m during winter, and ~1000-2000m during summer. The daytime BLH is in general relatively well simulated at most stations, as further illustrated by the ratios between modelled and observed BLHs, which are close to 1 (see **Figure 8**). An exception, however, are coastal sites (e.g., Angus, Mace Head), where apparently the model representation errors (e.g., transition between land and sea) are a limiting factor. In general, it should be expected that the model BLH extracted at the location of the IGRA station should agree better than that extracted at the InGOS station (see Section 3.2 for the definition of the model BLHs). However, e.g. at Egham the opposite is the case, since the IGRA station (Herstmonceaux) is closer to the coast, and the corresponding model BLH has more 'marine' character (and the transition zone between sea and land is not resolved by the model). For most stations far from the coast, however, the difference between the BLH at the InGOS station and the IGRA station is usually very small (**Figures 3 and 4 and Figures S2-S11** in the Supplement). Compared to the data for the nocturnal BLH, the daytime BLHs show much smaller difference between median and mean value, indicating a less skewed frequency distribution (**Figures 3 and 4**).

In the supplement (**Figures S2** to **S11**) we show the full time series for the 10 stations in 2009, illustrating that also the synoptic variability of the BLH is relatively well reproduced by the models (for both nocturnal and daytime BLH). Furthermore, we extend the analysis by using all IGRA stations over Europe (about 130 stations; see **Figure 1** and **Figures S12** and **S13** in the Supplement). This extended analysis confirms the major findings discussed above, especially (1) the good agreement between simulated and observed BLH during daytime, (2) the tendency for the simulated nocturnal BLHs to be too high during summer, and (3) larger differences between TM5 and IGRA BLHs for stations located close to the coasts.

In the following we include the ceilometer and lidar derived BLH at Cabauw and Trainou, respectively, in the analysis. As clearly visible from the correlation plot between ceilometer and IGRA data for Cabauw (**Figure S1**), the ceilometer BLHs during midday are usually lower than the IGRA data (especially for the period March to September), while modelled BLHs fall in between the two observational datasets (**Figure 5**). Part of this difference is likely due to the different methodologies. Hennemuth and Lammert (2006) pointed out that inconsistencies between the atmospheric thermal profile and the aerosol concentration profile can result in differences between radiosonde and lidar/ceilometer BLH retrievals. In addition, the spatial

separation between Cabauw and De Bilt (~23 km) combined with different surface characteristics (wetter soils in Cabauw and different large scale surface roughness) may play some role. While the correlation between IGRA BLHs and the ceilometer BLH retrievals at Cabauw is reasonable (r=0.63) during daytime, it is very poor during night (**Figure S1),** probably due to the issues of ceilometers to detect the shallow nocturnal BLH, as mentioned in Section 2.1.2. The lidar daytime data at Trainou for 2011 agree relatively well with the model BLHs (except May) (**Figure 5**). While no IGRA data are available for this period, the comparison between model simulations and IGRA for 2006-2010 at Trainou (**Figure 5**) shows similar (or slightly better) agreement as the comparison between lidar and model for 2011.

4.2. Simulated ^{222}Rn activity concentrations versus observations

Figures 6 and 7 show the mean seasonal variations of observed and simulated ^{222}Rn activity concentrations at each of the studied InGOS sites at 05 UTC (time around which typically the daily maximum ^{222}Rn activity concentration occurs) and at 14 UTC (^{222}Rn daily minimum), respectively. For most stations, TM5 simulated ^{222}Rn activity concentrations based on the InGOS ^{222}Rn flux map show significantly better agreement with observations than the simulations based on the constant ^{222}Rn flux, especially regarding the average seasonal variations. The improvement is largest during winter months, when TM5 simulations based on the constant ^{222}Rn fluxes often overestimate observations, while simulated concentrations based on the InGOS ^{222}Rn flux map are significantly lower owing to the lower ^{222}Rn fluxes (**Figure 6 and 7**). This, in turn, is driven mostly by the higher soil moisture and consequently lower permeability of the soil in winter. Furthermore, large differences are visible at many North European sites close to the coast (Angus, Lutjewad, Mace Head, Cabauw), where the water table can be very shallow, significantly reducing the ^{222}Rn fluxes (Karstens et al., 2015). Model simulations based on the InGOS ^{222}Rn flux map (which include modelled water table in the parameterization of ^{222}Rn fluxes) agree much better with observations than the control runs with constant ^{222}Rn fluxes. Despite the larger ^{222}Rn fluxes during summer, daily minimum ^{222}Rn concentrations in model and observations are usually lower at continental stations (e.g. Heidelberg, Gif-sur-Yvette) due to the much higher daytime boundary layer in summer compared to winter.
Figures S14 to S24 in the supplement show the full time series of simulated and observed ^{222}Rn concentrations at the 10 studied InGOS stations (with ^{222}Rn activity concentration observations available) for 2009.

Relationship between ^{222}Rn activity concentrations and boundary layer heights

In the following, we analyze the relationship between ^{222}Rn activity concentration and BLH in more detail. **Figure 9** shows the mean seasonal diurnal cycle of observed and simulated ^{222}Rn activity concentration and BLH for the four seasons at different sites. The figure illustrates the very strong anti-correlation between simulated BLH and ^{222}Rn activity concentration: The modelled BLHs increase sharply between 9:00 and 10:00 UTC (10:00/11:00 and 11:00/12:00 LT), resulting in an immediate decrease of modelled ^{222}Rn concentrations. In contrast, the ^{222}Rn activity concentration measurements show a slower decrease over several hours. Although this slow decrease may be partially due to the slow (45min) response time of the two-filter detectors, it is clear that the sharp changes in simulated BLHs and ^{222}Rn activity concentrations are due to the relatively coarse temporal resolution of ECMWF meteorological data (3-hourly for surface

data (e.g., BLHs) and 6-hourly for 3D fields (temperature, wind and humidity); see Section 3.1)..
Because the ceilometer data at Cabauw during night might be questionable, we included in
Figure 9 only the lidar measurements at Trainou (TR4). These show a much slower growth of
the BLH, starting in the morning and reaching its maximum in the late afternoon, as also
illustrated in Pal et al. (2012, 2015). Despite the obvious issue of the temporal resolution of the
model, however, inspection of **Figure 9** also indicates significant mismatches between simulated
and observed ^{222}Rn activity concentrations that cannot be explained wholly by problems with the
modeled BLH (even accounting for possible instrumental response time effects). Especially
during daytime, the TM5 BLHs are close to the IGRA measurements at most stations (as also
illustrated by the ratios of BLHs in **Figure 8**), whereas large differences are observed between
the simulated and measured ^{222}Rn activity concentrations at several stations. This is further
illustrated in **Figure 10**, where we compare the ratio of simulated to observed BLH with the ratio
of observed to simulated ^{222}Rn activity concentration during daytime for the different seasons. If
the ^{222}Rn activity concentration errors were purely due incorrect dilutions resulting from errors in
the modeled BLH at a given station, the two ratios would be similar. This is clearly not the case,
however, and the modelled afternoon concentration ratios range widely (from 0.2 to 1.8) from
station to station. These mismatches between observed and simulated ^{222}Rn activity
concentrations may be related to shortcomings of TM5 in correctly simulating the vertical ^{222}Rn
activity concentration gradients within the boundary layer (see below). Furthermore, it is
important to consider the uncertainties of the InGOS ^{222}Rn flux map. Karstens et al. (2015)
estimated that the most important uncertainty in the InGOS ^{222}Rn flux is due to the uncertainties
in the soil moisture data. Altogether, the uncertainties in modelled ^{222}Rn fluxes for individual
pixels ($0.083^{o} \times 0.083^{o}$) are estimated to be about 50%. Karstens et al. (2015) pointed out that the
uncertainty of the ^{222}Rn fluxes averaged over the footprint of the measurements might be
smaller. However, the uncertainties of neighboring pixels in the InGOS ^{222}Rn flux map are likely
to be strongly correlated, and therefore the reduction of the relative uncertainty (integrated over a
typical footprint on the order of 50-200km) is probably relatively small. Assuming an overall
uncertainty of ~50% of the regional ^{222}Rn fluxes, the model simulations could be considered
broadly consistent with observations at most sites.

Sensitivity of simulated ^{222}Rn activity concentrations to convection scheme

The use of the new ECMWF based convection combined with the "revised slopes scheme" (i.e.,
FI_CU acronym in Section 3.4) results in a small decrease of simulated ^{222}Rn concentrations at
most stations, typically on the order of ~10-30% (**Figures 6-9**). However, root mean square
(RMS) and correlation coefficients are very similar at most sites for both convection
parameterizations (**Figure 8**). Hence, no clear conclusions can be drawn, which parameterization
is more realistic. At the same time, **Figure 8** demonstrates again the improvement using the
InGOS ^{222}Rn flux map, resulting in (1) ratios between simulated and observed ^{222}Rn activity
concentration closer to one, (2) lower RMS, and (3) higher correlation coefficients at several
stations, compared to the model simulations using constant ^{222}Rn fluxes. This highlights the
challenge to validate model simulations. The difference of ~10-30% of ^{222}Rn activity
concentrations using a different convection parameterization is expected to result in a difference
of similar order of magnitude for the GHG emissions derived in inverse modelling. First GHG
inversions with the new ECMWF based convection confirm that derived emissions change
significantly (not shown).

Comparison of simulated and observed ^{222}Rn activity concentrations: Impact of sampling time

Figure 10 illustrates further that the ratio between observed and simulated daytime ^{222}Rn activity concentration also depends on the exact hour, decreasing significantly between 12:00 and 15:00 UTC at several stations (very pronounced at Trainou and Ispra). This is clearly due to the shortcomings of TM5 to simulate the diurnal cycle in the BLH discussed above (owing to the coarse temporal resolution of the meteorological data). In the current TM5-4DVAR system the average (observed and simulated) concentrations between 12:00 and 15:00 LT are used to derive emissions (Bergamaschi et al., 2010; 2015). Given the too fast increase of the BLH and consequently too fast decrease of simulated mixing ratios in the morning transition period, the choice of the assimilation time window may introduce some systematic errors in the flux inversions.

In the analyses shown in **Figure 10**, the data include all stability regimes. In addition, we performed this analysis separately for unstable, neutral, and stable vertical mixing conditions. We used the bulk Richardson number calculated at the first level of the model. This extended analysis, however, showed relatively similar model performance for these different weather conditions (results not shown). A limitation of this exercise is that for both stable and neutral stability regimes, we had at most stations only few cases by seasons.

Vertical gradients of ^{222}Rn activity concentrations in the boundary layer at Cabauw

Finally, we explore the vertical gradients of TM5 simulated ^{222}Rn activity concentrations at Cabauw where measurements are available at two vertical levels (20m [CB1] and 200m [CB4] height ; Table 1). The measurement height of 20m is within the first model layer, while 200m is within layer 3. **Figure 11** shows the monthly mean diurnal variations of modeled and observed vertical gradients of ^{222}Rn activity concentrations for each month for 2009. Although the InGOS ^{222}Rn flux based model simulations agree better with observations (in terms of ^{222}Rn activity concentrations; see **Figures 6, 7, and 8**) compared to the model simulations based on constant fluxes, this is not the case for the ^{222}Rn gradients for some months: between June and November the modelled gradients based on the constant fluxes agree better with observations, which could point to partially compensating systematic errors (e.g. too high ^{222}Rn fluxes might be compensated by too fast vertical mixing) During large parts of the year, the InGOS ^{222}Rn flux based model simulations underestimate the observed gradients. This is further illustrated in the scatter plots shown in **Figure 12** (separately for 00 and 12 UTC). For inverse modelling, especially the underestimated vertical gradient during daytime is critical and could lead to biases in the GHG inversions. Furthermore, **Figures 11** shows that during the transition phase in the morning the modelled ^{222}Rn activity concentration vertical gradient decreases faster than the observed gradient, which is probably largely due to the coarse time resolution of the meteorological data in TM5 together with the slow response time of the two-filter radon measurements, although it may also indicate that vertical mixing is proceeding too rapidly in the model.

5. Conclusions

In the first part of this study, we evaluated the boundary layer dynamics of the TM5 model by comparison with BLHs from the NOAA IGRA radiosonde data as well as with BLH retrievals from a ceilometer at Cabauw and lidar at Trainou.

TM5 reproduces reasonably well the IGRA BLHs during daytime within 10-20% (which is within the uncertainty of the IGRA data) for continental stations at low altitudes. During night, the model overestimates the shallow nocturnal BLHs, especially for very low BLHs (<100 m) observed during summer time. At coastal sites, the differences between simulated BLH and IGRA observations (both day and nighttime) are usually larger due to model representation errors (since the transition zone between the marine boundary layer over sea and the continental boundary layer over land is not resolved by the model).

The BLH retrievals at Cabauw show a reasonable correlation with IGRA data from De Bilt at 12 UTC, but are systematically lower. During night (00 UTC), however, the two data set show only a very poor correlation. Besides the fundamental differences in the BLH retrieval methods, however, also the spatial separation between Cabauw and DeBilt (~23 km) probably contributes to the differences in the derived BLH. For the lidar BLH data from Trainou, no direct comparison with the IGRA data is available (due to different time periods), but the comparison with the modelled BLH show similar agreement with the two different observational datasets [IGRA: for 2006-2010; lidar: 2011]. For the better exploitation of ceilometer / lidar data in the future, the further development of BLH retrievals is essential to ensure consistency between the different methods.

In the second part of this study, we compared TM5 simulations of ^{222}Rn activity concentrations with quasi-continuous ^{222}Rn measurements from 10 European monitoring stations.

The ^{222}Rn activity concentration simulations based on the new ^{222}Rn flux map show significant improvements compared to ^{222}Rn simulations using constant ^{222}Rn fluxes, especially regarding the average seasonal variability and generally lower simulated ^{222}Rn activity concentrations at North European sites close to the coast. These improvements highlight the benefit of the process-based approach, including a parameterization of water table (Karstens et al., 2015). Nevertheless, the (relative) differences between simulated and observed daytime minimum ^{222}Rn concentrations are larger for several stations (on the order of 50%) than the (relative) differences between simulated and observed BLH at noon. This is probably partly related to the uncertainties in the ^{222}Rn flux map (estimated to be on the order of 50%). In addition, however, also potential shortcomings of TM5 to correctly simulate the vertical ^{222}Rn activity concentration gradients are likely to play a significant role, which may be caused by the vertical diffusion coefficients and/or the limited vertical resolution in the model.

The comparison of simulated ^{222}Rn activity concentrations with measurements at Cabauw (20 m versus 200 m) shows that the model underestimates the measured vertical gradient (i.e., differences of concentrations between 20m and 200m levels) at this station. Furthermore, the sharp increase of the modeled BLH in the morning transition period results in a rapid decrease of the simulated ^{222}Rn activity concentrations, while ^{222}Rn measurements show a slower decrease at many stations. Although this latter timing effect may be partially due to the slow (45min) response time of the two-filter radon detectors, it is clear that the current coarse temporal resolution of the TM5 meteorological data (3-hourly for surface data and 6-hourly for 3D fields) limits the capability of simulating the diurnal cycle realistically. These issues probably lead to

systematic biases in inversions of GHG emissions. An updated TM5-4DVAR system is currently under development with increased temporal resolution of the meteorological data (3-hourly ECMWF data, interpolated to observational data time).

Finally, we evaluated the "revised slopes scheme" and the new ECMWF based convection scheme in the TM5 model. The results show a relatively small impact of the new slopes treatment, but a significant impact of the new ECMWF convection scheme, leading to significantly lower ^{222}Rn activity concentrations (about 20%) during daytime, especially in winter. While this is expected to have a significant impact on derived emissions in GHG inversions, the comparison with the available European ^{222}Rn activity concentration observations showed very similar performance. Hence, no clear conclusion about which parameterization is more realistic can be drawn from this study. These findings highlight the challenges of validating atmospheric transport models with the accuracy required to better evaluate and improve the quality of GHG flux inversions. In order to improve the validation capabilities it would be important (1) to increase the number of ^{222}Rn monitoring stations, (2) to perform vertical ^{222}Rn activity concentration profile measurements at tall towers and also from aircraft (e.g. Chambers et al., 2011; Williams et al., 2011, 2013), (3) to extend the validation of the ^{222}Rn inventories by local/regional ^{222}Rn flux measurements, (4) to further develop the BLH retrievals from ceilometer / LIDAR instruments and (5) to further extend the ceilometer / LIDAR network. More work is also needed to improve the representation of the nocturnal boundary layer in global and regional models. The use of ^{222}Rn in the diagnosis of the nocturnal mixing effects is one area showing promise in this regard (Williams et al., 2013).

Code availability

Further information about the TM5 code can be found at http://tm5.sourceforge.net/. Readers interested in the TM5 code can contact Maarten Krol (maarten.krol@wur.nl), Arjo Segers (arjo.segers@tno.nl) or Peter Bergamaschi (peter.bergamaschi@jrc.ec.europa.eu)

References

Angevine, W., Grimsdell, A., Hartten, L. M., and Delany, A. C.: The Flatland Boundary Layer Experiments, Bulletin of the American Meteorological Society 79, 419–431, 1998.

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., A. Segers, A., van Aardenne, J., Monni, S., Vermeulen, A., Schmidt, M. Ramonet, M., Yver, C., F. Meinhardt, F., Nisbet, E. G., R. Fisher, R., S. O'Doherty, S., and Dlugokencky, E.J.: Inverse modeling of European CH4 emissions 2001-2006, J. Geophys. Res., 115(D22309), doi:10.1029/2010JD014180, 2010.

Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S., and Dlugokencky, E.: Top-down estimates of European CH4 and N2O emissions based on four different inverse models, Atmos. Chem. Phys., 15, 715-736, doi:10.5194/acp-15-715-2015, 2015

Biraud, S., Ciais, P., Ramonet, M., Simmonds, P., Kazan, V., Monfray, P., O'Doherty, S., Spain, T. G., and Jennings, S. G.: European greenhouse gas emissions estimated from continuous atmospheric measurements and radon 222 at Mace Head, Ireland, J. Geophys. Res., 105(D1), 1351–1366, 2000

Broquet, G., Chevallier, F., Bréon, F.-M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmos. Chem. Phys., 13, 9039-9056, doi:10.5194/acp-13-9039-2013, 2013.

Carouge, C., Bousquet, P., Peylin, P., Rayner, P. J., and Ciais, P.: What can we learn from European continuous atmospheric CO2 measurements to quantify regional fluxes – Part 1: Potential of the network, Atmos. Chem. Phys. Discuss., 8, 18591-18620, doi:10.5194/acpd-8-18591-2008, 2008

Chambers, S., Williams, A.G., Zahorowski, W., Griffiths, A. and Crawford, J.: Separating remote fetch and local mixing influences on vertical radon measurements in the lower atmosphere. Tellus 63B, 843–859. doi:10.1111/j.1600-0889.2011.00565.x, 2011

Chevillard, A., Ciais, P., Karstens, U., Heimann, M., Schmidt, M., Levin, I., Jacob, D., Podzun, R., Kazan, V., Sartorius, H., and Weingartner, E.: Transport of 222 Rn using the regional model REMO: a detailed comparison with measurements over Europe, Tellus B, 54, 850–871, 2002

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., P. Kaufmann, P. and B. Calpini B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmos. Chem. Phys., 14, 13205-13221, 2014

Corazza, M., Bergamaschi, P., Vermeulen, A. T., Aalto, T., Haszpra, L., Meinhardt, F., O'Doherty, S., Thompson, R., Moncrieff, J., Popa, E., Steinbacher, M., Jordan, A., Dlugokencky, E., Brühl, C., Krol, M., and Dentener, F.: Inverse modelling of European N2O emissions: assimilating observations from different networks, Atmos. Chem. Phys., 11, 2381-2398, doi:10.5194/acp-11-2381-2011, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. doi: 10.1002/qj.828, 2011

Denning, A. S., Holzer, M., Gurney, K. R., Heimann, M., Law, R.M., Rayner, P. J., Fung, I. Y., Fan, S.-M., Taguchi, S., Friedlingstein, P., Balkanski, Y., Taylor, J., Maiss, M., and Levin, I.: Three-dimensional transport and concentration of SF6: A model intercomparison study (TransCom 2). Tellus 51B: 266-297, 1999

Dentener, F., Feichter, J., and Jeuken, A: Simulation of the transport of Rn222 using on-line and off-line global models at different horizontal resolutions: a detailed comparison with measurements, Tellus, 51B, 573-602, 1999.

Eresmaa, N., Karppinen, A.,Joffre, S. M., Räsänen, J. and Talvitie, H: Mixing height determination by ceilometers, Atmos. Chem. Phys., 6, 1485–1493, 2006.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D .E., Sturges, W. T., Moncrieff, J. B., and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmos. Chem. Phys., 15, 6393-6406, doi:10.5194/acp-15-6393-2015, 2015.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res., 108(D24), 4756, doi:10.1029/2002JD003018, 2003

Gregory, D., Morcrette, J. -J., Jakob, C., Beljaars, A. M., and Stockdale, T.: Revision of convection, radiation and cloud schemes in the ECMWF model, Q. J. R. Meteorol. Soc., 126, 1685–1710, 2000

Griffiths, A. D., Parkes, S. D., Chambers, S. D., McCabe, M. F., and A. G. Williams, A. G.: Improved mixing height monitoring through a combination of lidar and radon measurements, Atmos. Meas. Tech. , 6, 207–218, 2013

Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S.,Gobbi, G. P., Lolli, S., O'Dowd, C. D., Sauvage, L., Xueref. Remy, I., Wastine, B., and Feist, D. G.: Evaluation of mixing-height

retrievals from automatic profiling lidars and ceilometers in view of future integrated networks in Europe, Bound.-Layer Meteorol., 143, 49–75,doi:10.1007/s10546-011-9643-z , 2012

Haij, M. J. de., Wauben, W. M. F., and Baltink, H. K.: Determination of mixing layer height from ceilometer backscatter profiles, Proceedings of SPIE, Vol. 6362, 63620R (2006); doi:10.1117/12.691050, 11/9/2006-14/9/2006, James R. Slusser, Klaus Schäfer, Adolfo Comerón (Ed), 2006, Stockholm, Zweden, SPIE., 2006

Hatakka, J., Aalto, T., Aaltonen, V., Aurela, M., Hakola, H., Kompula, M., Laurila, T., Lihavainen, H., Paatero, J., Salminen, K., and   Viisanen, Y.: Overview of the atmospheric research activities and results at Pallas GAW station, Boreal Environ. Res., 8, 365–383, 2003

Hennemuth, B. and Lammert, A: Determination of the convective boundary layer height from radiosonde and lidar backscatter Boundary-Layer Meteorology, 120, 181-209, 2006

Holtslag, A. A. M., and Moeng, C. H.: Eddy diffusivity and counter-gradient transport in the convective atmospheric boundary layer, J. Atmos. Sci., 48, 1690-1698, 1991

Jacob, D. J., and Prather, M. J.: Radon-222 as a test of convective transport in a general circulation model, Tellus B, 42, 118–134, 1990

Jacob, D. J., Prather, M. J., Rasch, P. J., Shia, R.-L., Balkanski, Y. J.,Beagley, S. R., Bergmann, D. J., Blackshear, W. T., Brown, M., Chiba, M., Chipperfield, M. P., de Grandpré, J., Dignon, J. E., Feichter, J., Genthon, C., Grose, W. L., Kasibhatla, P. S., Köhler, I., Kritz, M. A., Law, K., Penner, J. E., Ramonet, M., Reeves, C. E., Rotman, D. A., Stockwell, D. Z., Van Velthoven, P. F. J., Verver, G., Wild, O., Yang, H., and Zimmermann, P.: Evaluation and intercomparison of global atmospheric transport models using 222Rn and other short-lived tracers, J. Geophys. Res., 102, 5953–5970, doi: 10.1029/96JD02955, 1997

Jakob, C., and Klein, S. A.:, A parametrization of cloud and precipitation overlap effects for use in General Circulation Models, Quart. J. Roy. Meteorol. Soc, 126, 2525-2544, 2000

Karstens, U., Schwingshackl, C., Schmithüsen, D., and Levin, I.: A process-based 222radon flux map for Europe and its comparison to long-term observations, Atmos. Chem. Phys., 15, 12845-12865, doi:10.5194/acp-15-12845-2015, 2015

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmos. Chem. Phys., 5, 417-432, doi:10.5194/acp-5-417-2005, 2005

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., D. B. C., Kaplan, J. O., Houweling, S.,and Wofsy, S. C.: Emissions of CH4 and N2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, Geophys. Res. Lett., 35, L18808, doi: 10.1029/2008GL034031, 2008

Levin, I., Born, M., Cuntz, M., Langendörfer, U., Mantsch, S., Naegler, T., Schmidt, M., Varlagin, A., Verclas, S., and D. Wagenbach, D.: Observations of atmospheric variability and soil exhalation rate of Radon-222 at a Russian forest site: Technical approach and deployment for boundary layer studies, Tellus, 54B, 462-475, 2002

Lopez, M., Schmidt, M., Yver, C., Messager, C., Worthy, D., Kazan, V., Ramonet, M., Bousquet, P., and Ciais, P.: Seasonal variation of N2O emissions in France inferred from atmospheric N2O and Rn-222 measurements, J Geophys. Res.-Atmos., 117, D14103, doi:10.1029/2012jd017703, 2012

Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreux, F., Grandpeix, J.-Y., Lefebvre, M.-P., Rio, C., Bergamaschi, P., Chambers, S. D., Karstens, U., Kazan, V., van der Laan, S., Meijer, H. A. J., Moncrieff, J., Ramonet, M., Scheeren, H. A., Schlosser, C., Schmidt, M., Vermeulen, A., and Williams, A. G.: Atmospheric transport and chemistry of trace gases in LMDz5B: evaluation and implications for inverse modelling, Geosci. Model Dev., 8, 129-150, doi:10.5194/gmd-8-129-2015, 2015

Louis, J. F. : A parametric model of vertical eddy fluxes in the atmosphere, Boundary Layer Meteorology, 17, 187-202, 1979

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., and R. G. Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, J. Geophys. Res., 116(D02305), doi:10.1029/2010JD014763, 2011

Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, Atmos. Chem. Phys., 8, 6341-6353, doi:10.5194/acp-8-6341-2008, 2008

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S., Dlugokencky, E., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, Proceedings of the National Academy of Sciences, doi:10.1073/pnas.1314392110, 2013

Morcrette, J.-J., Mlawer, E. J., Iacono, M. J., and Clough, S. A.: Impact of the radiation-transfer scheme RRTM in the ECMWF forecast system, Technical report in the ECMWF Newsletter, No. 91, 2001

Pal, S., Xueref-Remy, I., Ammoura, L., Chazette, P., Gibert, F., Royer, P., Dieudonné, E., J.-C. Dupont, J.-C., Haeffelin, M., Lac, C., Lopez, M., Morille, Y., Ravetta , F.: Spatio-temporal variability of the atmospheric boundary layer depth over the Paris agglomeration: An assessment of the impact of the urban heat island intensity, Atmospheric environment, Elsevier, 2012, 63, pp.261-275. <10.1016/j.atmosenv.2012.09.0, 2012

Pal, S., Haeffelin,M., and E. Batchvarova, E.: Exploring a geophysical process-based attribution technique for the determination of the atmospheric boundary layer depth using aerosol lidar and near-surface meteorological measurements, J. Geophys. Res. Atmos. 118, 9277–9295, doi:10.1002/jgrd.50710, 2013

Pal, S.: Monitoring Depth of Shallow Atmospheric Boundary Layer to Complement LiDAR Measurements Affected by Partial Overlap,Remote Sensing, 6(9), 8468-8493, 2014

Pal, S., Lopez, M., Schmidt, M., Ramonet, M., Gibert, F., Xueref-Remy, I., and Ciais, P.: Investigation of the atmospheric boundary layer depth variability and its impact on the 222Rn concentration at a rural site in France, J. Geophys. Res. Atmos., 120, 623–643, doi: 10.1002/2014JD022322, 2015

Patra, P. K., Houweling, S., Krol, M., Bousquet, P., Belikov, D., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Corbin, K., Fortems-Cheiney, A., Fraser, A., Gloor, E., Hess, P., Ito, A., Kawa, S. R., Law, R. M., Loh, Z., Maksyutov, S., Meng, L., Palmer, P. I., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C.: TransCom model simulations of CH4 and related species: linking transport, surface flux and chemical loss with CH4 variability in the troposphere and lower stratosphere, Atmos. Chem. Phys., 11, 12813-12837, doi:10.5194/acp-11-12813-2011, 2011

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, B. Am. Meteorol. Soc., 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004

Russell, G. L., and Lerner, J. A.: A new finite-differencing scheme for the tracer transport equation, Journal of Applied Meteorology, 20, 1483-1498, 1981

Scheeren, H. A. and Bergamaschi, P.: First Three Years of CO2 ,CH4, N2O, and SF6 Observations, and 222Radon-Based Emission Estimates from the JRC Monitoring Station at Ispra (Italy): What Have We Learned So Far?, proceeding of the 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2011), Wellington, New Zealand, World Meteorological Organization, 2012

Schmithüsen, D., Chambers, S., Fischer, B., Gilge, S., Hatakka, J., Kazan, V., Neubert, R., Paatero, J., Ramonet, M., Schlosser, C., Schmid, S., Vermeulen, A., and Levin, I: A European-wide ^{222}Radon and ^{222}Radon progeny comparison study, to be submitted to Tellus B, 2016

Seibert, P., Beyrich, F., Gryning, S. E., Joffre, S., Rasmussen, A., and P. Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmos. Environ, 34, 1001–1027, 2000

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J. -C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental United States and Europe, J. Geophys. Res.- Atmos, 117, 2012

Smallman, T. L., Williams, M., and Moncrieff, J. B.: Can seasonal and interannual variation in landscape CO_2 fluxes be detected by atmospheric observations of CO_2 concentrations made at a tall tower?, Biogeosciences, 11, 735–747, doi:10.5194/bg-11-735-2014, 2014

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, 666p, 1988

Taguchi, S., Law, R. M., Rödenbeck, C., Patra, P. K., Maksyutov, S., Zahorowski, W., Sartorius, H., and Levin, I.: TransCom continuous experiment: comparison of 222Rn transport at hourly time scales at three stations in Germany, Atmos. Chem. Phys., 11, 10071-10084, doi:10.5194/acp-11-10071-2011, 2011

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterisation in large scale models, Mon. Wea. Rev, 177, 1779–1800, 1989

van der Laan, S., Karstens, U., Neubert, R. E. M., van der Laan-Luijkx, I. T., and Meijer, H. A. J.: Observation-based estimates of fossil fuel-derived CO_2 emissions in the Netherlands using Delta 14C, CO and 222Radon, Tellus B, 62, 389–402, doi:10.1111/j.1600-0889.2010.00493.x, 2010

van der Veen, E., Optimizing transport properties in TM5 using SF6, Wageningen University, University of Twente, Master's thesis, 2013

Vermeulen, A. T., Hensen, A., Popa, M. E., van den Bulk, W. C.M., and Jongejan, P. A. C.: Greenhouse gas observations from Cabauw Tall Tower (1992–2010), Atmos. Meas. Tech., 4, 617–644, doi:10.5194/amt-4-617-2011, 2011

Vogelezang, D. H. P., and Holtslag, A. A., M.: Evaluation and model impacts of alternative boundary-layer height formulation, Boundary Layer Meteorol., 81, 245–269, 1986

Whittlestone, S., and W. Zahorowski, W.: Baseline radon detectors for shipboard use: Development and deployment in the First Aerosol Characterization Experiment (ACE 1), Journal of Geophysical Research, 103(D13), 16, 743-751, 1998

Williams, A. G,, Zahorowski, W, Chambers, S. D., and Griffiths, A.: The vertical distribution of radon in clear and cloudy daytime terrestrial boundary layers. J Atmos Sci 68, 155-174, 2011

Williams, A.G., Chambers, S.D. and Griffiths, A.D.: Bulk mixing and decoupling of the nocturnal stable boundary layer characterized using a ubiquitous natural tracer. Boundary-Layer Meteorol. 20 (149), 381-402. http://dx.doi.org/10.1007/s10546-013-9849-3, 2013

Xia, Y., Sartorius, H., Schlosser, C., St ohlker, U., Conen, F., and Zahorowski, W.: Comparison of one- and two-filter detectors for atmospheric 222Rn measurements under various meteorological conditions, Atmos. Meas. Tech., 3, 723–731, doi:10.5194/amt-3-723-2010, 2010

Yver, C., Schmidt, M., Bousquet, P., Zahorowski, W., and Ramonet, M.: Estimation of the molecular hydrogen soil up-take and traffic emissions at a suburban site near Paris through hydrogen, carbon monoxide, and radon-222 semi-continuous measurements, J. Geophys. Res., 114, D18304, doi:10.1029/2009JD012122, 2009

Zahorowski, W., Chambers, S. D., and Henderson-Sellers, A.: Ground based radon-222 observations and their application to atmospheric studies, J. Environ. Radioactivity, 76, 3–33, 2004

Table 1: Description of the different surface stations measuring ^{222}Rn activity concentrations. See Figure 1 for the locations of the stations shown by their ID. ANSTO stands for Australian Nuclear Science and Technology Organisation. The locations of the stations are shown in Figure 1. CB1 and CB4 are the level 1 and level 2 of the Cabauw tower, respectively. Altitude stands for terrain elevation referred to the mean sea level and Height is the above ground level

Station ID	Name	Country	Latitude (degree)	Longitude (degree)	Altitude/Height (m)	^{222}Rn instrument	Reference
PAL	Pallas	Finland	67.97	24.12	572/7	one-filter method	Hatakka et al. (2003)
TTA	Angus	United Kingdom	56.55	-2.98	363/50	ANSTO	Smallman et al. (2014)
LUT	Lutjewad	Netherlands	53.40	6.35	61/60	ANSTO	van der Laan et al. (2010)
MHD	Mace Head	Ireland	53.33	-9.90	40/15	one-filter method	Biraud et al. (2000)
CBW (CB1)	Cabauw	Netherlands	51.97	4.93	19/20	one-filter method	Vermeulen et al. (2011)
CBW (CB4)	Cabauw	Netherlands	51.97	4.93	199/200	ANSTO	Vermeulen et al. (2011)
EGH	Egham	United Kingdom	51.43	-0.56	45/10	one filter method	Levin et al. (2002)
GIF	Gif-sur-Yvette	France	48.71	2.15	167/7	one-filter method	Lopez et al. (2012), Yver et al. (2009)
HEI	Heidelberg	Germany	49.42	8.71	146/30	one-filter method	Levin et al. (2002)
TRN (TR4)	Trainou	France	47.95	2.11	311/180	ANSTO	Schmidt et al. (2014)
IPR	Ispra	Italy	45.80	8.63	223/3.5 (15)[1]	ANSTO	Scheeren and Bergamaschi (2012)

[1]measurements at 3.5m 'normalized' to sampling height of 15 m based on wind-speed dependent correction (see Section 2.2)
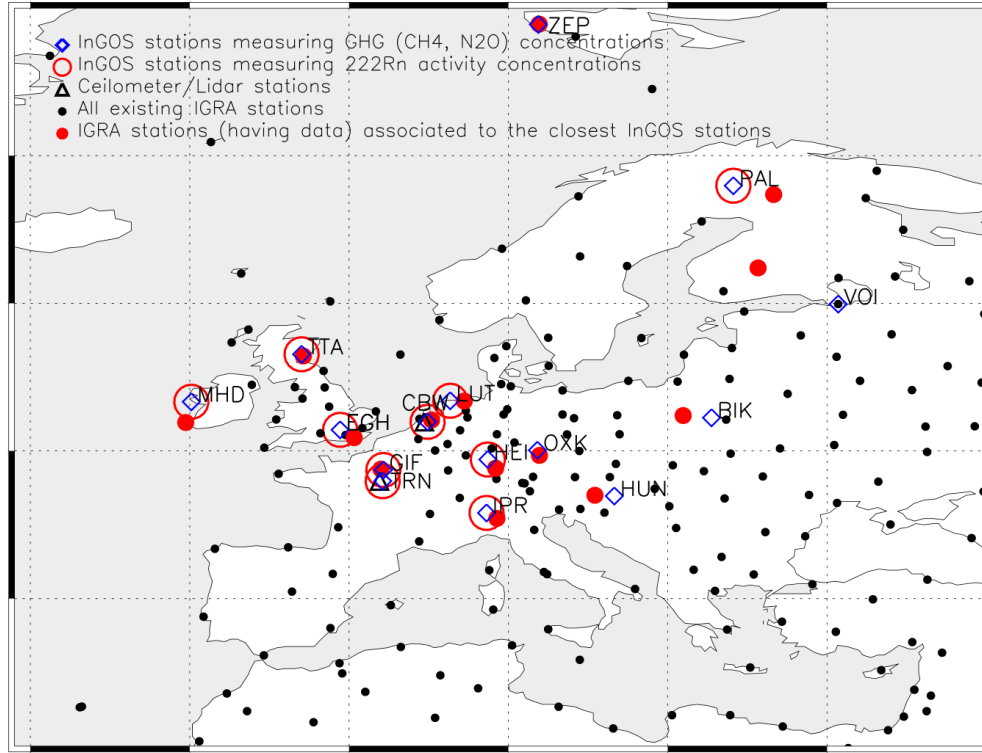
Figure 1: Observational network of InGOS greenhouse gas (CH$_4$, N$_2$O) and radon (^{222}Rn) concentration measurements and boundary layer height observations, blue diamonds ($\diamond$):INGOS stations that measure CH$_4$ and/or N$_2$O concentrations; red circles ($\circ$):InGOS stations that measure radon (^{222}Rn) activity concentrations; black dots ($\bullet$): all existing IGRA stations; red dots ($\bullet$): IGRA station closest to InGOS station; triangles (Δ):ceilometer/lidar measurement sites (i.e., Cabauw/Trainou). The acronyms for the stations measuring ^{222}Rn activity concentrations are compiled in Table 1.
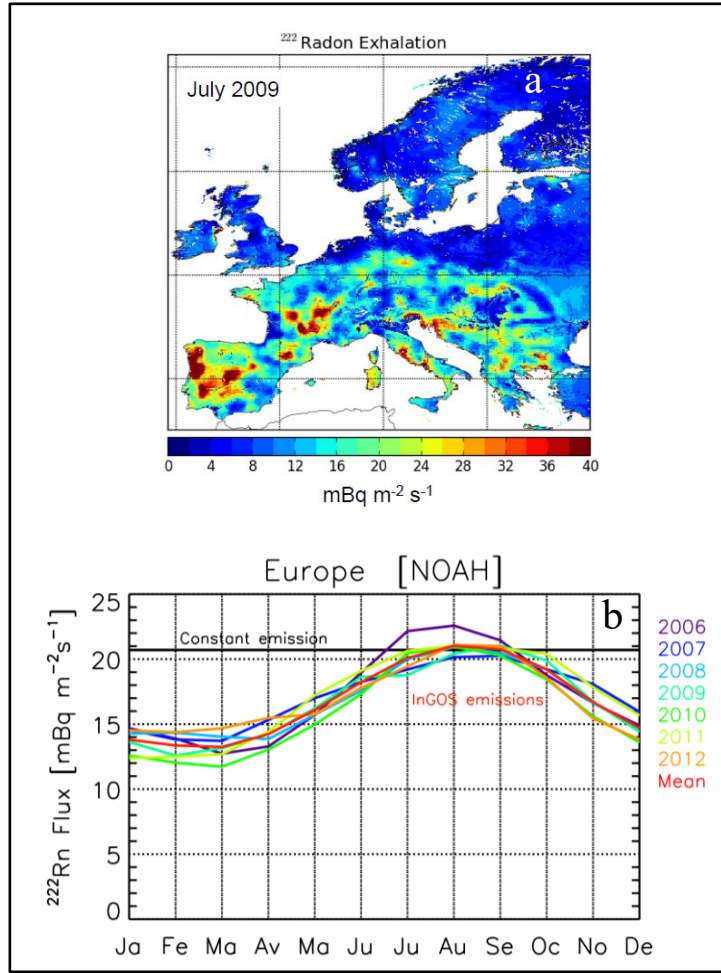
Figure 2: Radon (^{222}Rn) emissions used for the model simulations, (a) spatial distribution of InGOS emissions over Europe during July 2009, (b) seasonal and inter-annual variations of InGOS emissions (in different colors for different years; mean in red) and the commonly used constant emissions (black). The mean seasonal variations are averaged over the geographic domain between 10°W and 30°E longitude and between 35°N and 70°N latitude.
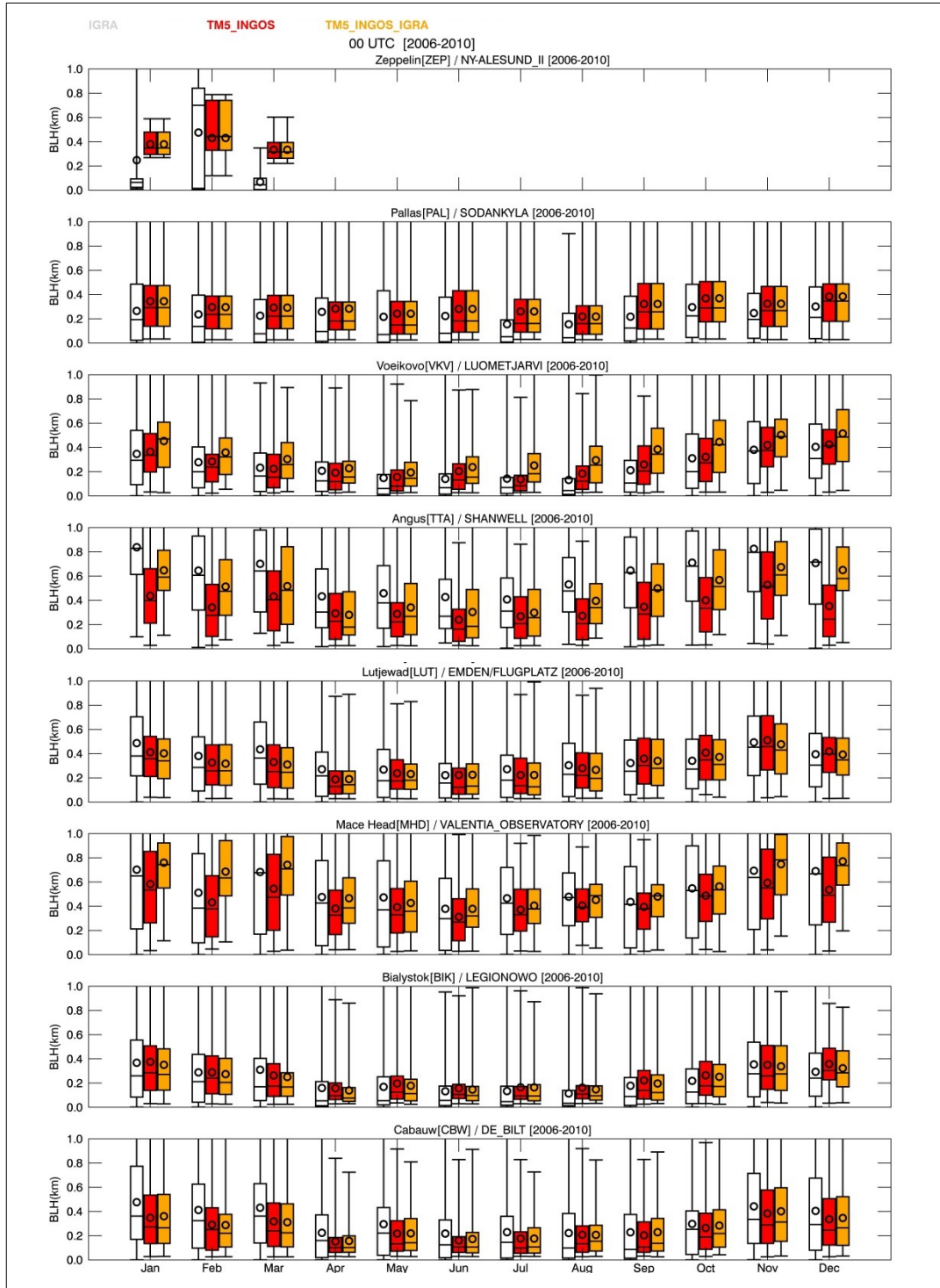
Figure 3: Observed (IGRA; blank) and modelled (TM5_INGOS; red and TM5_INGOS_IGRA; orange) BLHs for InGOS stations at 00 UTC (2006-2010). The titles of each panel show the names and acronyms of the InGOS station, and the names of the nearest IGRA station used for comparison. The Whisker plots show the monthly minimum and maximum values (bars), and the 25% and 75% percentiles (boxes). The median values are given by the horizontal line and the mean values by the open circles in the boxes. The different acronyms of the model data are defined in Section 3.2 of the text
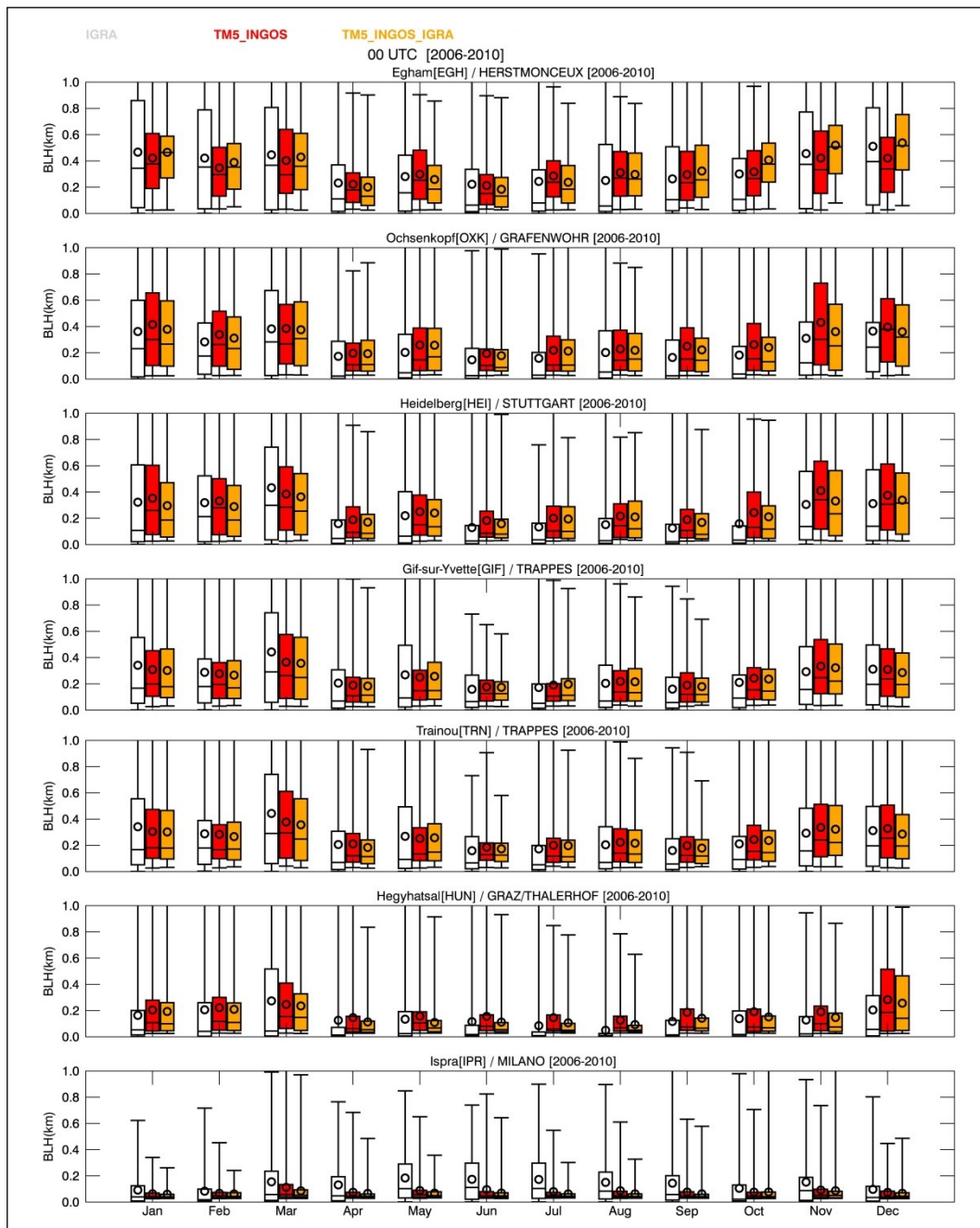
Figure 3: continued
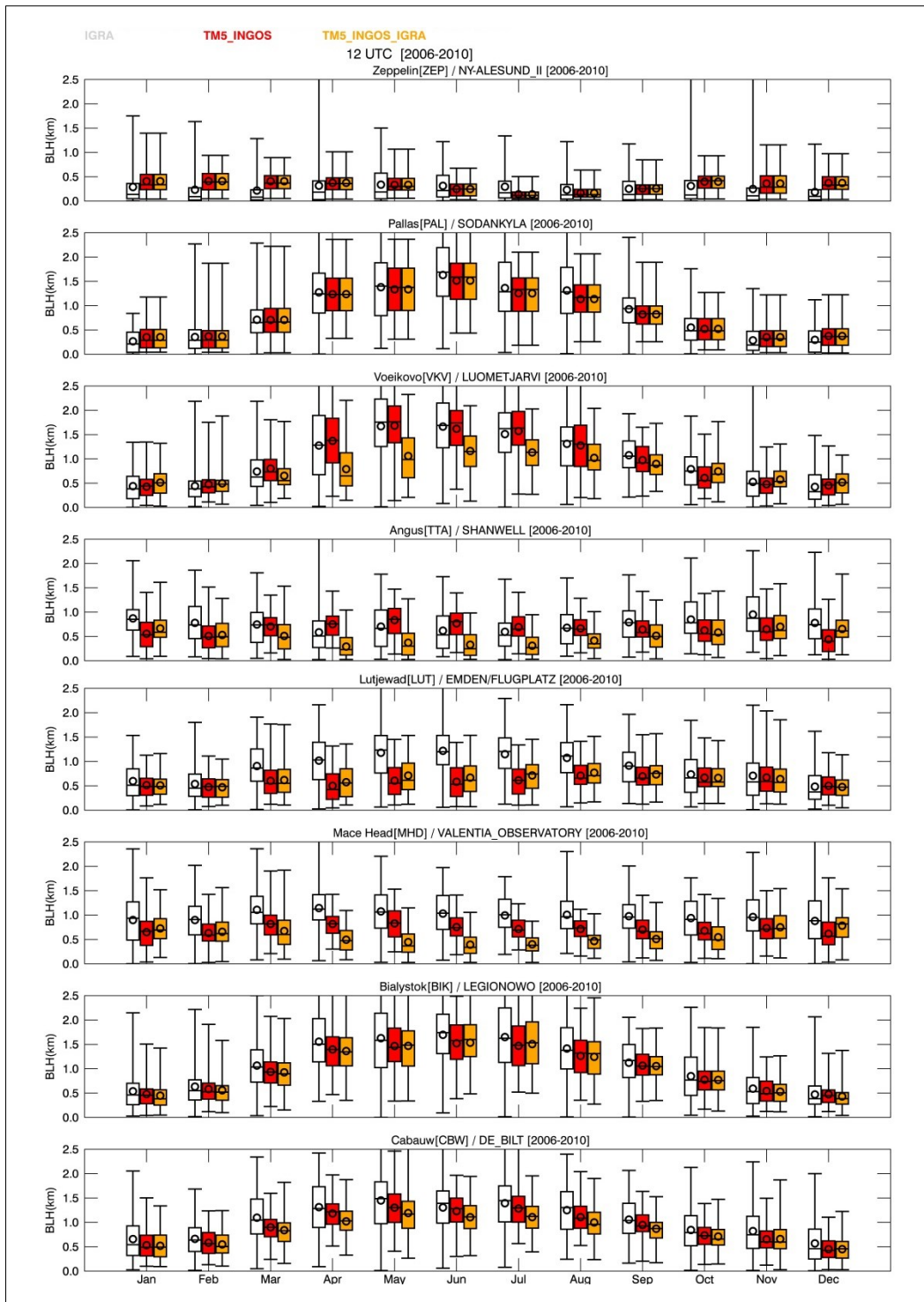
Figure 4: As Figure 3, but at 12 UTC
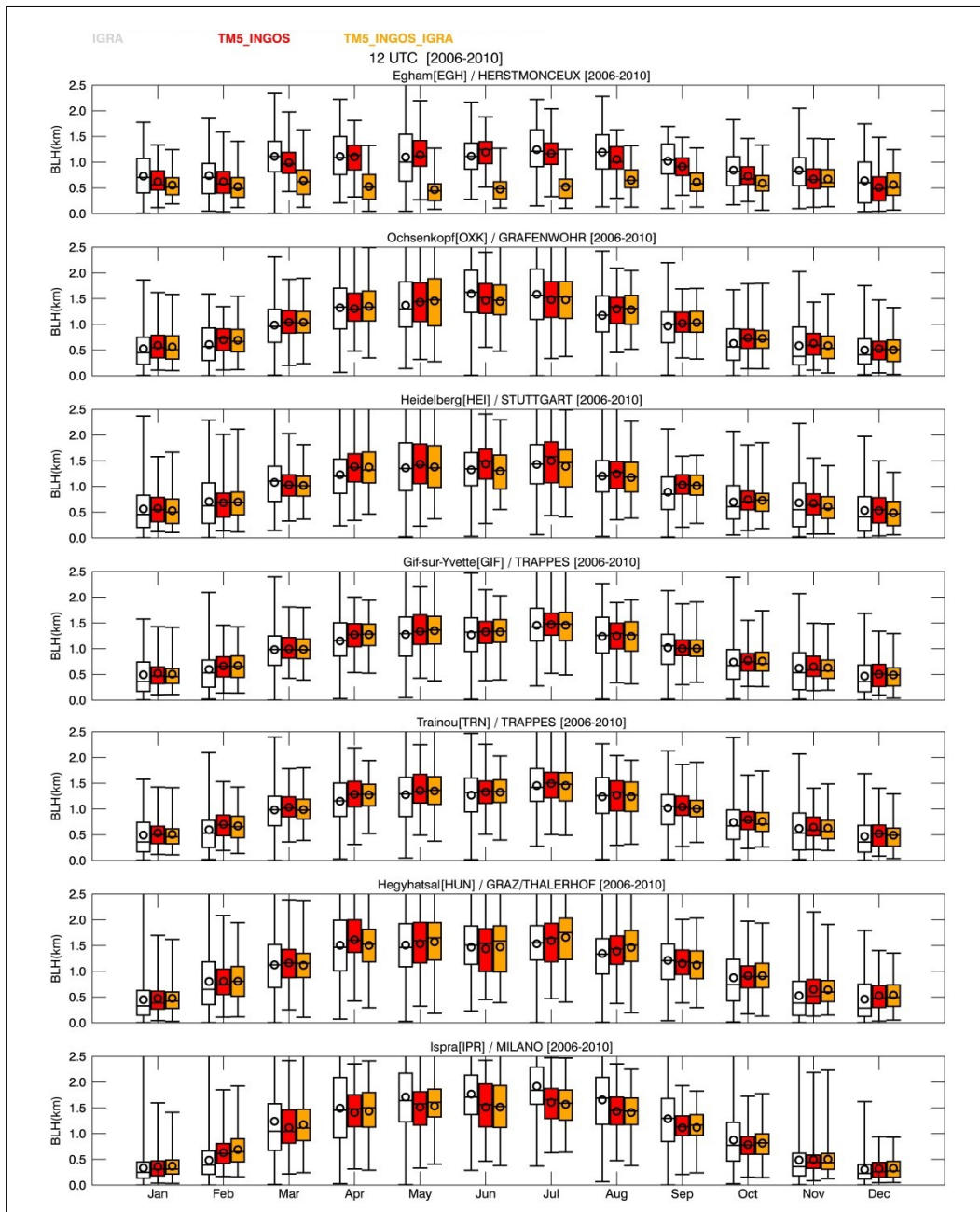
Figure 4: continued
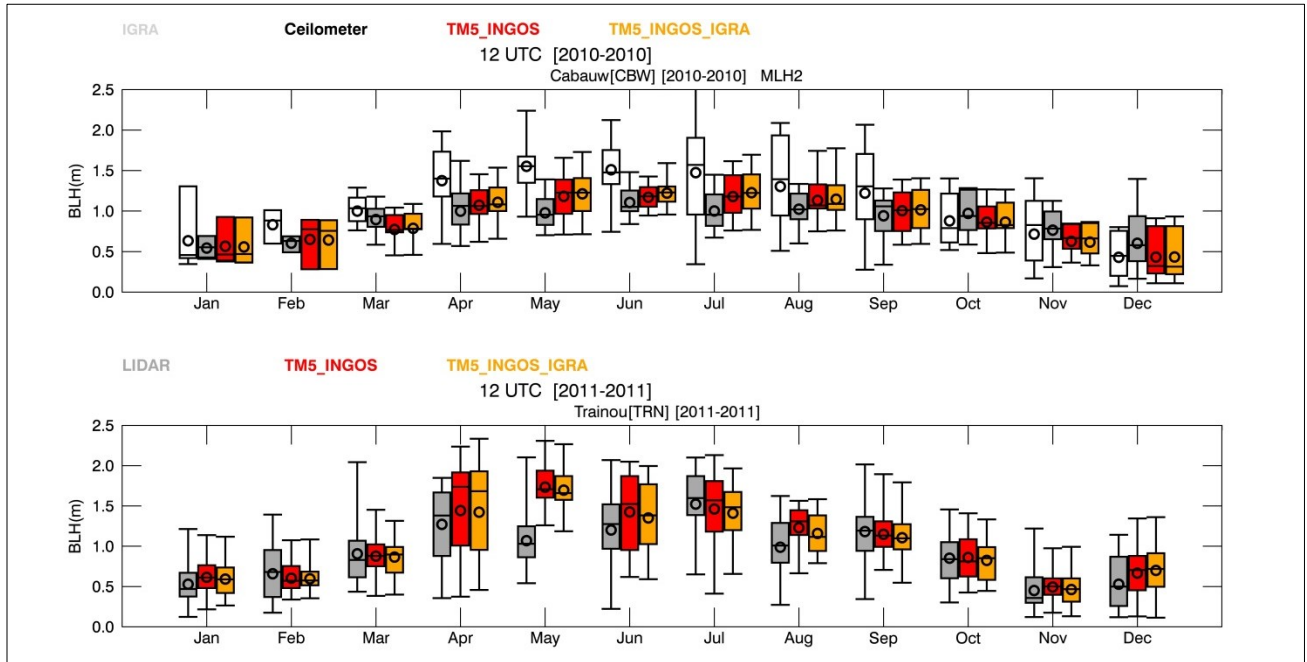
Figure 5: As Figure 3, but on the top) Cabauw (CBW) where both ceilometer and nearby IGRA observations (from De Bilt) are available. Observed (IGRA in blank; ceilometer in dark grey) and simulated (colors) boundary layer heights at 12 UTC and for 2010 are shown. On the bottom, Trainou (TRN) lidar based boundary layer heights (dark grey) at 12 UTC during 2011 are shown. The model boundary layer heights are represented by the colored boxes. The different acronyms are defined in Section 3.2 of the text
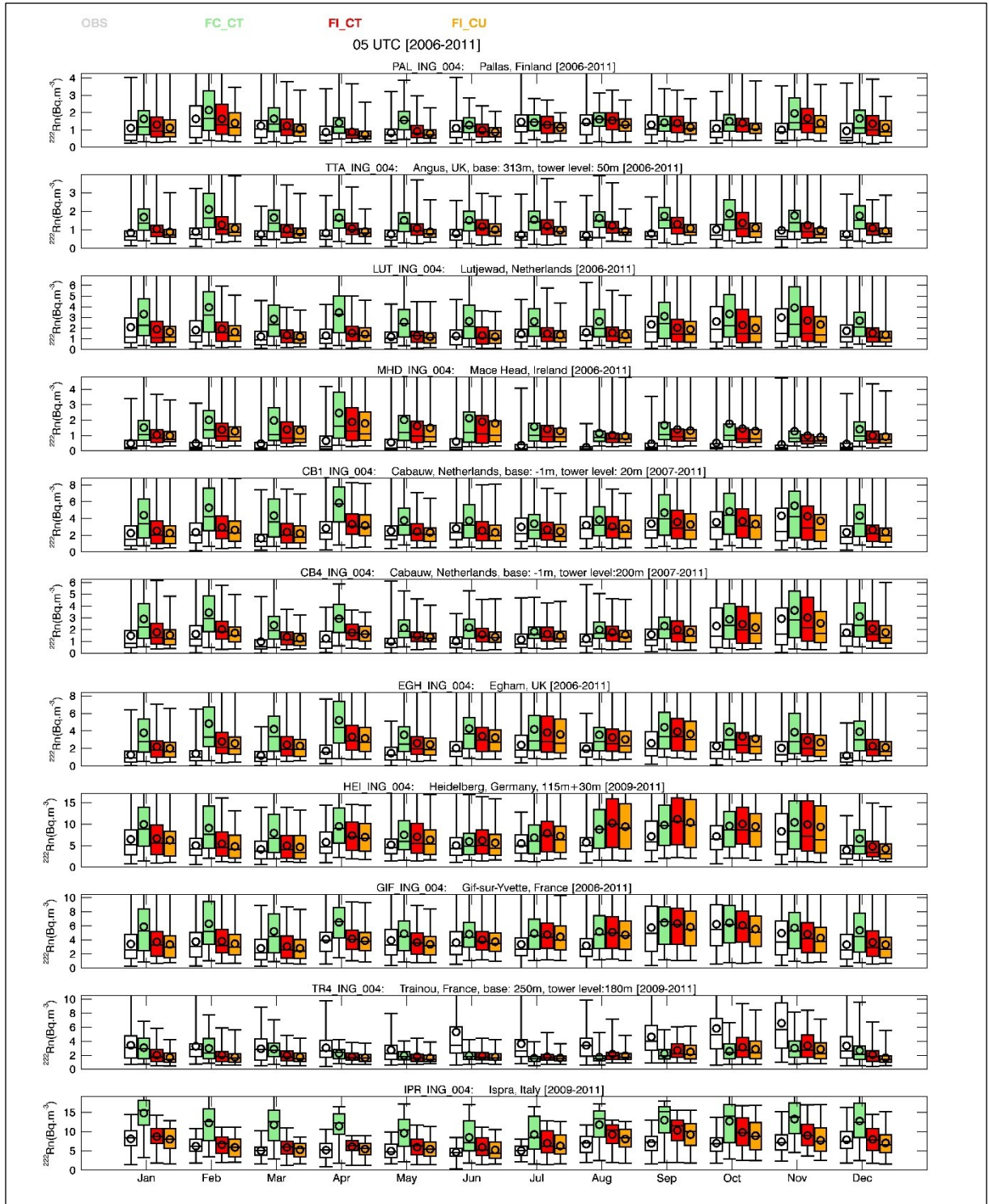
Figure 6: Seasonal variations of daily maximum of observed and simulated radon (^{222}Rn) activity concentrations at InGOS sites at 05 UTC (2006-2011). The Whisker plots show the monthly minimum and maximum values (bars), and the 25% and 75% percentiles (boxes). The median values are given by the horizontal line and the mean values by the open circles in the boxes. The observed radon activity concentrations are shown in blank, and the model simulations are represented by the colored boxes (the different acronyms are defined in Section 3.4). FC uses constant ^{222}Rn fluxes and FI the InGOS flux map
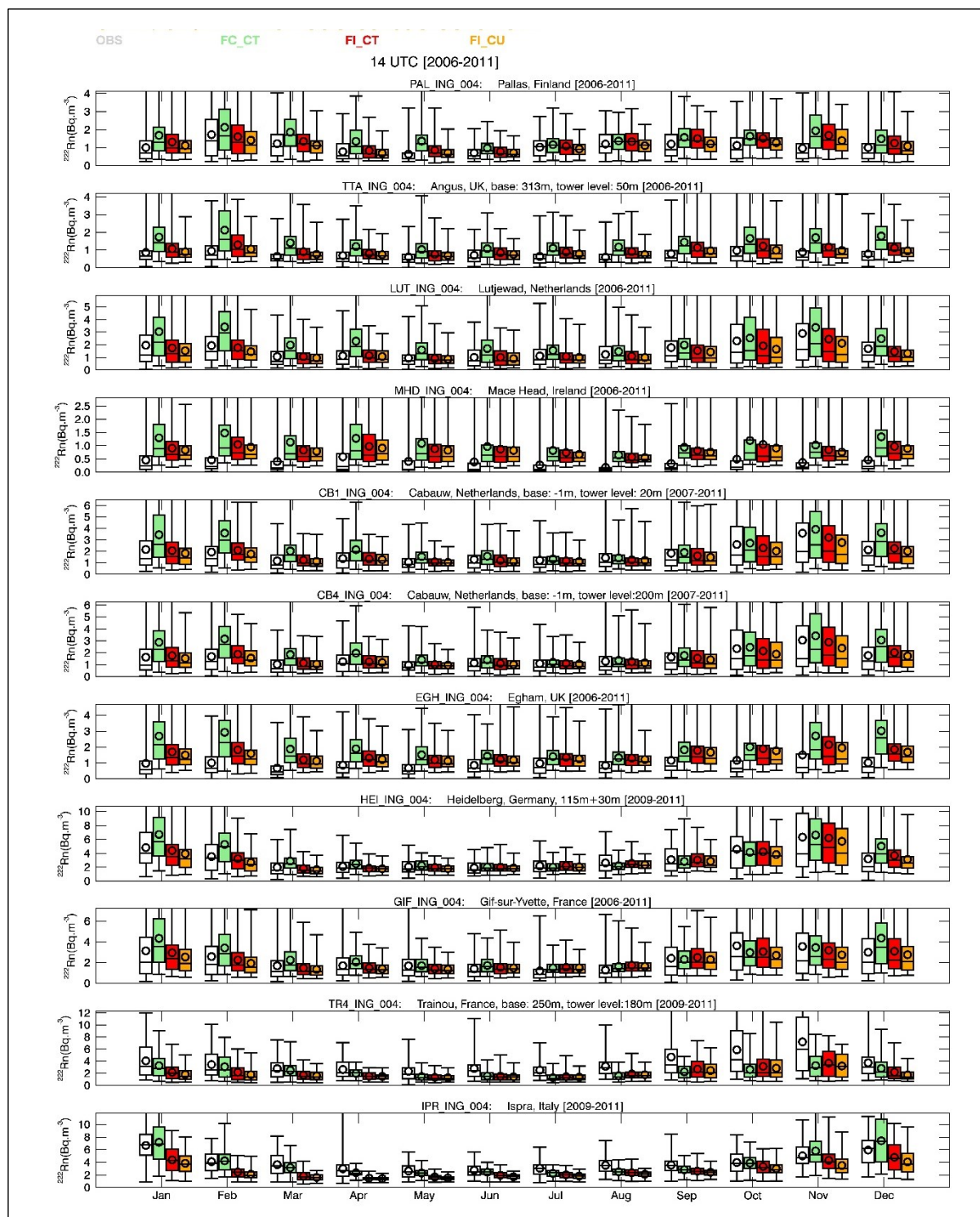
Figure 7: As for Figure 6, but at 14 UTC illustrating the seasonal variations of daily minimum of radon (^{222}Rn) activity concentrations
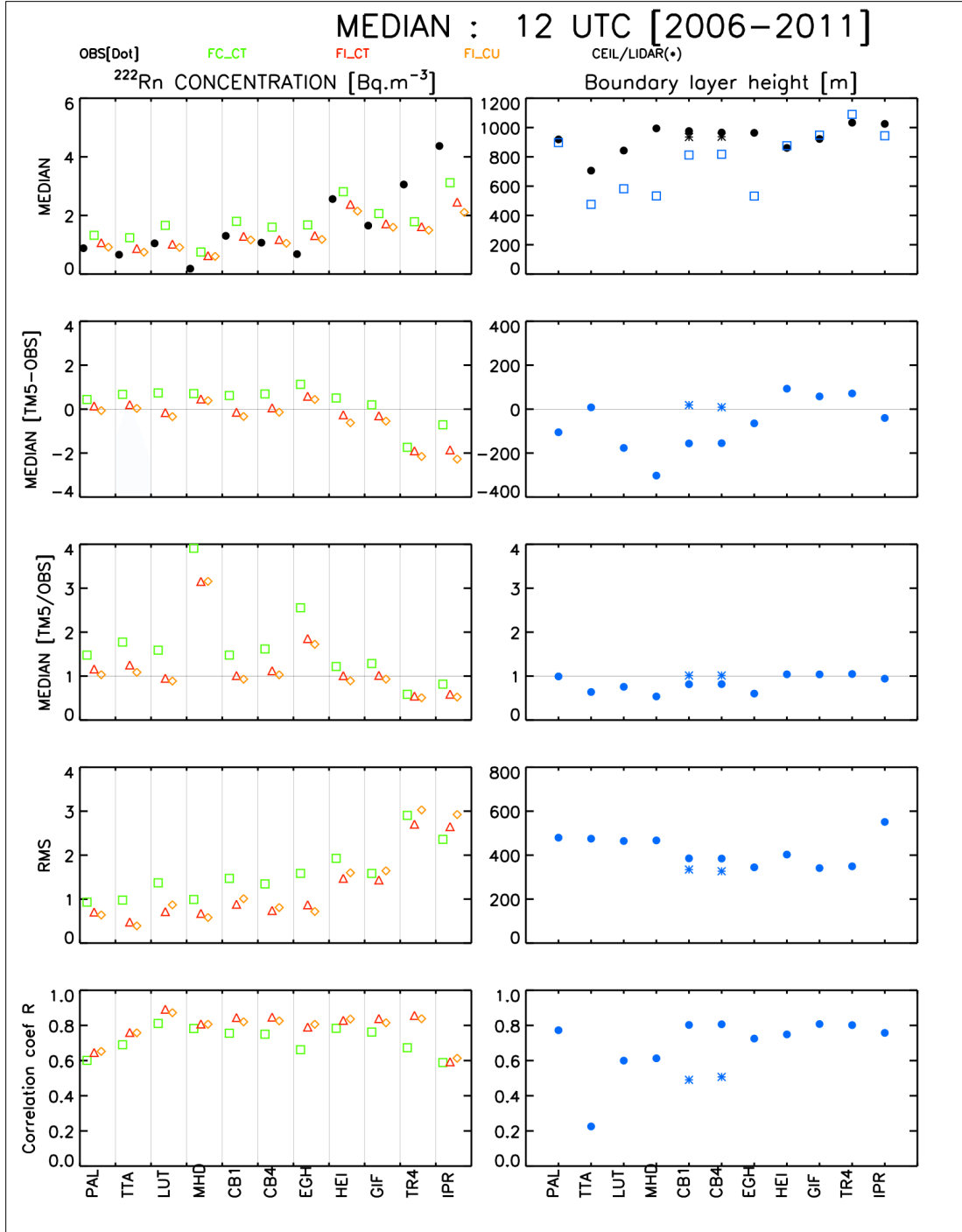
Figure 8: Left: statistics of observed vs. simulated ^{222}Rn activity concentrations for the different stations (12 UTC). Right: statistics of observed (IGRA (●) and ceilometer (CEIL)/LIDAR (*)) vs. simulated boundary layer heights (TM5_INGOS_IGRA). (12UTC). The acronyms of the stations (x-axis) are given in Table 1. For the median and RMS values, the unit of the y-axis is given on the top of the relevant graphs. The different model settings are given on the top of the graphs. The number of pair of data for each station is larger than 500
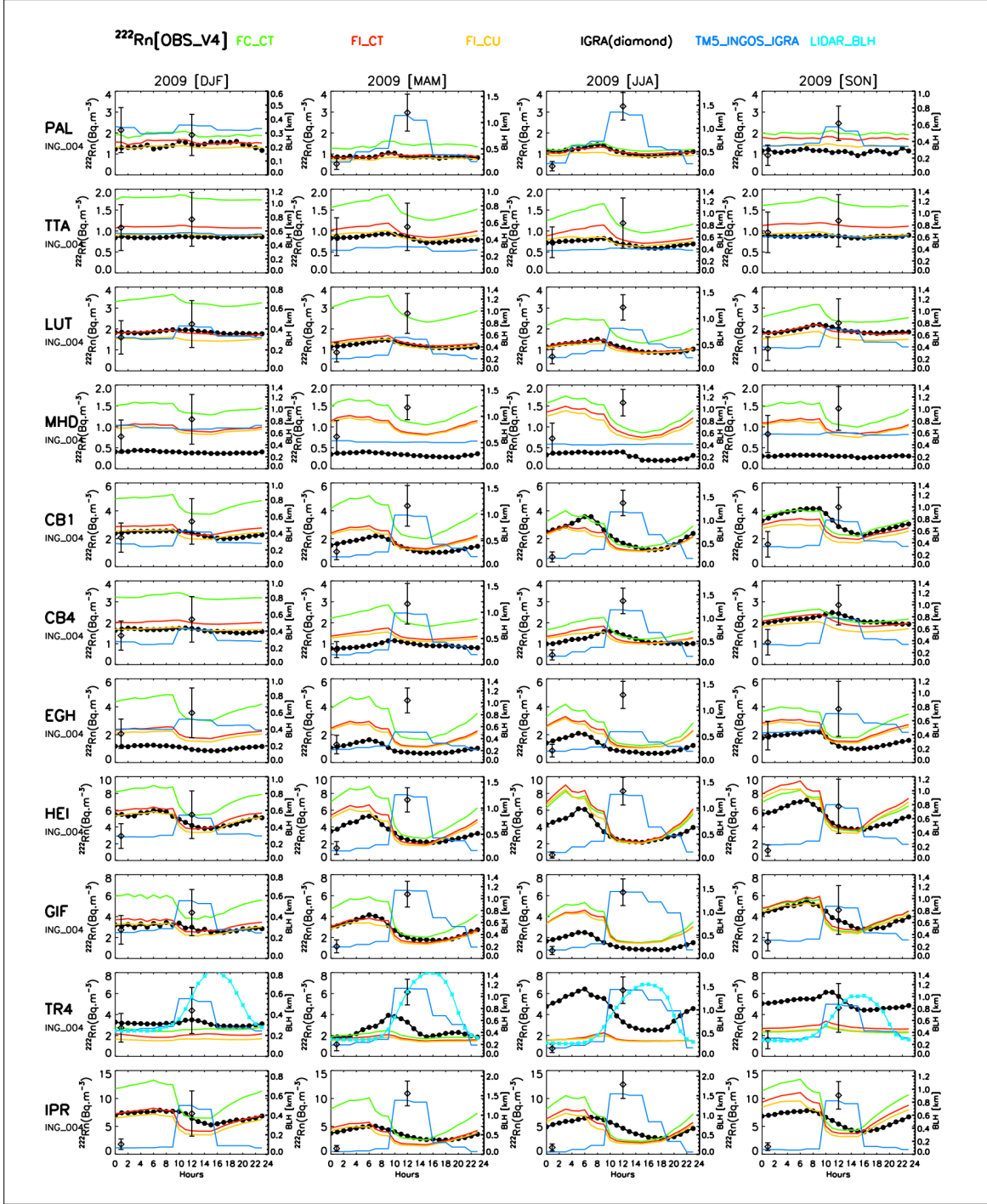
Figure 9: Seasonal variations of ^{222}Rn activity concentrations and boundary layer heights (BLH) at the InGOS stations that measure ^{222}Rn activity concentrations. The observed concentrations are represented by the black solid line with dots. Three model simulations are considered: FC_CT, the model simulations using constant emissions, FI_CT using the InGOS emissions and the default convection scheme of TM5; FI_CU using the InGOS emissions and the combination of the "revised slopes scheme" and the new convection scheme based on ECMWF reanalyses. The BLHs of TM5 (TM5_INGOS_IGRA) are in dark blue, while observed IGRA BLHs at 00 and 12 UTC are shown by the black diamonds together with their uncertainties. The lidar BLHs at Trainou (for 2011) are shown by the light blue line
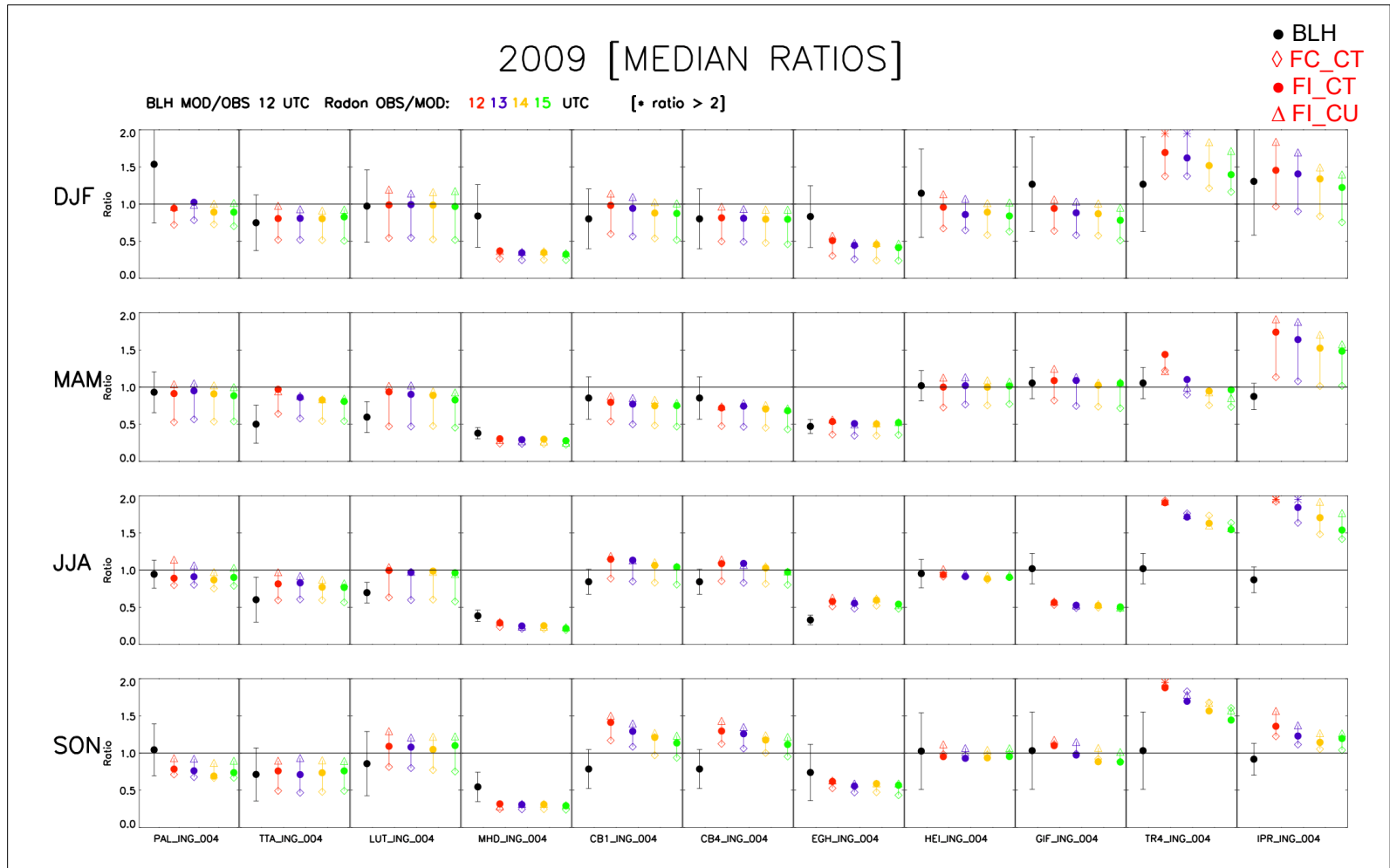
Figure 10: The seasonal variations of the ratios of BLHs (TM5/IGRA; black dot with solid line) at 12 UTC and the ratios of ^{222}Rn activity concentrations (OBS/TM5) at 12, 13, 14, and 15 UTC for the 4 seasons [DJF; MAM; JJA; SON] of the year 2009 for all InGOS ^{222}Rn measurement sites. The closest IGRA station to the radon measurement site is considered (see Figure 1). Three TM5 simulations are shown here: The model simulations using the constant emissions (FC_CT; colored diamond), InGOS emissions and using the default convection scheme of TM5 (FI_CT; colored filled circles) and using the new convection scheme (FI_U; colored triangles)
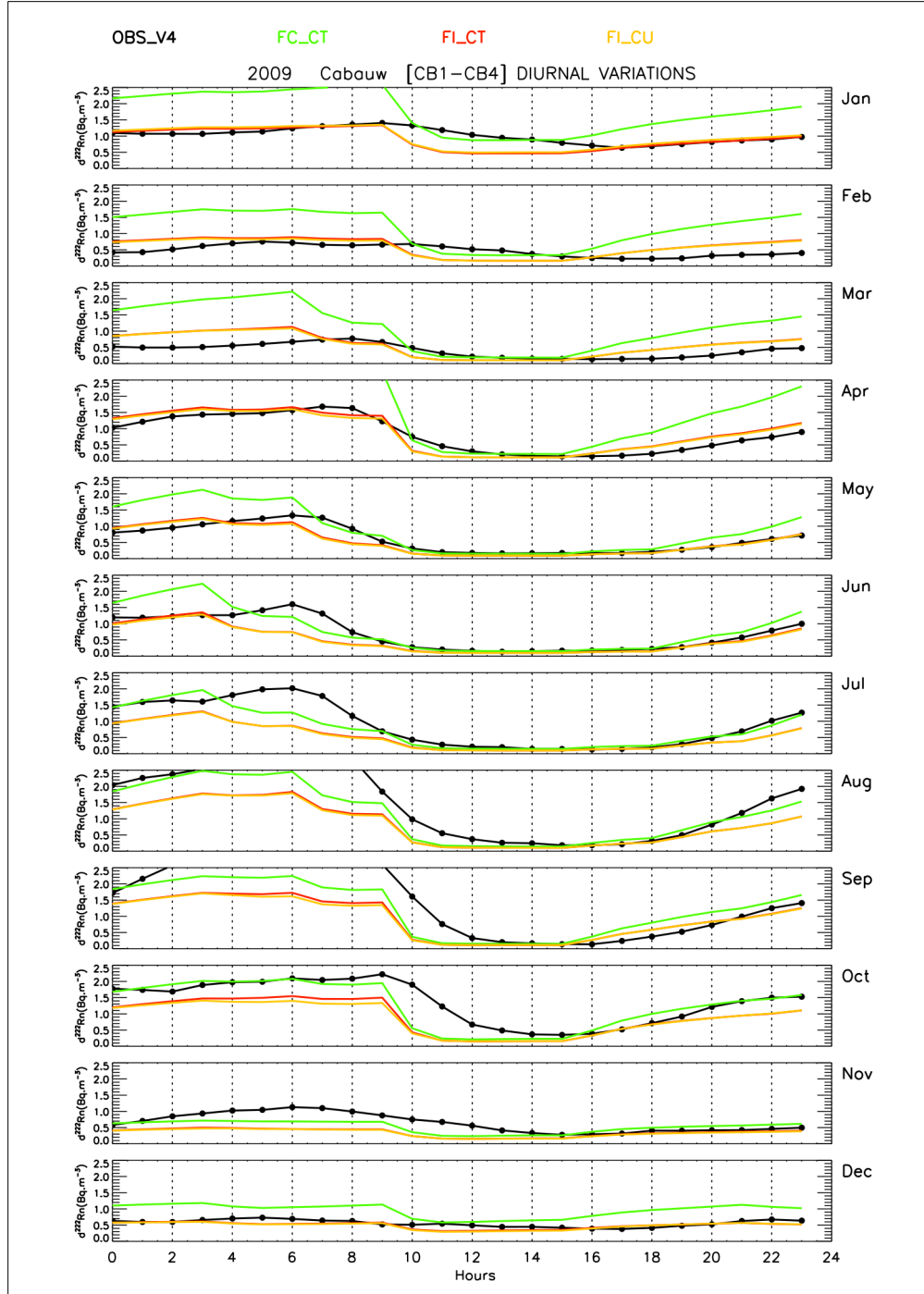
Figure 11: Mean diurnal variations of the radon activity concentration differences between the two measurement levels at Cabauw (20 m [CB1], 200 m [CB4]). The observed gradient is shown by the black solid line with dots (for each month of the year 2009), and the modelled gradient by the solid green line for the constant emissions (FC_CT), by the solid red line for the InGOS emissions (FI_CT), and for the solid orange line for the simulations using the InGOS emissions and the combination of the "revised slopes scheme" and the new convection scheme based on ECMWF reanalyses (FI_CU), respectively.
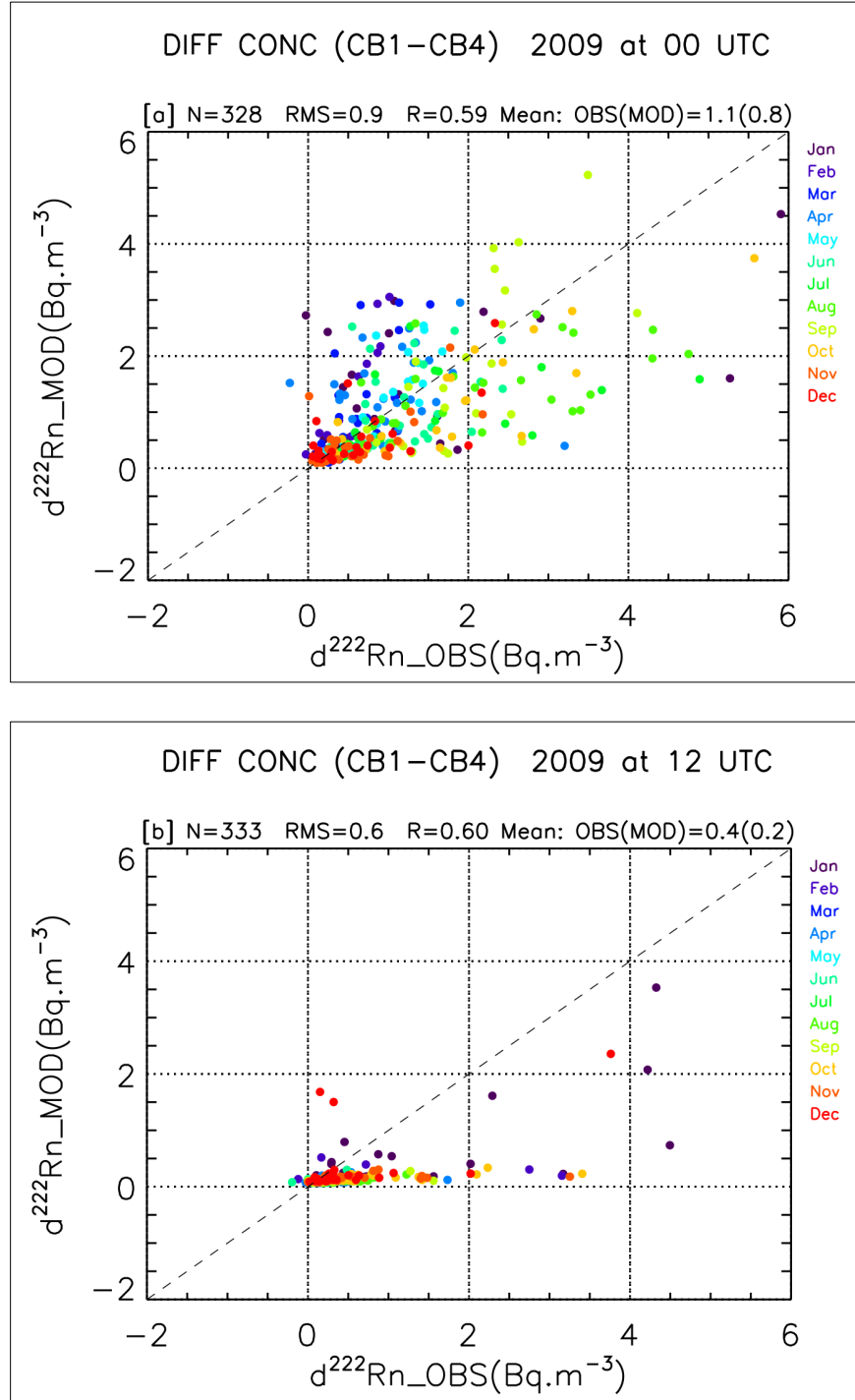
Figure 12: Correlation plots between the simulated and observed vertical [222]Rn activity concentration gradients (difference between 200m (CB4) and 20m (CB1) at Cabauw at 00 UTC (top) and 12 UTC (bottom). Model simulations using InGOS emissions (FI_CT termed as MOD) are shown. Each color indicates the month at which the data are obtained.