Remote Sensing of Environment 169 (2015) 418–437

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing

Jean-Luc Widlowski [a], Corrado Mio [a], Mathias Disney [b,c], Jennifer Adams [a,b], Ioannis Andredakis [d], Clement Atzberger [e], James Brennan [b,c], Lorenzo Busetto [f], Michaël Chelle [g], Guido Ceccherini [h], Roberto Colombo [i], Jean-Francois Côté [j], Alo Eenmäe [k,l], Richard Essery [m], Jean-Philippe Gastellu-Etchegorry [n], Nadine Gobron [a,*], Eloi Grau [n], Vanessa Haverd [o], Lucie Homolová [p], Huaguo Huang [q], Linda Hunt [r], Hideki Kobayashi [s], Benjamin Koetz [t], Andres Kuusk [l], Joel Kuusk [l], Mait Lang [k,l], Philip E. Lewis [b,c], Jennifer L. Lovell [u], Zbyněk Malenovský [v], Michele Meroni [w], Felix Morsdorf [x], Matti Mõttus [y], Wenge Ni-Meister [z], Bernard Pinty [a], Miina Rautiainen [aa], Martin Schlerf [ab], Ben Somers [ac], Jan Stuckens [ad], Michel M. Verstraete [ae,af], Wenze Yang [ag], Feng Zhao [ah], Terenzio Zenone [ai]

[a] Institute for Environment and Sustainability, European Commission, Joint Research Centre (JRC), Land Resources Management Unit, Via E. Fermi 2749, 21027 Ispra, VA, Italy
[b] Department of Geography, University College London, UK
[c] NERC National Centre for Earth Observation (NCEO), UK
[d] Institute for the Protection and Security of the Citizen, European Commission, Joint Research Centre (JRC), Global Security and Crisis Management Unit, Via E. Fermi 2749, 21027 Ispra, VA, Italy
[e] Institute of Surveying, Remote Sensing and Land Information, University of Natural Resources and Life Sciences, Vienna, Austria
[f] Institute on Remote Sensing of Environment, National Research Council (CNR-IREA), Via Bassini, 15, Milano 20133, Italy
[g] Institut National de la Recherche Agronomique, UMR1402 Ecosys, 1 route de la Ferme, F-78850 Thiverval-Grignon, France
[h] Institute for Environment and Sustainability, European Commission, Joint Research Centre (JRC), Water Resources Unit, Via E. Fermi 2749, 21027 Ispra, VA, Italy
[i] Remote Sensing of Environmental Dynamics Laboratory, Department of Earth and Environmental Science (DISAT), University of Milano-Bicocca, Piazza della Scienza 1, 20126 Milano, Italy
[j] Natural Resources Canada, Canadian Wood Fibre Centre, 1055 du P.E.P.S., P.O. Box 10380 Stn. Sainte-Foy, Quebec, QC G1V 4C7, Canada
[k] Estonian University of Life Sciences, Institute of Forestry and Rural Engineering, Kreutzwaldi 5, 51014 Tartu, Estonia
[l] Tartu Observatory 61602 Tõravere, Tartumaa, Estonia
[m] School of Geosciences, University of Edinburgh, UK
[n] Centre d'Etudes Spatiales de la BIOsphère, Toulouse, France
[o] CSIRO Oceans and Atmosphere Flagship, GPO Box 3023, Canberra, ACT 2601, Australia
[p] Global Change Research Centre, Academy of Sciences of the Czech Republic, v. v. i., Bělidla 986/4a, 603 00 Brno, Czech Republic
[q] Beijing Forestry University, Beijing, China
[r] Science Systems and Applications, Inc., Hampton VA 23666, USA
[s] Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[t] European Space Agency, ESRIN D/EOP-SEP, Via Galileo Galilei, I-00044 Frascati, (RM), Italy
[u] CSIRO Oceans and Atmosphere Flagship, GPO Box 1538, Hobart, TAS 7001, Australia
[v] Institute for Conservation Biology, School of Biological Sciences, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia
[w] Institute for Environment and Sustainability, European Commission, Joint Research Centre (JRC), Monitoring Agricultural Resources Unit Via E. Fermi 2749, 21027 Ispra, VA, Italy
[x] Remote Sensing Laboratories, Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[y] Department of Geosciences and Geography, University of Helsinki, P.O. Box 68, FI-00014, Finland
[z] Department of Geography, Hunter College, 695 Park Ave., New York, NY 10021, USA
[aa] Department of Real Estate, Planning and Geoinformatics and Department of Radioscience and Engineering, PO BOX 15800; 00076 Aalto University, Finland
[ab] Department for Environmental Research and Innovation, Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg
[ac] Division Forest, Nature and Landscape, KU Leuven, Heverlee, Belgium
[ad] Merkator nv, Vliegwezenlaan 48, 1731 Zellik, Belgium
[ae] South Africa National Space Agency, SANSA Earth Observation, Enterprise Building, Mark Shuttleworth Street, Innovation Hub, Pretoria 0087, South Africa
[af] Global Change and Sustainability Research Institute (GCSRI), University of the Witwatersrand, Johannesburg, South Africa
[ag] Cooperative Institute for Climate and Satellites, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740-3823, USA
[ah] School of Instrument Science and Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, PR China
[ai] Department of Biology, Research Group of Plant and Vegetation Ecology, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium

* Corresponding author.

http://dx.doi.org/10.1016/j.rse.2015.08.016

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form 10 July 2015
Accepted 13 August 2015

Keywords:
Conformity testing
Radiative transfer
Model benchmarking
3D virtual plant canopy
Digital hemispherical photography
Optical remote sensing
Shared risk
Guarded acceptance
GCOS
ISO-13528

ABSTRACT

The RAdiative transfer Model Intercomparison (RAMI) activity focuses on the benchmarking of canopy radiative transfer (RT) models. For the current fourth phase of RAMI, six highly realistic virtual plant environments were constructed on the basis of intensive field data collected from (both deciduous and coniferous) forest stands as well as test sites in Europe and South Africa. Twelve RT modelling groups provided simulations of canopy scale (directional and hemispherically integrated) radiative quantities, as well as a series of binary hemispherical photographs acquired from different locations within the virtual canopies. The simulation results showed much greater variance than those recently analysed for the abstract canopy scenarios of RAMI-IV. Canopy complexity is among the most likely drivers behind operator induced errors that gave rise to the discrepancies. Conformity testing was introduced to separate the simulation results into acceptable and non-acceptable contributions. More specifically, a shared risk approach is used to evaluate the compliance of RT model simulations on the basis of reference data generated with the weighted ensemble averaging technique from ISO-13528. However, using concepts from legal metrology, the uncertainty of this reference solution will be shown to prevent a confident assessment of model performance with respect to the selected tolerance intervals. As an alternative, guarded risk decision rules will be presented to account explicitly for the uncertainty associated with the reference and candidate methods. Both guarded acceptance and guarded rejection approaches are used to make confident statements about the acceptance and/or rejection of RT model simulations with respect to the predefined tolerance intervals.

1. Introduction

The RAdiative transfer Model Intercomparison exercise (RAMI) was launched in 1999 as an open, self-organizing activity of the canopy radiative transfer (RT) modelling community (Pinty, Gobron, Widlowski, Gerstl, Verstraete, Antunes, et al., 2001). For over 15 years now, RAMI has served as a common platform for the evaluation of models simulating bidirectional reflectance factors (BRFs) as well as radiative fluxes in 1-D and 3-D vegetation canopies. Starting with relatively simple and intentionally abstracted plant environments, the first three phases of RAMI were used by participants to identify coding bugs and to improve some of the RT formulations in their models (Pinty et al., 2001; Pinty, Gobron, Widlowski, Lavergne, & Verstraete, 2004; Widlowski, Taberner et al., 2007). By giving participants the opportunity to run the same experiments in subsequent phases of RAMI (together with one or two new ones), allowed to gradually eliminate operator induced errors in the setup and execution of the RT simulations. As a consequence, the overall model agreement increased and – by the end of the third phase of RAMI – was found to lie below 1% for the participating 3-D Monte Carlo models (Widlowski, Taberner, Pinty, Bruniquel-Pinel, Disney, Fernandes, (2007). This in turn enabled the definition of a surrogate truth data set and the subsequent development of a web-based benchmarking facility known as the RAMI On-line Model Checker (ROMC), accessible via http://romc.jrc.ec.europa.eu/ (Widlowski, Robustelli, Disney, Gastellu-Etchegorry, Lavergne, & Lewis (2007).

Building on these achievements, the fourth phase of RAMI was planned along two major axes, namely 1) the application of existing standards and procedures to enhance the rigour of the evaluation protocols, and 2) the expansion of test case complexity toward more realistic representations of plant environments. While the former featured prominently in the recent analysis of the abstract canopy scenarios of RAMI-IV (Widlowski, Pinty, Lopatka, Atzberger, Buzica, & Chelleet, 2013), the latter will be the focus of this present contribution. Realistic-looking canopy representations are relevant in a variety of contexts, among others, 1) to visually convince users of RT models about the relevance of these simulation tools – in particular if the use of effective variables in order to maintain radiative equivalence between abstracted and actual canopy representations is less familiar, (e.g., Pinty et al. (2004); Mõttus, Sulev & Lang, 2006); 2) to act as architectural reference in efforts aiming to assess the degree of canopy abstraction that still permits the simulation of RT quantities in different spectral bands and across different spatial scales to within predefined levels of tolerance from those of the true canopy, (e.g., Widlowski et al., 2013); and 3) to mimic in situ measurement protocols in an effort to a) characterise their inherent uncertainties with respect to their target variables at coarser spatial scales, e.g. Jonckheere, Nackaerts, Muys, van Aardt, and Coppin (2006), b) optimize sampling schemes and measurement devices for specific biomes, test sites, and/or illumination conditions, and c) support validation efforts of satellite-derived essential climate variables as suggested by the Global Climate Observing System, e.g., Bojinski, Verstraete, Peterson, Richter, Simmons, and Zemp (2014).

Efforts to assess the quality of remotely sensed data (whether real or simulated) inevitably require access to a reliable reference. However, this in itself is not sufficient to determine whether a candidate method or dataset is actually good enough for a particular task or not. To make that decision also requires information on how close an item must match the reference, or better still, how often the candidate method must fall within a predefined tolerance interval from the reference in order to be declared acceptable. This type of quality evaluation is deeply rooted in the manufacturing sector and also in legal metrology where it is referred to as conformity testing (e.g., JCGM-106, 2012). Over the years these communities have developed methods and rules (of thumb) allowing to explicitly account for or to limit the influence of uncertainty in either the reference and candidate methods or datasets. This body of knowledge is also applicable to a remote sensing context, where the idea of compliance testing is gradually gaining in prominence due to the regular publication of accuracy criteria for quantitative remote sensing products of relevance in a climate change context (e.g., GCOS, 2011).

This contribution will apply conformity testing to RT simulation results pertaining to the actual canopy scenarios of RAMI-IV. More specifically, Section 2 describes the reconstruction of three highly detailed virtual (deciduous and coniferous) forest stands, as well as two agricultural sites on the basis of data from intensive field measurement campaigns carried out in Europe and South Africa. In Section 3, the results of domain-level BRF and flux simulation from a dozen RT models will be presented for both summer and winter (snow) conditions and a series of spectral regimes. Conformity testing is then introduced in Section 4 and both the shared risk and the guarded risk approaches will be applied to the simulated BRF and flux quantities. Finally in Section 5, the main findings of this study are discussed and placed in context.

2. RAMI-IV actual canopy scenarios

In an effort to increase the realism of test scenes used in 3D radiative transfer simulations, RAMI-IV proposed a small number of actual

canopy scenarios. These highly detailed canopy representations were based on available inventory data from a series of test sites described in Table 1. In total, six different canopy scenarios covering deciduous and coniferous environments, regular and clumped tree patterns, as well as summer and winter conditions, were presented. For each canopy scenario, the RAMI website (http://rami-benchmark.jrc.ec.europa.eu/) provided the necessary structural and spectral information to assign a colour, shape, location and orientation to every individual leaf, twig, needle and branch within a nominally 1 ha sized area.

While this wealth of detail was not necessarily usable by all of the participating 3D models, it was a direct corollary of the RT community's request to increase the realism of the virtual plant environments used in RAMI. In addition, it was hoped that the actual canopy scenarios might permit documenting the impact that different levels of canopy abstraction may have on the simulated radiative properties of a canopy, e.g., Widlowski et al. (2013), as well as on the response of virtual sensors placed within these modelled plant environments.

It is imperative to note that the RAMI-IV actual canopy scenes are not identical to the real world environments that gave rise to the data on which the reconstructions are based. This is because even the most intensive field campaigns are not yet able to gather spatially-resolved information on the spectro-directional and architectural characteristics of all objects within (and adjacent to) a particular test site, e.g., Kuusk, Kuusk, and Lang (2009); Schneider, Leiterer, Morsdorf, Gastellu-Etchegorry, Lauret, Pfeifer et al. (2014). Information gaps encountered in the reconstruction of the RAMI-IV actual canopy environments concerned, for example, the shape, curl and size of individual foliage elements, the shape of individual tree crowns, the vertical distribution of foliage in a tree, the branching angles and patterns of individual trees, the orientation distribution of the foliage elements in a tree crown, the Bi-directional Reflectance Distribution Function (BRDF) of individual foliage elements, the BRDF of the background, the spatial variability of the background BRDF, the transmission of the foliage elements, the BRDF of tree branches, twigs and trunks, as well as the directionality of incident radiation. These information gaps consequently had to be replaced where necessary by information obtained by other means. As such the reconstructed scenes are not identical copies of their real world originals.

2.1. Architectural reconstructions

The virtual trees of the RAMI test cases Järvselja Pine Stand (HET07), Summer Järvselja Birch Stand (HET09) and Winter Järvselja Birch Stand (HET15) were generated on the basis of site-specific allometric equations for tree height, crown radius and crown length (Kuusk, Nilson, Paas, Lang, & Kuusk, 2008; Kuusk, Lang, & Kuusk, 2013). However, the number of unique virtual tree reconstructions was much smaller than that counted at the actual test sites. More specifically, the available tree height range was discretized into a small number of height intervals, for which a single virtual tree – centred on the origin of a Cartesian coordinate system – was then constructed on the basis of the available average architectural information and the corresponding tree libraries of the *xfrog* software (Lintermann & Deussen, 1999). For example, in the case of the HET07 scene only 10 unique Scots Pine tree models were generated instead of the 1114 individual pine trees that were contained in the stand inventory. To build the RAMI scene, first each inventoried tree was mapped into its corresponding height class, then a random azimuthal rotation was applied to the model tree

Table 1
Overview of the test sites that served as blueprint for the generation of the RAMI-IV actual canopy scenes. Indicated are the locations of the test sites, their characteristics, as well as any relevant publications and the site identifier (ID) used in RAMI-IV. More detailed information are available from the RAMI web site, namely http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI-IV/RAMI-IV.php.

Name	Location	Description	Reference	RAMI-IV ID
Järvselja Birch Stand	Järvselja, Estonia, 58° 16′ 50″ N, 27° 19′ 52″ E	49 year old deciduous stand dominated by birch (57%), common alder (29.5%) and aspen (11%). A lower tree layer contains linden and spruce. The average height in the top layer is 25 m. The overall stand density is 1017 trees/ha.	Kuusk et al. (2009); Kuusk, Nilson, Kuusk, and Lang (2010), (2013) Pisek, Lang, Nilson, Korhonen, and Karu (2011)	HET09_JBS_SUM for the summer conditions and HET15_JBS_WIN for the scene with winter conditions
Järvselja Pine Stand	Järvselja, Estonia, 58° 18′ 41″ N, 27° 17′ 48″ E	124 year old Scots pine stand (99%) with a stand height of 15.6 m. The stand density is 1122 trees/ha.	Kuusk et al. (2008, 2010, 2013) Pisek et al. (2011)	HET07_JPS_SUM
Ofenpass Pine Stand	Ofenpass, Swiss National Park, 46°39′ 45″ N 10°13′ 46″ E	150 year old coniferous stand in a dry alpine valley dominated by Mountain Pine. 20% of the stand consists of upright dead trees. The understorey is low and dense vegetation. The stand height is 15 m.	Morsdorf et al. (2004, 2006); Kötz, Schaepman, Morsdorf, Bowyer, Itten, and Allgöwer (2004)	HET08_OPS_WIN for the scene with winter conditions
Wellington Citrus Orchard	Near Wellington, South Africa 33° 35′ 00″ S 18° 55′ 30″ E	9-year-old orange grove with a row spacing of 4.5 m, a tree spacing of 2 m, and a row azimuth of 7.31°. The average tree height was 3 m.	Somers, Delalieux, Verstraeten, and Coppin (2009), Stuckens, Somers, Delalieux, Verstraeten, and Coppin (2009)	HET14_WCO
Short Rotation Forest	Near Vigevano, Lombardy, Italy, 45°17′ 42″ N 8° 52′ 38″ E	1 year old poplar plantation in the basin of the Ticino river consisting of Pegaso clones planted in double rows (0.45 × 0.74 × 2.8 m) for a density of 12,500 plants/ha.	Zenone, Migliavacca, Montagnani, Seufert, and Valentini (2008)	HET16_SRF

corresponding to the selected height class, and finally the resulting tree instance was translated to its inventoried location within a 1 ha area. For the HET07 scene, a further 6 birch trees were added in a similar manner to make the final tree count equal to 1120.

While the tree number of the thus constructed HET07 scene was in good agreement with that reported by Kuusk et al. (2008), the tree density of the virtual pine stand scene was however somewhat smaller. This is because field inventories use the location of tree trunks when reporting on the number of trees within a given area, while the description of the virtual RAMI-IV stands necessitates a background area that contains not only the trunks but also the crowns of all the trees in a scene. As a result the actual size of the HET07 stand increased to 1.12 ha, and the effective tree density reduced to 996 tree by hectare. In the case of the Järvselja Birch Stand (HET09) a total of 1029 tree (trunk) positions were located within a 1 ha, while their crowns covered 1.12 ha thus reducing the effective tree density to 919 tree by hectare. Despite these apparent differences between actual field inventory and RAMI-IV tree densities, the tree count statistics were maintained within the central 1 ha of the virtual canopy scenes. This applies thus also to any other statistics that require normalisation by the stand area, e.g., the leaf area index.

The virtual trees of the Ofenpass Pine Stand (HET08) scene were generated with the *xfrog* software (Lintermann & Deussen, 1999) using tree crown information extracted from LiDAR observations over the test site, e.g., Morsdorf, Kötz, Meier, Itten, and Allgöwer (2006); Morsdorf, Meier, Kötz, Itten, Dobbertin and Allgöwer (2004). Similar to the Järvselja scenes, the available tree height range of the several hundred live Mountain Pine trees was subdivided into 12 height intervals for each of which a single model tree was then built. After a random azimuthal rotation, the modelled trees – having a maximum height of 15 m – were then placed in accordance with location coordinates derived from an airborne study (Morsdorf et al., 2006). In addition to these mature live trees, a total of 250 shrubs of about 1 m height were generated and randomly distributed across the scene in an attempt to mimic the substantial understorey of the Ofenpass pine stand. Finally, a total of 120 dead trees of circa 10 m height were also included to match the 20% of dead (standing) trees that made up the overall tree cover at the actual Ofenpass site. Last but not least, and also contrary to the actual site, the background of the HET09 scene in RAMI-IV was assumed to be completely flat.

Among the actual canopy scenes with regular planting patterns, the Wellington Citrus Orchard (HET14) scene was built with the virtual plant models previously created by Stuckens et al. (2008). Here, individual trees were constructed with the *arbaro* software (Weber & Penn, 1995) on the basis of field data collected by Somers et al. (2009) in a citrus grove near Wellington, South Africa and calibrated by comparing measured and simulated gap fraction distributions. The generated 10 unique models of the *Citrus sinensis* L. species were randomly rotated around their vertical axis before being positioned at circa 2 m intervals in rows separated by approximately 4.5 m in an effort to resemble the actual orchard spatial distribution. About 11% of the potential tree positions were omitted in order to mimic the percentage of missing trees in the real orchard. The rows were parallel to the Y-axis of the scene which corresponded to having an azimuth of 90° when counted counterclockwise from the X-axis. The background was assumed free of grass and completely flat.

The second actual canopy scene with a regular planting pattern was the Short Rotation Forest (HET16). It featured 11924 poplar clones, which occurred every 0.45 m along double rows separated by about 3.55 m from each other (Zenone et al., 2008). The rows in each double row were separated by roughly 0.75 m. Six unique representations of 1 year old Pegaso plants were generated with the *arbaro* software (Weber & Penn, 1995). These model plants were rotated in a random manner around their vertical and then distributed according to the double-row planting scheme described above. Overall, 2% of the potential tree positions were omitted to match the percentage of missing trees in the original test site. As with the other actual canopy scenes, the background was assumed completely flat and the rows were oriented parallel to the Y-axis of the scene which corresponded to having an azimuth of 90° when counted counterclockwise from the X-axis. The leaf area index of the scene was set such as to represent a fully developed 1 year old canopy.

For almost all of the above actual canopy scenes the geometry of the various foliage units were modelled on the basis of information obtainable from the internet. This meant that leaf curl was generally not accounted for with exception of the citrus tree (*C. sinensis* L.) where leaf curl information was available from Stuckens, Somers, Delalieux, Verstraeten, and Coppin (2009). In addition, the shoots of the Scots pine, Norway Spruce and Mountain Pine species were all reconstructed in a three-dimensional manner following the model of Stenberg, Palmroth, Bond, Sprugel, and Smolander (2001) and Smolander and Stenberg (2003). Fig. 1 provides a graphical overview of the structural elements that contribute (at different scales) to the architectural realism of the various actual canopy scenes in RAMI-IV. All of the structural and spectral properties to reconstruct these scenes are maintained on the RAMI website.

Table 2 lists the number of geometric primitives – such as triangles, cylinders, discs, ellipsoids – needed to describe each and every one of the structural attributes in the above scenes. One can see that the representation of a visually realistic one hectare forest stand required between 90 million (in the case of the citrus orchard) and almost 900 million (for the Järvselja pine stand) definitions of geometric primitives. Graphical depictions of the resulting architectural detail are shown in Fig. 2. While the RAMI website contained all the necessary information to address a comprehensive scene generation, it also provided architectural information of a more statistical nature that were derived from the detailed plant environments. This included, for example, information on the height, leaf and wood area of individual trees, the vertical profile of foliage and wood area in individual trees, the leaf orientation distributions within individual tree crowns, the number of dead branches below the crown of trees, as well as the height to the bottom of the tree crown and the diameter at breast height. In addition, it was explicitly mentioned on the RAMI website that RAMI participants were to treat the information presented in the same manner as they would deal with actual 'inventory data'. In other words, they should identify and extract those parameters and characteristics that were required by their RT models to mimic the radiative environments of the actual canopy scenes.

2.2. Spectral properties and illumination conditions

It is widely known that foliage reflects light anisotropically, e.g., Breece and Holmes (1971); Brakke, Smith, and Harnden (1989); Kaasalainen and Rautiainen (2007). In recent years this has prompted the development of dedicated leaf goniometers, e.g., Bousquet, Lachérade, Jacquemoud, and Moya (2005); Biliouris, Verstraeten, Dutré, van Aardt, Muys, and Coppin, (2007); Lolli, Pisani, Rajteri, Widlowski, Bialek, Greenwell et al. (2014). However, the acquisition of reliable bi-conical reflectance and transmission information in the field is far from trivial. As such, the vast majority of currently available leaf and wood reflectance data does not originate from leaf goniometric measurements but rather consists of either hemispherically integrated quantities or narrow field-of-view spectrometric data acquired under specific view and illumination geometries under ambient light conditions, e.g., Milton, Schaepman, Anderson, Kneubühler, and Fox (2007). When such reflectance data are used within RT modelling contexts, a Lambertian scattering law is typically assumed (in order to simulate directional scattering). Evidence exists that this assumption may bear little on the quality of canopy absorption estimates, e.g., Chelle (2006), however little is known as to how this assumption relates to other radiative quantities and the spatial scale of interest (i.e., local versus domain-level radiative properties).

As a consequence of missing spectro-directional inventory data, a gap-filling approach had to be used to assign the necessary spectro-directional reflectance (and transmittance) properties. More specifically, Lambertian

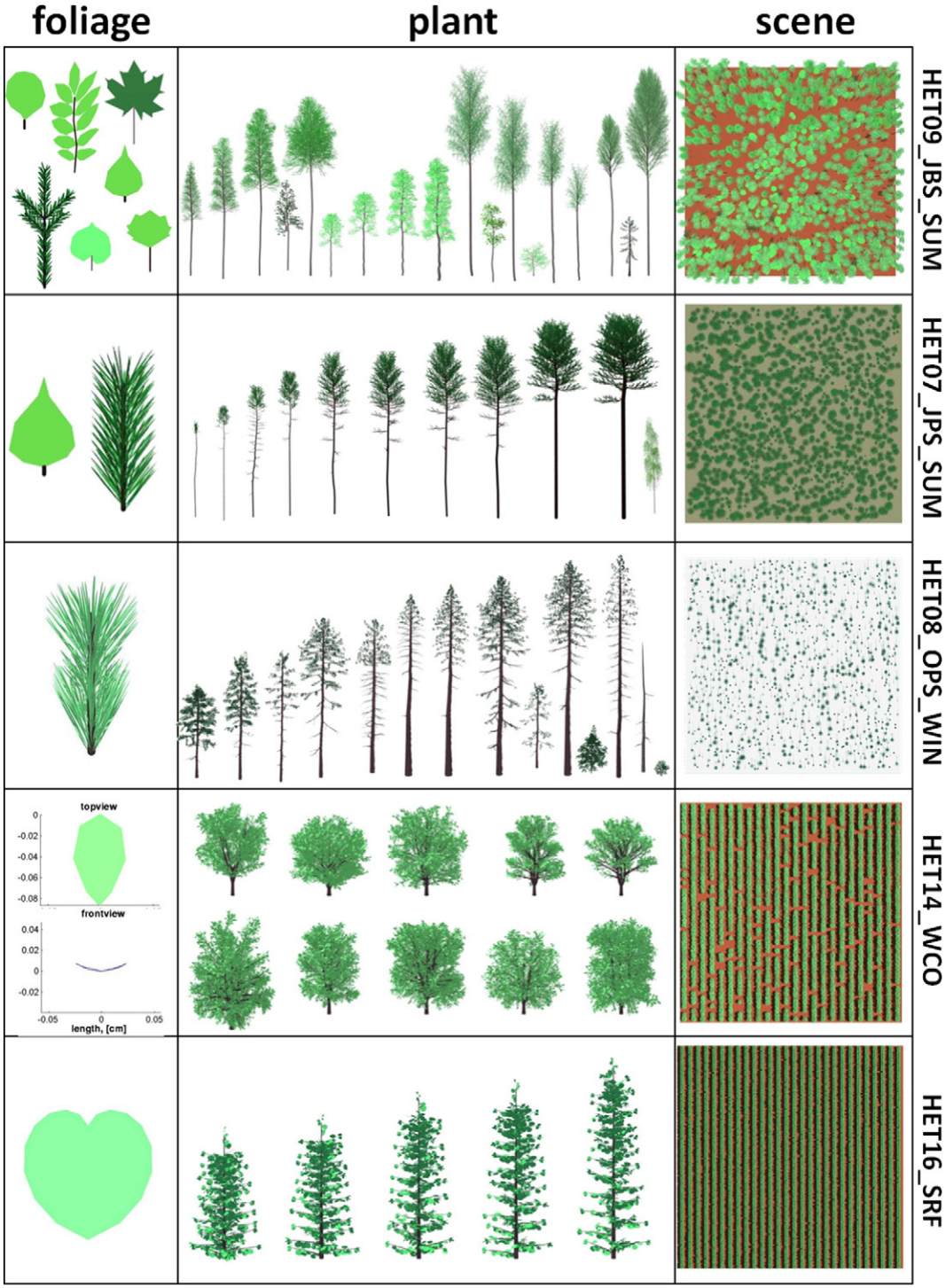

Fig. 1. Overview of the structural elements that contribute at different scales to the architectural realism of the actual canopy scenes in RAMI-IV. The winter birch stand (not shown) is structurally identical to the summer version bar the foliage which is removed from deciduous trees. Note that the graphics below are for illustration purposes only, i.e., the colours do not reflect the actual spectral properties of the depicted items.

scattering laws were assumed throughout all of the RAMI-IV actual canopy scenes, i.e., for all of the geometric primitives that made up the canopy. The only exceptions to this were related to the anisotropy of the backgrounds in the two winter scenes. Here, the BRDF of the snow-covered backgrounds were mimicked with the semi-empirical RPV model (Rahman, Pinty, & Verstraete, 1993; Rahman, Verstraete, & Pinty, 1993) using parameter values obtained from fitting the RPV model with the procedure of Lavergne, Kaminski, Pinty, Taberner, Gobron, Verstraete, et al. (2007) to a directional reflectance dataset acquired with a portable goniometer (Sandmeier, 2000). For the summer scenarios, the prescribed background albedos were based on in situ measurements at the respective test sites (and ignored any spatial variability). In addition, while actual reflectance measurements of species-specific wood and foliage spectra were available for most of the tree species, this was not always the case for transmission data. Whenever reflectance or transmission data were missing, spectra from similar species was used to fill in the gaps (again using the Lambertian scattering assumption).

For the various forest stands, only a single illumination scenario was defined (see rightmost column in Table 2). However, due to a lack of information about the directional distribution of the incident radiation

Table 2
Overview of architectural characteristics for the six *reconstructed* RAMI-IV actual canopy scenes together with the illumination conditions (SAA = sun azimuth angle, SZA = sun zenith angle) prescribed for the radiative transfer simulations. Note that the structural properties here may differ from those inventoried at the actual test sites. The density accounts for all plant objects (live trees, dead trees, understorey if defined) within one hectare. The leaf area index (LAI) is computed as half the total area of all virtual foliage elements within a scene divided by the background area of that scene. The fractional cover is identical to the direct transmission (in percentage) along the vertical through the canopy. The fractional cover is defined as 1 - direct transmission at zero solar zenith angle. The number of geometric primitives (triangles, ellipsoids, cylinders, discs) that described a given actual canopy scenario on the RAMI website are also indicated. More detailed information are available from the RAMI web site, namely http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI-IV/RAMI-IV.php.

Name	Density of plants [ha^{-1}]	Scene LAI [m^2/m^2]	Fractional cover [%]	Maximal height [m]	Primitives in scene [−]	SZA; SAA [°]
Järvselja Birch Stand (SUM & WIN)	919	3.442	50.4	30.51	350,050,476 and 138,251,607	36.6; 270.69 and 54.0; 291.3
Järvselja Pine Stand	996	2.302	40.6	18.56	895,635,743	36.6; 299.06
Ofenpass Pine Stand	931	0.745	12.5	15.02	169,120,314	47.0; 151.3
Wellington Citrus Orchard	991	2.691	39.2	4.12	89,618,249	(0,20,50); 0 and (20, 50); 90
Short Rotation Forest	11,841	3.219	39.2	3.41	146,665,201	(0,20,50); 0 and (20, 50); 90

(and thus across all relevant spectral bands in the visible and near-infrared), only the spectrally varying ratio of (isotropic) diffuse to total incident radiation was reported on the RAMI website. These data were based on measurements acquired either directly at the site (Kuusk et al., 2013) or using the nearest AERONET (Holben, Eck, Slutsker, Tanré, Buis, & Setzer, 1998) station to the test sites. For the actual canopy scenes with regular planting patterns, 5 different sun positions (parallel and perpendicular to the plant rows) were provided. For these scenarios the diffuse component was set to zero.

3. Generation of the simulation results

3.1. Participating models

The setup and coordination of RAMI-IV followed those of previous phases. Participation was on a voluntary basis and a feedback was provided to participants in case of obvious model blunders (regarding BRF simulations). This often resulted in a renewed submission of the simulation results. Overall, a total of 12 different RT models participated in the actual canopy scenarios of RAMI-IV. Table 3 lists RT model names, relevant publications, operator names and the type of measurements that were submitted. As in previous phases of RAMI, model participation was somewhat heterogeneous, in the sense that no single RT model performed all of the measurements and test cases. In addition, several models submitting flux quantities did not provide BRFs and vice versa. As a result, there were 6–7 RT models typically contributing to the BRF measurements whereas only 4–5 models contributed to hemispherical flux measurements on average (albeit some cases had up to 7 models contributing, while others received only one or even none at all). Not accounting for multiple submissions, the number of measurement files that were ultimately used in the analysis amounted to 700 for ACTS, 692 for CanSPART, 3957 for DART, 4380 for FLiES, 1149 for FRT, 76 for inform, 2334 for librat, 2660 for pbrt, 36 for price, 6198 for rayspread, 3728 for RGM and 1368 for row. In general, participation decreased when sub-components of BRFs or fluxes were to be simulated (i.e., the uncollided by vegetation, the single collided by vegetation, and the multiple-collided by the background and canopy components). As a consequence, this contribution will focus primarily on total BRF and flux quantities.

3.2. Simulation results

Widlowski et al. (2013) provides a detailed description of the various BRF and flux quantities prescribed in the context of RAMI-IV (see their Table 2 for a summary). As far as the actual canopy scenarios were concerned, participants had been encouraged to provide simulations

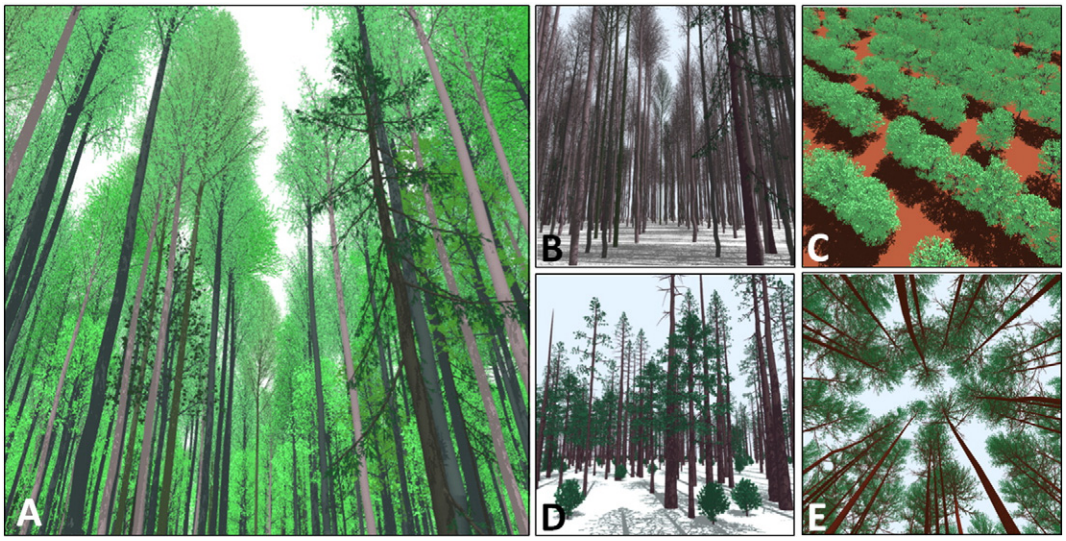

Fig. 2. Graphical depictions of the architectural realism contained in the actual canopy scenes of RAMI-IV. Panel A) depicts the summer version of the Järvselja birch stand, panel B) relates to the winter version of the Järvselja birch stand, C) displays the Wellington Citrus orchard, D) provides a view of the Ofenpass Mountain pine stand, and E) shows the Järvselja Pine stand. Note that the colour of the sky, the background as well as any other object in these images does not correspond to the spectral properties assigned for the radiative transfer simulations.

Table 3
Overview of RT models, relevant publications and operators that submitted simulation results for the RAMI-IV actual canopy scenarios. THP refers to thresholded hemispherical photographs. More detailed information are available from the RAMI web site, namely http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI-IV/RAMI-IV.php.

Model name	Model reference(s)	Model operator	Measurement types
ACTS	Ni-Meister, Yang, and Kian (2010) Yang, Ni-Meister, Kiang, Moorcraft, Strahler, and Oliphant, (2010)	W. Ni-Meister W. Yang	Fluxes
CanSPART	Haverd, Lovell, Cuntz, Jupp, Newnham, and Sea (2012) Lovell, Haverd, Jupp, and Newnham (2012)	V. Haverd J. Lovell	Fluxes
DART	Gastellu-Etchegorry, Demarez, Pinel, and Zagolski (1996); Gastellu-Etchegorry, Martin, and Gascon (2004); Gastellu-Etchegorry, Yin, Lauret, Cajgfinger, Gregoire and Lopes (2015)	E. Grau J-P Gastellu	BRFs, fluxes
FLiES	Kobayashi and Iwabuchi (2008)	H. Kobayashi	BRFs, fluxes, THP
FRT	Kuusk and Nilson (2000) Kuusk et al. (2008)	A. Kuusk	BRFs, fluxes
inform	Atzberger (2000) Schlerf and Atzberger (2006)	C. Atzberger M. Schlerf	BRFs
librat	Lewis (1999) Disney, Lewis, Bouvet, Prieto-Blanco, and Hancock (2009)	M. Disney P. Lewis	BRFs, THP
pbrt	Pharr and Humphreys (2010)	J. Stuckens	BRFs, fluxes
price	Essery, Bunting, Hardy, Link, Marks, and Melloh (2008)	R. Esserey	THP
RGM	Liu, Huang, Qin, Fu, and Li (2007) Huang, Chen, and Liu (2009)	H. Huang	BRFs, fluxes, THP
Rayspread raytran[1]	Widlowski, Lavergne, Pinty, Verstraete, and Gobron (2006) Govaerts (1995)	J-L Widlowski C. Mio	BRFs, fluxes, THP
row	Zhao, Gu, Verhoef, Wang, Yu, Liu, et al. (2010)	F. Zhao	BRFs

[1] In order to simplify notation, we always refer to rayspread model.

of these canopy-scale quantities in 19 different spectral bands corresponding to 1) the 18 spectral bands of the CHRIS-Proba instruments reported in Kuusk et al. (2009) and listed again in Table 4 here, and, 2) a so called purist scenario (Pinty et al., 2001) where soil and bark reflectance were set equal to 1 and leaf reflectance and transmittance were both fixed at 0.5 (here labelled as band 19: B19). In addition, participants had also been asked to provide 1) total BRF simulations at a fixed view zenith angle (i.e., 37°) covering the full range of possible view azimuth angles (at 2° interval), 2) the bi-hemispherical reflectance (white sky albedo), and 3) canopy transmission for diffuse incident radiation only. Furthermore, model simulations were encouraged that mimicked the result of an upward looking, perfectly thresholded, hemispherical photograph (THP) taken at 5 cm above the background level (for a total of 9 predefined locations – labelled A to I – within the various actual canopy scenes). Detailed definitions of these quantities as well as pointers to relevant scientific reference publications were available on the RAMI website.

3.3. Example of BRF simulations

Fig. 3 depicts RT model simulated canopy-level BRFs along the orthogonal plane (top row) as well as the principal plane (bottom row) for a variety of spectral bands and different actual canopy scenarios. These examples document the rather large dispersion that generally exists between the RT model simulations in RAMI-IV, as well as the recurring lack of a unique cluster of RT models exhibiting similar results. In a sense, this is not surprising since all of the actual canopy test cases are new and rather complex architecturally. Earlier phases of RAMI have shown that new experiments increase the likelihood of implementation glitches (Pinty et al., 2001) and that one or two reruns of the same scenarios are needed before this error component is satisfactorily reduced (Pinty et al., 2004; Widlowski et al., 2007). In fact, the simulation results presented here indicate that a similar approach might also have to be adopted here.

One can see in Fig. 3, that the angular pattern of the simulated BRFs is often rather similar between models. In general, there are no more than one or two models that differ in their simulated BRF shapes. Occasionally, the BRFs of the RGM model were susceptible to rapid fall-offs beyond view zenith angles of about 60°. Similarly, the BRF simulations of the pbrt model could appear somewhat noisy while the DART model occasionally showed a somewhat broader hot spot region, and the FLiES model simulations feature occasional kinks (especially along the principal plane). Somewhat different in Fig. 3 is the top middle panel, which depicts simulation results for the purist case where no absorption is permitted within the canopy (here the Wellington Citrus Orchard). As such, the albedo (which is equal to the hemispherically integrated BRF) must be equal to one for this scenario, which implies that any BRF values smaller than unity must be balanced by others that exceed this number. This does not always appear to be the case though and will be addressed in more detail in the model consistency section below.

Fig. 4 depicts the results of BRF simulations carried out in the red (left panels) and near-infrared (right panels) spectral domains at a constant view zenith angle of 37° and for azimuth angles ranging from 0° to 360°. The central column shows a graphical representation of the target canopy (the Wellington Citrus Orchard scene) when viewed from directly above together with a (white coloured) polar coordinate system indicating the viewing geometries (black ring), and illumination condition (yellow disc) with respect to the orientation of the tree rows. While the overall azimuthal BRF pattern is similar between different RT models (especially for illumination directions parallel to the tree rows), the magnitude of the generated BRF quantities however, may be rather different. One exception to this are the rayspread and pbrt model simulations which ingested all the available architectural information from the RAMI website and agree very closely. Given the large

Table 4
List of the RAMI-IV spectral band labels together with their corresponding central wavelengths (in nanometer). More detailed information are available from the RAMI web site, namely http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI-IV/RAMI-IV.php.

Band ID	B01	B02	B03	B04	B05	B06	B07	B08	B09
Wavelength	442	490	530	551	570	631	661	674	697
Band ID	B10	B11	B12	B13	B14	B15	B16	B17	B18
Wavelength	706	712	741	752	781	872	895	910	1019

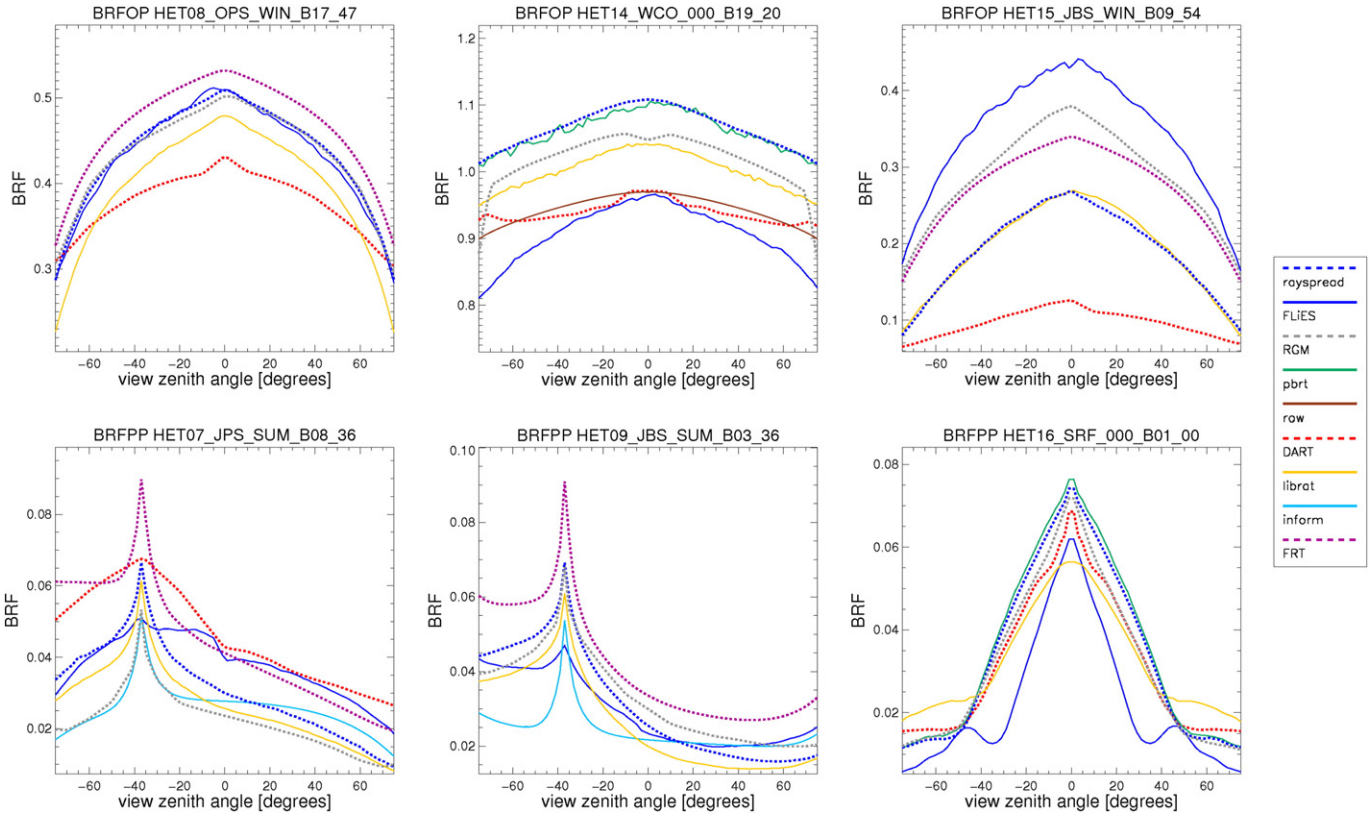

Fig. 3. Examples of model simulated domain-level BRFs along the orthogonal plane (top row) and principal plane (bottom row) for different spectral bands and for all of the six actual canopy test cases. Band 1, 3, 8, 9 and 17 relate to 442 nm, 530 nm, 674 nm, 697 nm and 910 nm, respectively. Band 19 relates to the purist case where no absorption is possible.

degree of BRF dispersion in general, the latter agreement is important and will be made use of later on.

To understand the BRF patterns in Fig. 4, it is often sufficient to consider the changes affecting the visible fractions of illuminated and shadowed scene components when the view azimuth angle is varied. Consider, for example, the BRF simulations depicted in the top row of Fig. 4 where the illumination direction was parallel to the orientation of the plant rows (i.e., at an azimuth of 90°) and the illumination zenith angle was 50°. Here, all shadows are cast along the direction of the tree rows, that is, either onto the architectural elements contained within the crowns or onto the background below the trees in a row. A viewer looking (down along a 37° zenith angle) in an azimuthal direction perpendicular to the tree row orientation, can see some of this shadowed background area because the tree crowns do not reach all the way down to the ground. Hence, the BRF values at view azimuths of 0° and 180° are noticeably lower in this configuration than if the view direction were parallel to the orientation of the tree rows and little of the shadowed background area would be visible (i.e., at view azimuths of 90° and 270°). The fact that the magnitude of the latter BRF peaks was smaller at 270° than at 90° view azimuth arises because the visible fraction of shadowed background – particularly in the gaps caused by missing trees in a row – is larger when the viewer looks toward the light source, rather than if the viewer looks at the target from the same side as the source that illuminates it. Obviously, this effect vanishes when the light source is directly overhead and the gaps in a tree row do not contain any shadow (compare with the panels in the third row of Fig. 4).

When the illumination azimuth is perpendicular to the tree rows, then the background area in between the tree rows becomes covered by shadows and different azimuthal BRF patterns emerge (compare with the two bottom rows in Fig. 4). Whether the solar zenith angle is at 20° or 50°, the smallest amount of visible background shadow always occurs when the viewing azimuth coincides with that of the radiation source (i.e., at 0° azimuth here). By contrast, the azimuth angle at which the largest amount of background shadow becomes visible to a viewer depends on the zenith angle of the light source. At a solar zenith angle of 20°, when the area between the tree rows is only partially covered by shadow, the minimum BRF value occurs at a view azimuth angle of 270° (because this maximises the shadowed background area directly below the tree crowns that now becomes visible). If the solar zenith angle is equal to 50° (and the area between tree crowns is almost completely filled with shadow), then two BRF minima occur at spectrally dependent azimuths somewhat after 90° (or before 270°), where the visible background shadow in between the tree rows and below the tree crowns is largest. As the viewer continues to vary its azimuthal position from 90° (or 270°) to 180°, increasing amounts of the illuminated background directly below the tree crowns come into view with the net effect of raising the BRF again to a small local maximum at 180° view azimuth.

To understand the differences in the azimuthal location of the BRF maximum in the last two rows in Fig. 4 one must also consult the spectral properties of the various scene components. More specifically, if the illumination zenith angle is set to 50° then the BRF in the red spectral band decreases monotonically as the view azimuth varies from 0° to 90 (or to 270°), whereas for an illumination zenith angle of 20° the BRF rises first to its maximum value before falling off again to its overall minimum. The latter is because only a fraction of the width between tree rows is covered with shadow such that – as increasing amounts of the background area between the tree rows come into view – the BRF may actually rise shortly, if the illuminated background is much brighter than the foliage or wood. In the near-infrared, where the background is darker than the foliage, no BRF peak is generated at view azimuth angles larger than 0°. Interestingly it is for illumination scenarios perpendicular to the rows that the FLiES and to a lesser extent also the row models generate azimuthal BRF patterns that differ noticeably from those of the other participants. Here it should be pointed out that the row model uses a statistical representation of the Wellington Citrus Orchard scene – which makes use of only 6 structural parameters – rather

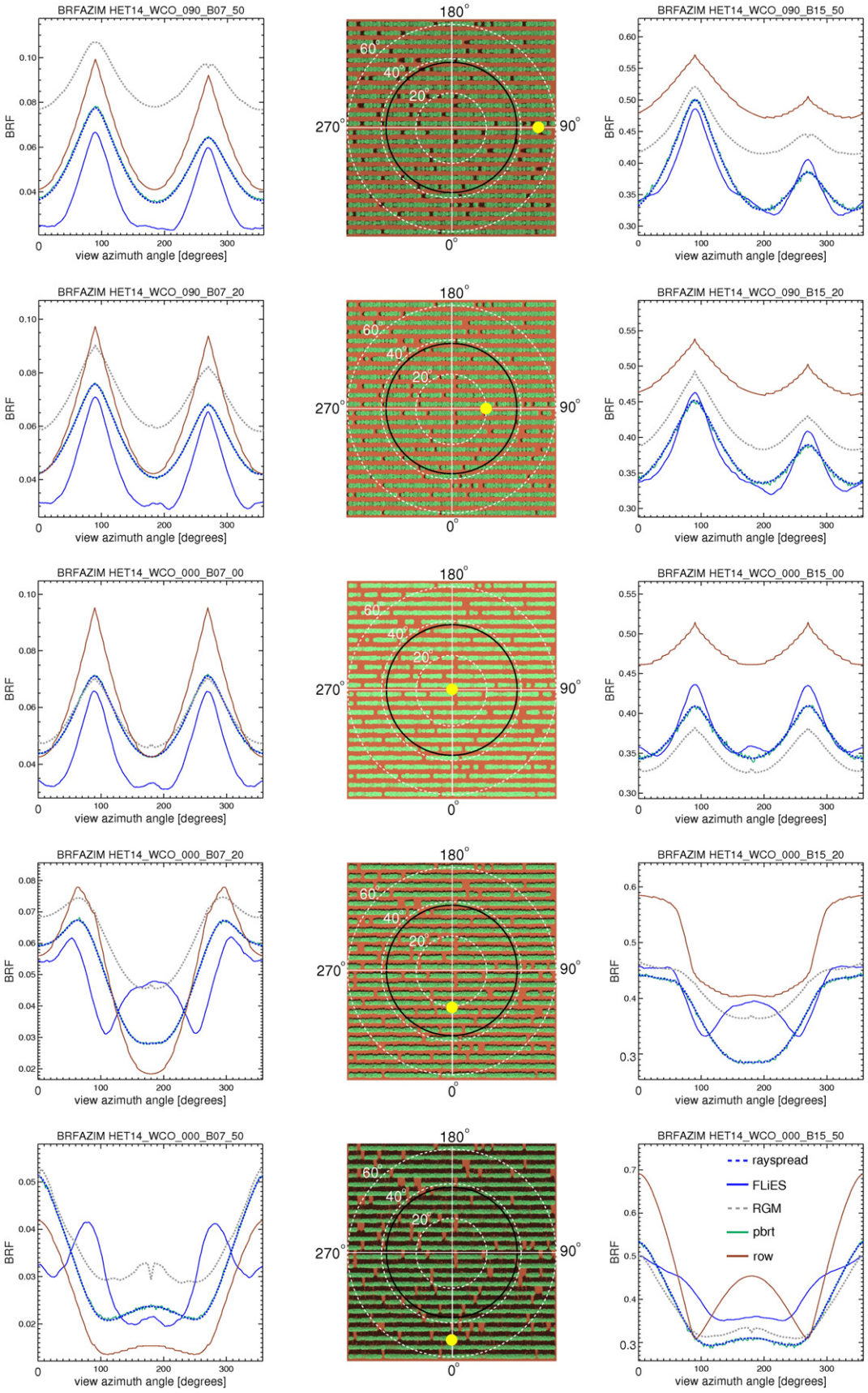

Fig. 4. Model simulations in band 7 (661 nm; left column) and band 15 (865 nm; right column) of the BRF leaving the Wellington Citrus Orchard scene at a constant view zenith angle of 37° and view azimuth angles from 0 to 360°. Graphical depictions of the target canopy when viewed from nadir are shown in the middle column, together with a (white coloured) polar coordinate system indicating the viewing (black ring) and illumination (yellow disc) configuration with respect to the planting pattern.

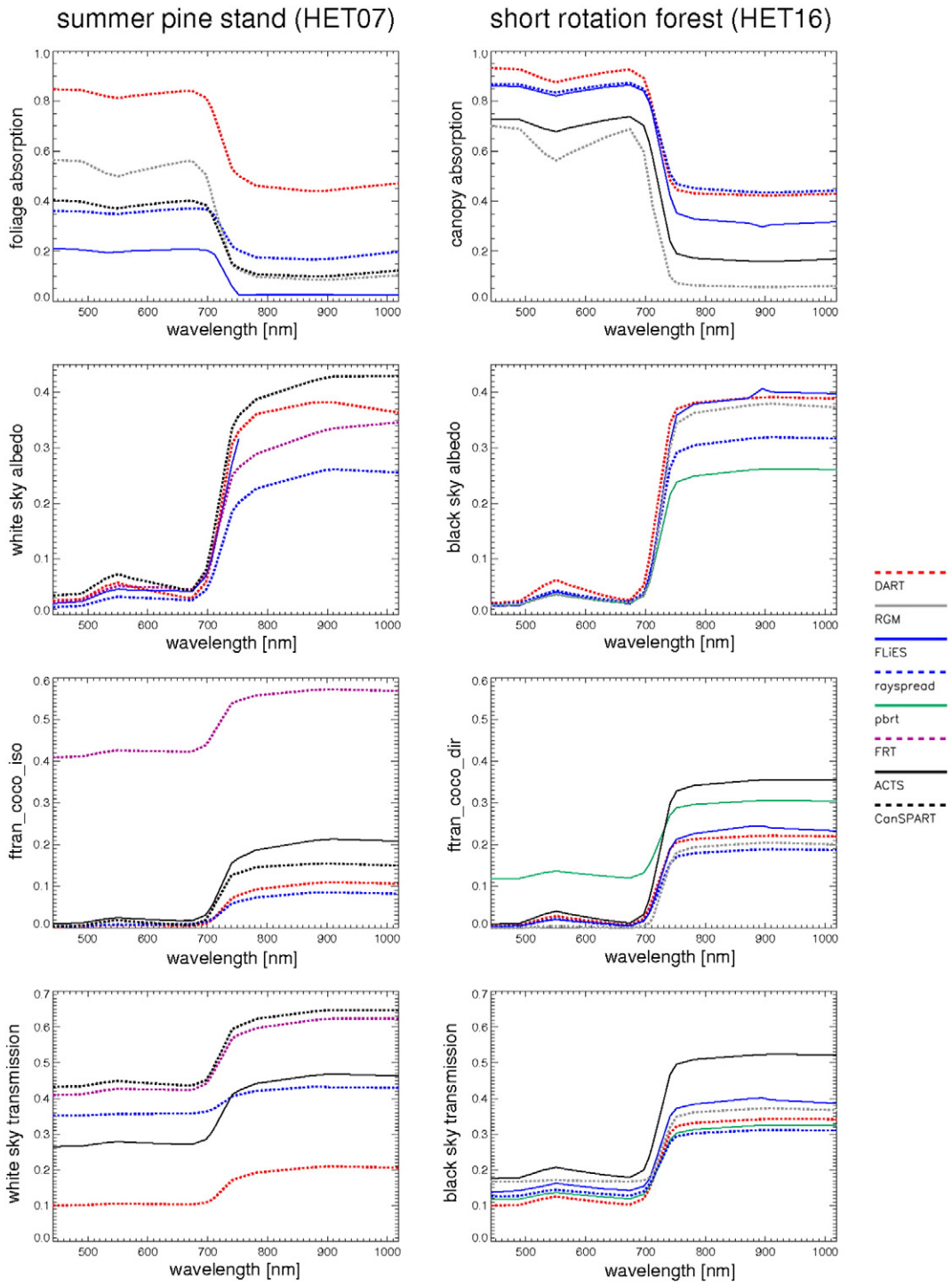

Fig. 5. Examples of model simulated flux spectra pertaining to the HET07 Järvselja pine stand (left column) and the HET16 Lombardy short rotation forest illuminated with a zenith angle of 50° perpendicular to the row orientation (right column). The simulations cover the wavelength range from 442 nm (B1) to 1019 nm (B18). The first row depicts absorption quantities, the second albedo quantities and the last total transmission quantities. The third row depicts that transmission component which has undergone at least one scattering interaction on its way down through the canopy for isotropic diffuse incident radiation only (left panel) and direct incident radiation only (right panel).

than the thousands of geometric primitives available on the RAMI website.

3.4. Spectral flux simulations

Fig. 5 shows the spectra of RT model simulated flux quantities pertaining to the HET07 Järvselja summer pine stand (left column) as well as the HET16 Lombardy short rotation forest (right column) illuminated with a zenith angle of 50° perpendicular to the row orientation. The spectral bands cover a wavelength range from 442 nm (B1) to 1019 nm (B18). Every graph in Fig. 5 relates to a different radiative flux quantity. The first row pertains to absorption fluxes, the second to hemispherical reflectances (i.e., spectral albedo), and the fourth to canopy transmission quantities. The third row depicts quantities that are not measureable in the field, namely, the component of the transmission that has undergone at

least one scattering interaction on its way down through the canopy if the incident radiation at the top-of-canopy level was composed of isotropic diffuse radiation only (left panel) and of direct solar radiation only (right panel).

As expected, both the foliage and canopy absorption are the highest in the blue (B1–B2) and red (B7–B9) part of the solar spectrum with a small absorption dip occurring in the green (B4) followed by a more substantial decrease in the near-infrared region (B12–B18). This spectral behaviour is coherent and subsequently inverted for the albedo and transmission components (last 3 rows of Fig. 5). The dispersion of simulation results, however, appears to be far more substantial for canopy absorption and transmission quantities – where both the visible and the near-infrared spectral domains are affected – than for the canopy albedo simulations, where the outcome is rather similar in the visible. In the absence of obvious operator glitches and model bugs, the latter could be simply due to the contribution of the background (dominating the reflected signal in the visible), whereas the former could be a consequence of different representations of the target canopies in the RT models (because plant interception is highly correlated with their absorption in the visible (e.g., Widlowski, 2010) and RT quantities are particularly affected by the wood structure representation in the near-infrared (Widlowski et al., 2013)).

3.5. Binary hemispherical photographs

Hemispherical photography is one of the most common tools used in field validation efforts of satellite derived Leaf Area Index (LAI) products, e.g., Claverie Vermote, Weiss, Baret, Hagolle, and Demarrez (2013). In these efforts the canopy cover is separated from the sky or background with the help of thresholding operations, e.g., Jonckheere, Nackaerts, Muys, and Coppin, (2005); Côté, Fournier, and Verstraete, (in press). The top and bottom rows of Fig. 6 display graphical depictions of model simulations emulating thresholded hemispherical photographs for the nominal location C ($x = 20$ m, $y = 80$ m,$z = 0.05$ m) within the Järvselja Scots Pine scene (HET07) as well as the nominal location E ($x = 50$ m, $y = 50$ m,$z = 0.05$ m) within the Ofenpass Mountain Pine scene (HET08), respectively. The central panels depict the directional gap probability that was derived from these data using zenith angle bins of 1° width. The examples shown in Fig. 6 are typical in the sense that 1) camera locations often differed from the nominal positions e.g., (FLiES for the HET07:C case and librat for the HET08:E case), and 2) blunders often occurred in the simulations (e.g., tree floating above background (RGM in the HET08:E case), crowns floating above trunks (FLiES), no indefinitely repeating cyclic boundary conditions were applied (RGM and to a lesser extend also price for the HET07 case).

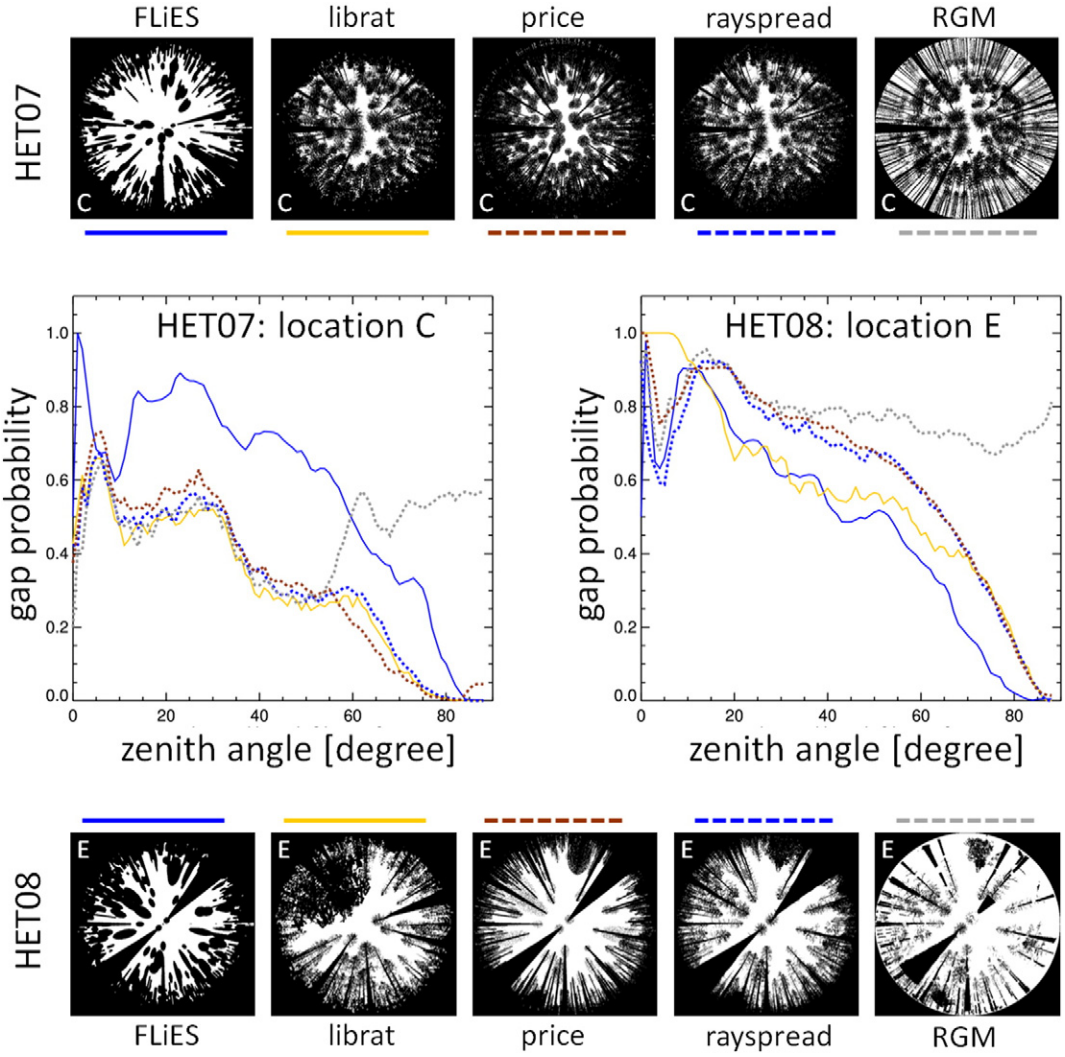

Fig. 6. Model simulated thresholded hemispherical photographs for camera location C inside the Järvselja summer pine stand (top row) and camera position E inside the Ofenpass Pine stand (bottom row). The middle row shows graphs depicting the directional gap probability derived from the thresholded hemispherical photograph simulations depicted in the top and bottom rows. The gap probability estimates were computed using zenith angle bins of 1° width.

In absence of user error, the simulations of librat, rayspread and price are rather similar. For the HET07 case, one notices that the gap probability of price is slightly larger for zenith angles below ~40° and also above ~85°. The first of these is caused by the rounded tree crowns used by price, while the second one is caused by a too low number of scene duplications. The gap probability of RGM, which up to a zenith angle of 45–50° agrees rather well with that of the previous 3 models, rises rapidly thereafter due to missing duplications of the HET07 scene.

The gap probabilities of the FLiES model are rather different from those of the other models. Among the RT models simulating the THP images, FLiES was the only one that represented foliage elements as a gas of scatterers within some enclosing volume. Following the measurement definition of the RAMI website, this hence resulted in the entire tree crown volume becoming an area of zero transmission (unlike the crowns that used discrete foliage representations where gaps could exist between foliage elements). This in turn explains the somewhat lower gap probabilities of the FLiES model in the HET08 case. Future phases of RAMI should refine the definition of this measurement type to permit non-binary probabilities of transmission when tree crowns are represented statistically as a gas of oriented scatterers. However, since the number of error-free THP simulations were rather small (2 models on average) it was decided not to include this measurement type in the detailed analysis stage below.

4. Analysis of simulation results

In order to assess the quality of the RT simulation methods that participated in RAMI-IV, the submitted BRF and/or flux quantities must be compared against fiduciary reference quantities. For consistency purposes these references must relate 1) to identical definitions of RT quantities and measurements setups, 2) to the same architectural and spectro-directional target characteristics, and 3) to identical illumination conditions as those described on the RAMI website (and used as input to the RT simulation methods). The use of reference quantities pertaining to other target and/or observation conditions is not considered appropriate here, since this introduces new (and often poorly characterised) sources of uncertainty that conceal the true quality of the simulation methods under study. Accordingly, rather than turning toward in situ observations as an evaluation criteria (e.g., Kuusk et al., 2010; Kuusk, Kuusk, & Lang, 2014), this study follows the approach of previous RAMI phases and evaluate the quality of simulation methods in terms of 1) the consistency between the generated RT quantities, 2) deviations from an absolute truth criteria (where available), and 3) the proximity to a reference acting as surrogate truth and originating from among the available RT simulation results.

4.1. Consistency of RT simulations

Model consistency tests were applied separately to BRF and flux simulations. Already during the active submission phase of RAMI-IV, the coordinators scrutinized the angular shapes of the BRF simulations to identify obvious simulation errors. If detected, then the relevant operators were informed and asked to verify their results and – if deemed necessary – to submit new simulation results. Next, the consistency of the BRF simulations was verified, that is, given that the single-collided and multiple-collided BRF components were provided, they were added up and compared to the value of the total BRF (for every view angle, spectral band, and canopy scenario). BRF consistency was generally found to exist with the exception of 1) RGM where some minor discrepancies were occasionally found for the largest (positive and negative) view zenith angles, 2) librat which after providing new simulation results suffered from a BRF inconsistency of around 0.15 in the near infrared for both the HET07 and HET09 test cases, and 3) DART which showed a deviation of about 0.15 between the total BRF and the sum of BRF components for the HET07 purist case (B19) only. In the case of librat it is possible that this was caused by post-processing of spectrally invariant components of the overall BRF quantity.

The single-uncollided BRF contribution (ρ_{uc}), corresponding to radiation not having interacted with the canopy architecture, but only with the underlying background, can be subjected to spectral consistency testing. For this, one must first compute the ratio of ρ_{uc} in two different spectral bands (for the same view geometry and canopy scene). This value must be equal to the ratio of the background BRDFs from the canopy scene of interest (in the same two spectral bands and viewing geometry). The mean absolute spectral bias (MASB) is, therefore, the absolute difference between these two quotients averaged over all relevant viewing conditions, spectral bands and also canopy scenarios. The MASB (and the maximum value of the absolute spectral bias) was found to be equal to 0.001 (0.030) for DART having n = 6764, 0.018 (0.277) for FLiES having n = 8208, 0.033 (0.099) for FRT having n = 2736, 0.075 (0.488) for librat having n = 2736, 0.000 (0.001) for pbrt having n = 5472, 0.002 (0.047) for rayspread having n = 8208, and 0.010 (0.071) for RGM having n = 8208 and 0.000 (0.001) for row having n = 2736 – where n denotes the number of available ρ_{uc} values per models. These numbers clearly indicate that several of the model do not behave in a consistent manner here.

Flux quantities could also be evaluated for spectral consistency. More specifically the uncollided transmission component should be a spectral constant. However only DART, rayspread and pbrt provided this information in all 19 bands. While the data for the last two was spectrally flat, the DART simulations for the Järvselja pine stand were slowly increasing from 0.19 in B01 to about 0.21 in B05 and with a spurious dip (down to 0.18) occurring at B17.

On the basis of the RT quantities prescribed for RAMI-IV, it was possible to assess the energy conservation of simulation methods in a given spectral band. Ignoring the impact of non-zero horizontal radiation fluxes (Widlowski, Lavergne, Pinty, Verstraete, & Gobron, 2006) the deviation from energy conservation can be written as:

$$\Delta F = 1A + [f_{dir} \cdot R_{dir} + f_{iso} \cdot R_{iso}] - (1-\alpha) \cdot [f_{dir} \cdot T_{dir} + f_{iso} \cdot T_{iso}] \qquad (1)$$

where the fractions of incident direct and isotropic diffuse radiation are related as $f_{dir} = 1 - f_{iso}$ and A is the total absorption of the canopy. The Lambertian background albedo is given by α, while T relates to the canopy transmission and R is the spectral albedo of the scene (separated into a component for the direct and the diffuse incident radiation). While CanSPART showed a perfect energy balance record, and FLiES exhibited only minimal deviations, the models DART, rayspread and RGM showed some deviations. In the case of rayspread, the ΔF values increased together with the albedo of the canopy to reach ~2% in B18. This was because the measurement specification had ignored the contribution of radiation being scattered multiple times down to the background – and explains, for example, why the rayspread transmission simulations were among the lowest in the near-infrared, but not in the visible in Fig. 4. In the case of DART the energy conservation bias, i.e., ΔF, showed no systematic spectral dependency and varied in magnitude from 5% to about 40%. For RGM the deviation from energy conservation exhibited a spectral pattern similar to the shape of a leaf reflectance spectra, that is, it was lowest in the blue and red spectral ranges (or bands B1–B2 and B6–B8) with a small rise in the green spectral range (bands B3–B5) and a major increase in the near-infrared (B10–B18). While the value of ΔF was always negative for DART, it could take on either signs in the case of RGM.

Fortunately, the purist case scenario of band 19 allowed one to evaluate the consistency of RT simulation methods against absolute reference values. In this band the spectral properties had been set such as to ensure that foliage and canopy absorption were both zero and the black sky albedo was equal to unity (no diffuse light source was present). Overall 8 different RT models provided results related to band 19. Defining bias as the mean difference between the true (T) and simulated (S) flux quantities, i.e., as $\beta = \langle T - S \rangle$, the following results were obtained for black sky

albedo simulations: $\beta = 0.216$ for DART, 0.062 for FLiES, 0.157 for pbrt, 0.000 for rayspread, 0.296 for RGM, 0.073 for FRT, and 0.000 for CanSPART. For canopy (foliage) absorption the bias was: −0.167 (−0.137) for DART, −0.053 (−0.046) for FLiES, and 0.000 (0.000) for FRT, rayspread and CanSPART. Hence most of the RT simulation methods underestimated the true albedo and overestimated the true (canopy and foliage) absorption under these purist case conditions.

When computing the dispersion of all models simulated BRF values around their ensemble mean for all of the spectral bands and actual canopy scenarios in RAMI-IV, one obtained a standard deviation equal to 0.050 on average (using around 246,000 data points). In the visible part of the spectrum the standard deviation was typically around 0.033 while in the near-infrared it was 0.067. These statistics may be indicative of the typical uncertainties that can be expected when actual canopy sites are reconstructed within RT models in support of specific scientific studies and/or applications.

4.2. Conformity testing

Reliable assessments of the conformity of a candidate *item* require 1) a clear definition of the nature of the item under consideration, 2) an unambiguous specification of the requirements that the item must satisfy in order to be declared compliant (or non-compliant), 3) access to an ideally unbiased reference item, and 4) reliable estimates of the uncertainty associated with both the candidate and reference items, e.g., JCGM-106 (2012). In the context of RAMI-IV, an item refers to the overall *method* that was used to obtain simulated estimates of local or domain-level RT quantities for a series of realistic looking virtual plant environments. Hence, the *method* relates to the assumptions and shortcuts that are inherent to the participating model, as well as to the choices and decision that the model operator(s) have made in the process of obtaining the simulation results. Among others, model operators may have been prompted by the high degree of realism of the various actual canopy scenarios a) to adapt the available test case descriptions to the capabilities of the model, b) to implement a computing strategy enabling the timely generation of the prescribed physical quantities within the limits imposed by the available hardware (for example, through the use of approximate calculations that are computationally more efficient), and c) to reconfigure the computing environment in which the model is running.

As such, during the analysis of the first set of results from RAMI-IV (abstract canopies), Widlowski et al. (2013) subdivided the combined standard uncertainty (u_c) of the candidate simulation methods into two conceptual contributions:

$$u_c = \sqrt{u_{c_{op}}^2 + u_{c_{mod}}^2} \tag{2}$$

where $u_{c_{mod}}$ refers to combined standard uncertainty of the RT *model*; that is, the uncertainty contributions arising from simplifications, parameterizations, and errors in the mathematical formulation of the physical concepts used by the model. $u_{c_{op}}$ relates to the combined standard uncertainty of the *operator*, that is, uncertainty contributions which are the consequence of operator choices and errors in the implementation of the test cases, the running of the model, and any subsequent post-processing steps.

In manufacturing contexts conformity testing is used to establish whether the properties of candidate items fall within permissible bounds from the ideal characteristics that the product should have. This is somewhat different from RAMI, where the true magnitude of the simulated RT quantities is generally unknowable a priori and also different for every virtual canopy scene, spectral band, viewing and illumination geometry. An adequate framework for conformity testing in a context like RAMI, thus, is to compare the *apparent error* (ε = R–C) – that exists between the reference item (R) and the candidate item (C) – to some predefined maximum permissible deviation limit (Δ). However, when conformity is sought over an ensemble of different scenarios (e.g., for RT simulations covering different spectral bands, canopy architectures, as well as view and illumination geometries), then a minimum required compliance rate ($\breve{R}_c$) must also be specified before an item – that is a particular RT simulation method – can be declared compliant. Different decision rules may be used for this depending on whether the combined standard uncertainty that is associated with the apparent error estimate, i.e., $u_\varepsilon = \sqrt{u_R^2 + u_C^2}$ – in the case of independent methods – is accounted for or not.

4.2.1. Shared risk approach

In cases where the combined standard uncertainty of both the reference and the candidate items are negligible, a so called shared risk approach may be used to determine conformity, e.g., JCGM-106 (2012). In such a scheme, compliance is asserted whenever the rate with which the estimated apparent error (ε) falls within the permissible tolerance interval (Δ) is higher than the minimum required compliance rate ($\breve{R}_c$):

$$P(|\varepsilon| \le \Delta) \ge \breve{R}_c \tag{3}$$

where $|\varepsilon|$ relates to the absolute deviation between the candidate and reference items. Selecting a shared risk approach implies that the producer of an item (i.e., the author of a RT simulation method here) and its consumer (i.e., the user selecting that RT simulation method for a particular task) both share the risk of wrongly rejected or wrongly accepted items, respectively. In order to keep these risks within reasonable levels, a shared risk approach should only be applied if the uncertainty of the apparent error estimate is sufficiently small. While this is widely understood, there are however situations where little or no (reliable) uncertainty information is available and where the shared risk approach seems to offer a carefree solution to more tedious alternatives. One such case is the use *of ensemble averages* as a means to evaluate candidate methods in inter-comparison efforts like RAMI. For demonstration purposes, the shared risk approach will thus be applied to the actual canopy simulations of RAMI-IV without prior assessment of the uncertainty that is associated with the chosen reference data.

Rather than selecting specific values of Δ and $\breve{R}_c$ for each one of the RT quantities addressed in RAMI-IV, Fig. 7 documents how the compliance rate of the RT simulation methods, i.e., $P(|\varepsilon| \le \Delta)$ changes as a function of the width of the tolerance interval (Δ). For ease of representation, the latter is defined as a percentage of the reference value, i.e., $\Delta = (f/100) \cdot R$ and varied from f = 1% to f = 50% of the reference (R). Due to the substantial spread between the contributing RT simulation results (see Fig. 3), as well as the lack of an obvious cluster of methods providing closely agreeing simulation results for all of the actual canopy scenarios of RAMI-IV, it was decided to define a qualified reference dataset on the basis of all submitted simulation results. This is in accordance with the procedure defined in ISO-13528 for cases where neither an absolute reference value is available nor a trustworthy participant can be identified. However, rather than working with a simply ensemble average, the robust mean algorithm proposed in annex C of ISO-13528 (2005) was used instead. The visual adequacy of this algorithm as well as the angular smoothness of the resulting reference BRF data can, for example, be seen in Fig. 7 of Widlowski et al. (2013).

Each panel in Fig. 7 relates to a different radiative quantity. Simulation results from all of the RAMI-IV actual canopy scenarios are included, with N_{max} indicating the maximum number of simulation results that could be submitted for a given RT quantity. In the case of flux quantities pertaining to isotropic illumination conditions, N_{max} corresponds to a mere 76 data points whereas for BRF simulations along an azimuthal ring, N_{max} was of the order of 48000. As a consequence the curves of the BRF quantities are smoother than those of the flux quantities. Horizontal lines and abrupt vertical jumps are indicative of small numbers of submitted simulation results occupying disjoint quality regimes with respect to the location of the reference. An obvious example of such a

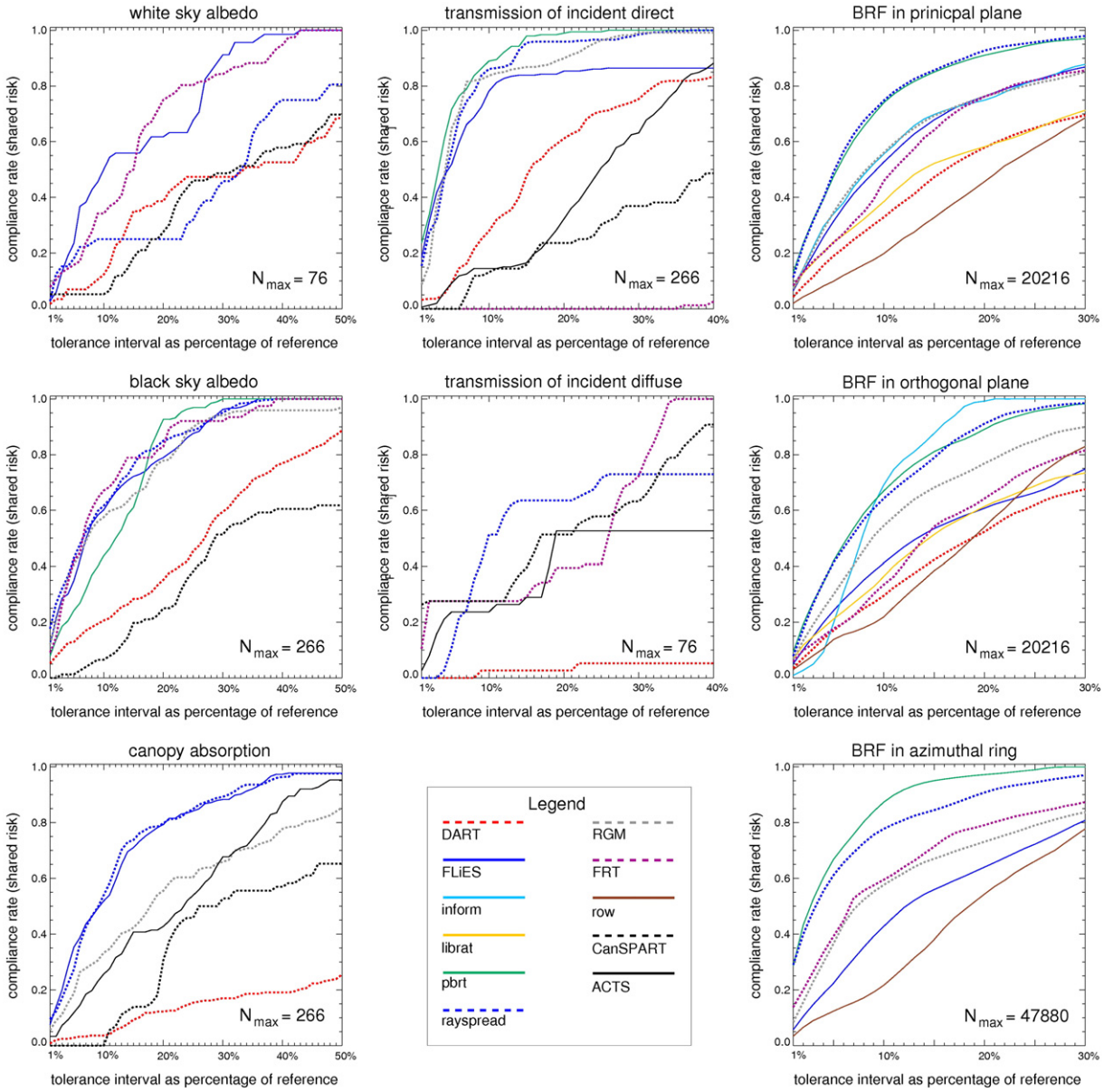

Fig. 7. Change in the compliance rate of the RT simulation methods as a function of the width of the tolerance interval (Δ) around the reference. Compliance is determined here using a shared risk approach, i.e., as |R–C| < Δ without accounting for the uncertainty associated with the reference (R) and candidate (C) methods. The value of the reference was determined using the robust mean algorithm proposed in annex C of ISO-13528. The size of the tolerance interval was varied from 1% to 50% of the reference, i.e., $\Delta = (f/100)\cdot R$. Different panels relate to different radiative quantities. All actual canopy scenarios of RAMI-IV are included. N_{max} indicates the maximum number of simulations that could contribute to the computation of the compliance rate. The smoother a curve the more simulations were provided.

behaviour is provided by the ACTS model in the central panel of Fig. 7 where half of the submitted simulation results are off the graph.

The more rapid a curve rises, in Fig. 7, the closer it is to the qualified reference. If the compliance rate reaches 1 for a particular value of Δ, then all results submitted by that RT simulation method do fall within the specified tolerance interval. The first simulation methods to reach the 100% compliance rate are inform (middle right panel) followed by pbrt (bottom right panel) when the tolerance interval $\Delta = (f/100)\cdot R$ is equal to f = 20% and 30% of the reference BRF value, respectively. For transmission simulations of direct radiation, pbrt is the first to reach a 100% compliance rate if f = 20%, while FRT reaches full compliance at f = 35% for transmission simulations of diffuse radiation. Strangely enough the same FRT method delivers rather different results with respect to the robust mean when the transmission of direct radiation is to be simulated. White sky albedo simulations reach full compliancy at f = 43% (FLiES) while black sky albedo results do so at f = 30% (pbrt). Canopy absorption simulations (left bottom panel in Fig. 7) are the only ones where full compliancy is not reached for tolerance intervals up to 50% of the reference value.

The Global Climate Observing System (GCOS) has provided indicative values for the accuracy of remotely sensed broadband surface albedo quantities (i.e., 0.0025 or 5% of the true value – whichever is larger) and fraction of absorbed photosynthetically active radiation (FAPAR) products (i.e., 0.05 or 10% of the true value – whichever is larger), e.g., GCOS (2011). GCOS does not indicate what the minimum required compliance rate ($\breve{R}_c$) for these products must be nor by what method compliance should be evaluated (Widlowski, 2015). However, if the retrieval algorithm is bias free, then in the absence of other error sources the distribution of albedo and FAPAR biases can be assumed Gaussian and $\breve{R}_c$ may thus be set equal to 68.3% or 95.5% depending on whether the GCOS accuracy criterion relates to 1 or 2 standard deviations of

the distribution of apparent biases. From the relevant graphs in Fig. 7 one can read a compliance rate of at most 30% for the white sky albedo (and no larger than 42% for the black sky albedo) if the width of the tolerance interval is set to f = 5% of the reference. Similarly if f = 10%, then for canopy absorption simulations the best compliance rate will be around 53%. While the GCOS accuracy criteria relate to broader spectral ranges than the RT quantities depicted in Fig. 7, it is nevertheless surprising that the participating RT simulation methods – in a controlled environment like RAMI – do not seem capable of matching the GCOS accuracy criteria for albedo and FAPAR. Since several of the RT simulation methods in Fig. 7 were shown capable of meeting (a more stringent interpretation of) the above GCOS accuracy criteria during the recent analysis of RAMI-IV abstract canopy simulations (Widlowski et al., 2013), the current dispersion thus must come from 1) errors in the software or theory behind the simulation methods (this concerns models that are new to RAMI or have changed their RT formulation since the last comparison effort), 2) operator choices and inadvertent errors in the preparation, running and post-processing stages, and/or 3) unreliable reference data.

This picture is confirmed when examining the results for BRF simulations. Here, one would like to aim for similar levels of accuracy as are currently achievable for radiometric measurements acquired by optical instruments on board of Earth orbiting satellites. If one assumes a bias free instrument (and an absence of other perturbing factors), then the distribution of absolute calibration biases can be thought of as being Gaussian with a standard deviation of 2–6% depending on the spectral band and calibration method (e.g., Bruegge , Chrien, Ando, Diner, Abdou, & Helmlinger 2002; Govaerts, Clerici, & Clerbaux, 2004; Helder, Thome, Mishra, Chander, Xiong, & Angal, 2013; McCorkel, Thome, & Lockwood, 2013; Thome, Arai, Tsuchida, & Biggar, 2008; Wang, Czapla-Myers, Lyapustin, Thome, & Dutton, 2011). If the width of the tolerance interval in the relevant graphs of Fig. 7 is thus set to f = 5% of the reference, then one would expect a compliance rate of about 68.3% assuming that the RT model simulations were of a similar quality than real world vicarious calibration efforts. However, for BRF simulations along the principal plane the best compliance rate is around 44%, whereas in the orthogonal plane it is 40%. Only for simulations along the azimuthal plane does the pbrt simulation method reach a compliance rate of about 66%. Again, these findings contrast with the results that were obtained for several of the present participants during the analysis of the abstract canopy BRF simulations (Widlowski et al., 2013).

Earlier phases of RAMI have shown that the likelihood of identifying (and correcting) faulty RT formulations as well as operator induced errors increases as models contribute repeatedly to the same test cases in subsequent inter-comparison phases, particularly if they also provide sub-components of the required BRF and flux quantities. In this contribution our focus thus must be on the quality of the reference data that was used in the above shared risk analysis. One of the most lenient rules of thumb in legal metrology states that the maximum permissible uncertainty (MPU) – of the reference and candidate items together – should not exceed a third of the maximum permissible deviation, e.g., Sommer and Kochsiek (2002). In other words:

$$2u_\varepsilon = MPU \le \Delta/3 \tag{4}$$

where in the case of independent reference and candidate methods $u_\varepsilon = \sqrt{u_R^2 + u_C^2}$, as seen before. Intuitively, the uncertainty of the reference should ideally also be smaller than that of the candidate. For proficiency testing of laboratories (ISO-13528, 2005), as well as for some sectors of the automobile industry (Pendrill, 2014), the uncertainty of qualified reference method should satisfy:

$$u_R \le 0.3 \cdot u_\varepsilon \tag{5}$$

In other words the uncertainty of the reference method must be adequate with respect to the desired tolerance interval. Since u_ε is constrained by the size of Δ, Eqs. (4) and (5) thus imply that the tolerance interval should be 20 times larger than the combined standard uncertainty associated with the reference method.

As a thorough estimate of u_R is beyond the scope of this study (see for example JCGM-100(2008) or the various components of ISO-5725 (1994)), the uncertainty of the robust mean was estimated using the robust standard deviation provided by the algorithm of ISO-13528 (2005). As it turned out, the mean (maximum) value of the robust standard deviation was equal to 18% (56%) of the reference value for black sky albedo, 33% (68%) for white sky albedo, 32% (156%) for canopy absorption, 17% (80%) for the transmission of direct incident radiation, 34% (55%) for the transmission of diffuse incident radiation, as well as 16% (78%) of BRFs along the principal plane, 17% (63%) for BRFs along the orthogonal plane and 13% (56%) for BRFs along the azimuthal ring. These numbers make it clear that the robust mean estimate – that was used in the above shared risk analysis – was too uncertain to function as a reliable reference in conformity testing where the width of the tolerance interval does not exceed 50% of the reference. A more precise reference dataset and decision rule that accounts for the uncertainty in both the reference and candidate methods is therefore needed.

4.2.2. Guarded risk approaches

A guarded risk approach is used when the likelihood of a false acceptance (or a false rejection) is to be kept within well controlled limits. This necessitates a descriptor of the distribution of likely values for the target quantity (i.e., ε here). If a full description of the probability density function (PDF) of the likely values of ε were available – for example through a Monte Carlo propagation scheme (JCGM-101, 2008) – then conformance could be attributed to only those estimates of ε for which more than a predefined minimum fraction of the PDF falls within the specified tolerance interval (±Δ). In a 'Guide to the Uncertainty in Measurements' (GUM) framework, e.g., JCGM-100 (2008), where the PDF of likely values of the target quantity is assumed to be symmetric and unimodal, a coverage interval together with an associated probability statement can be used for the same purpose (JCGM-106, 2012). More specifically, the latter uses an expanded uncertainty $U(\varepsilon) = k \cdot u_\varepsilon$ to determine the width of the coverage interval [ε - U; ε + U]. The coverage probability that is associated with this coverage interval is defined by the choice of coverage factor k and depends on the actual PDF of ε. For normal distributions of likely values, $k = 1$ relates to a 68.3% coverage probability, while $k = 2$ relates to 95.5% coverage probability. Compliance may then be decreed if the coverage interval (at a predefined coverage probability) is fully contained within the tolerance interval, that is, when $|\varepsilon| + k \cdot u_\varepsilon \le \Delta$, e.g., (ISO-10576, 2003).

The expanded uncertainty thus acts as *a guard band* to reduce the risk of accepting a non-conforming item, e.g., ISO-14253 (1998). Conceptually speaking these guard bands can be placed on the inside of the tolerance interval if a *guarded acceptance* of items is wanted (i.e., when clear evidence is required before an item is accepted as conforming), or else, on the outside of the tolerance interval if a *guarded rejection* of items is the goal (i.e., when clear evidence is required that a limit has been exceeded before any negative action is taken) (JCGM-106, 2012). The guarded acceptance approach thus reduces the tolerance interval to $\pm\,|\Delta - k \cdot u_\varepsilon|$ before applying a shared risk approach within the resulting *acceptance interval*. Contrarily, the guarded rejection approach uses the shared risk approach to identify items that fall outside of the enlarged interval covering $\pm\,|\Delta + k \cdot u_\varepsilon|$. Fig. 8 provides a schematic overview of these two guarded risk approaches.

When dealing with candidate methods that are applied to ensembles of different scenarios, the guarded acceptance approach will lead to a declaration of conformity if:

$$P(|\varepsilon| \le \Delta - k \cdot u_\varepsilon) \ge \breve{R}_c \tag{6}$$

while a guarded rejection approach will result in a declaration of non-

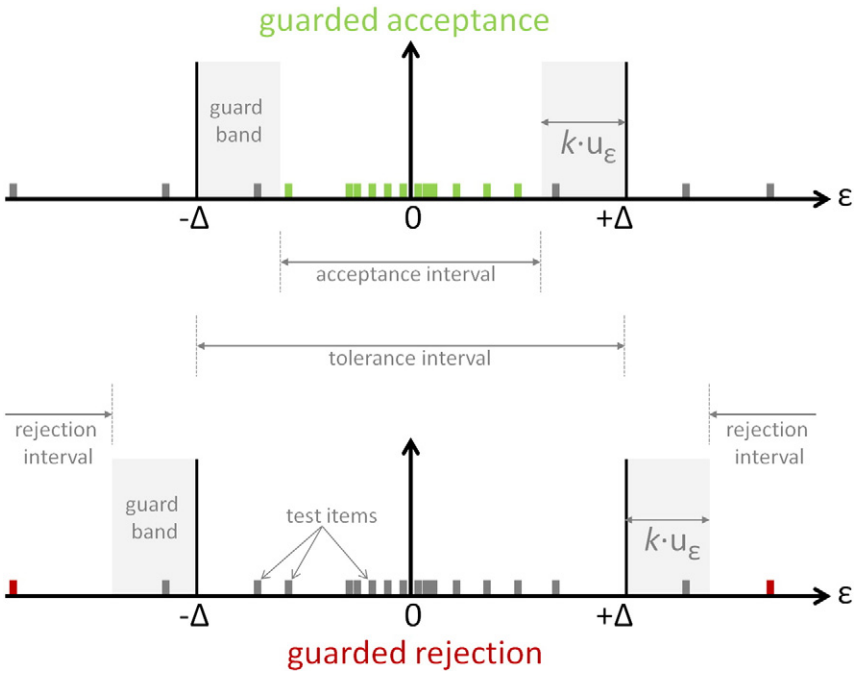

Fig. 8. Schematic depiction of two guarded risk decision rules applied to a symmetric tolerance interval (±Δ) in the context of apparent errors (ε) between a candidate and reference method. Both approaches make use of so called guard bands having a width $k \cdot u_\varepsilon$ where k is a coverage factor typically set to 2 and u_ε is the combined standard uncertainty associated with ε. The upper schematic depicts a guarded acceptance effort, where the guard bands are located on the inside of the tolerance interval. The (green) ε items that fall within the thus defined acceptance interval possess a higher likelihood of being truly compliant. The lower schematic shows a guarded rejection effort where the guard bands are added to the outside of the tolerance interval. The (red) ε items that fall within the resulting rejection interval thus possess a greater likelihood of being truly non-compliant. Since the width of the guard band in these two schematics was drawn in a static manner, the combined standard uncertainty must be assumed constant between different test items.

compliance whenever:

$$P(|\varepsilon| \geq \Delta + k \cdot u_\varepsilon) \geq \hat{R}_E \tag{7}$$

Here $k \cdot u_\varepsilon$ denotes the width of the guard band (which may be different for each one of the tested scenarios), $\breve{R}_c$) is the minimum required rate of compliance, and $\hat{R}_E$ is the maximum permissible rate of exceedance. If the condition in Eq. (6) were satisfied, then one could state with 95% confidence (if k = 2) that the results of the candidate method comply with the prescribed tolerance interval in at least $\breve{R}_c$) percent of all cases. Conversely, if the condition in Eq. (7) were found to be true, then one could be 95% certain (if k = 2) that the results of the candidate method do not satisfy the quality criteria.

In an effort to reduce the uncertainty of the reference data, the following examples will focus exclusively on total BRF simulations for the Wellington Citrus Orchard (HET14) scenario. From the examples in Fig. 4 it is known that the BRF simulation results of pbrt and rayspread agree very closely for this test case. This is corroborated by RMSE values of 0.0020 and 0.0021 for the (7220) BRF simulations along the principal and orthogonal planes. The rayspread simulation method was found to agree to within a fraction of 1% with analytical reference solutions obtained for single-scattered directional BRF components (see Fig. 7 in Widlowski et al. (2013)) and has been selected as one of 6 credible RT simulation methods in earlier phases of RAMI. At the same time, the pbrt simulation method has been extensively validated against data collected at the Wellington Citrus Orchard site (which coincidentally was the only actual canopy scenario for which pbrt submitted simulation results). The pbrt simulation results thus serve as independent evidence that operator induced errors are unlikely to have affected the HET14 BRF simulations of rayspread. Since the pbrt simulation results are generally noisier than those of rayspread (see the top middle panel in Fig. 3 as well as Widlowski et al., 2013), it was decided to base the reference data for the forthcoming analysis solely on the BRFs simulated with rayspread. Due to the large number of incident rays used in the rayspread simulation method (compare with Fig. 6 in Widlowski et al., 2013), its BRF data can be expected to have an experimental standard deviation equivalent to 0.24% of the true value. This subsequently fulfils the requirements of Eqs. (6) and (7), provided that the width of the tolerance interval is at least 5% of the reference BRF value.

Fig. 9 shows contour plots of the compliance rate that was obtained by applying a guarded acceptance approach to BRF simulations (along the orthogonal and principal plane) over the Wellington Citrus Orchard scenario (HET14). The abscissa relates to the tolerance interval Δ (expressed as a percentage of the reference, i.e., $\Delta = (f/100) \cdot R$), while the ordinate relates to the width of a guard band ($k \cdot u_\varepsilon$) (expressed as a percentage of the tolerance interval, i.e., $k \cdot u_\varepsilon = (g/100) \cdot \Delta$). As the width of the tolerance level increases then the compliance rate will increase (especially for small values of the guard band width). Conversely, as the guard band width increases then the compliance rate will decrease (especially for large values of the tolerance interval width). If the guard band width is equal to that of the tolerance interval, then the compliance rate will obviously be zero.

According to Eq. (4) the expanded uncertainty should not exceed one third of the tolerance interval. In Fig. 9 this corresponds to a maximum guard band width of 33% and is indicated by a light grey rectangle in each graph. At the same time, however, it would be meaningful if the RT simulation methods could achieve similar accuracy levels as radiometric calibration efforts. As such the tolerance interval in Fig. 9 should be of the order of 5% of the reference (for a nominal compliance rate of 68.3%). However, the only simulation method that is capable of meeting these criteria is pbrt. The next simulation method to reach this compliance rate is RGM (for a tolerance interval that equals 17% of the reference) followed by FLiES (here the tolerance interval equates 23% of the reference) under the assumption that there are no guard bands. The proximity of the 0.68 and 0.95 contour lines for the librat simulation

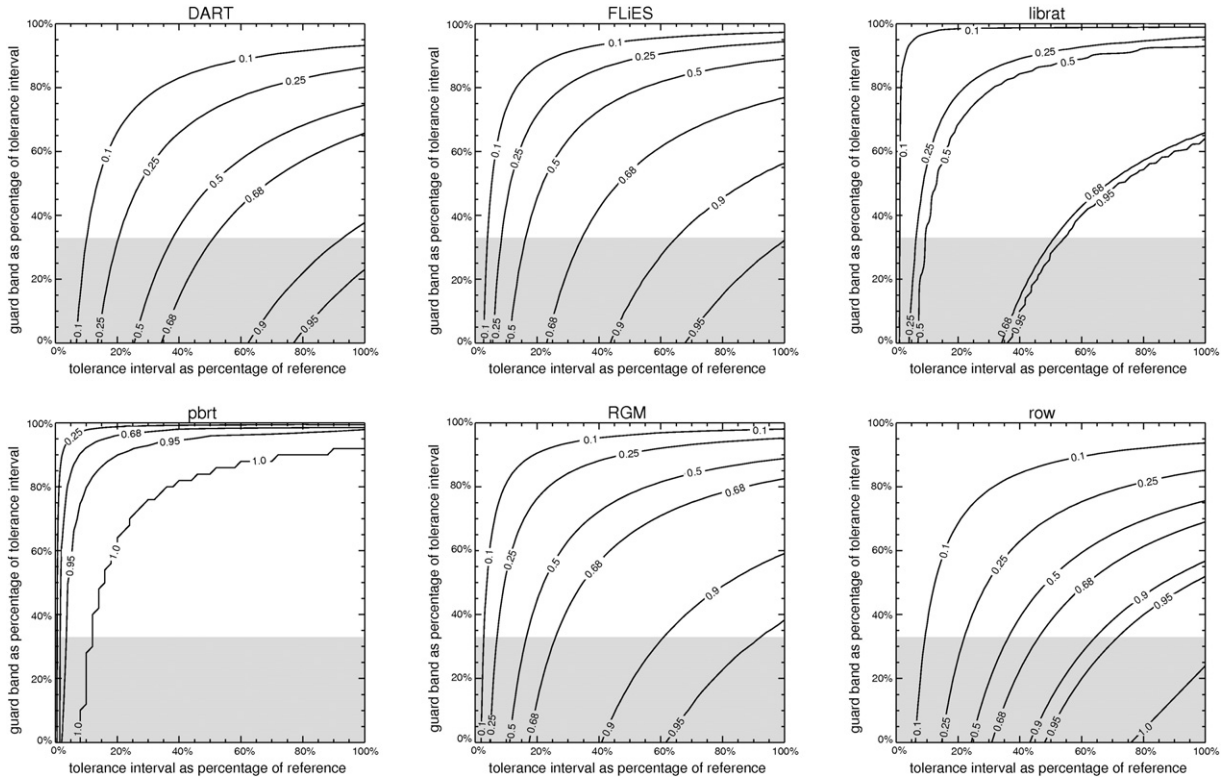

Fig. 9. Contour plots of the compliance rate of individual RT simulation methods for BRF simulations (in the orthogonal and principal plane) over the Wellington Citrus Orchard test case (HET14). Compliance is determined using a guarded acceptance approach, that is, if $|R–C| < \Delta - k \cdot u_\varepsilon$, where k is a coverage factor, u_ε is the combined standard uncertainty of the reference (R) and candidate (C) methods, and Δ is the tolerance interval (expressed as a percentage of the reference, i.e., $\Delta = (f/100) \cdot R$ with the percentage f featuring as the abscissa of the graphs). Note that $k \cdot u_\varepsilon$ is also the guard band width (expressed as a fraction of the tolerance interval, i.e., $k \cdot u_\varepsilon = (g/100) \cdot \Delta$ with the percentage g featuring as the ordinate of the graphs). The reference is provided by the rayspread simulation method. The grey area denotes the range of permissible uncertainty according to Eq. (4).

method (together with their large separation from the 0.5 contour line) arises from the fact that its BRF simulations at an illumination zenith angle of 50° follow the shape of the reference BRF but at a quasi constant offset. Given the excellent performance of the librat simulation method during the analysis of the abstract canopy cases of RAMI-IV (Widlowski et al., 2013) this pattern seems to point toward some sort of post-processing error or else an operator glitch in the setup of the scene and run. All in all, the results presented in Fig. 9 thus seem to confirm that, for all those models which have performed well in earlier phases of RAMI and that have not undergone any changes to their RT formulations, the most likely cause for the observed differences will be operator glitches in the setup and running of models (as well as during any post-processing stages). For models that are new to RAMI, the possibility of software errors in the RT formulations and/or theoretical flaws in the concept of the simulation method have also to be envisaged.

To complete the analysis, Fig. 10 provides an illustration of the rejection rates that were obtained when a guarded rejection approach was applied to the BRF simulations of the (HET14) Wellington Citrus Orchard test case. The meaning of the abscissa and ordinate is identical to that of Fig. 9. Without guard bands the rejection rate is the complement of the compliance rate, and thus the intersection of the 0.5, 0.9 and 0.95 contour lines with the abscissa of Fig. 9 occurs at the same tolerance intervals as that of the 0.5, 0.1 and 0.05 contour lines in Fig. 10, respectively. As the guard band width increases then the sum of the rejection and compliance rate will become increasingly smaller than unity. The missing items are those that fall within the guard bands – implying that not enough evidence exists for their respective rejection or acceptance.

Plausible values for the maximum permissible rate of exceedance $\hat{R}_E$ are of the order of 0.1, 0.5 or even 0.01 depending on context. If the width of the guard band is set equal to 33% of the tolerance interval (see Eq. (4)), then for $\hat{R}_E = 0.05$, the various RT simulation methods in Fig. 10 should be rejected if the tolerance interval Δ is required to be less than 62% (DART), 67% (FLiES), 35% (librat), 63% (RGM) and 48% (row) of the reference. According to this scenario, all model simulations (with the exception of those by pbrt) would thus have to be rejected for HET14. In a sense this conclusion is identical to that made when all actual canopy scenarios were compared against a robust mean reference using a shared risk approach in Fig. 7. The variability between the various RT quantities is simply too large, even between veteran RAMI participants, thus suggesting that the complexity of the proposed architectural scenarios has dramatically increased the likelihood of operator errors in the setup and running of models. The best means to assess the validity of this statement seems to be a rerun of the same canopy scenarios in a future phase of RAMI.

At the same time, it is however strongly recommended that the first time participants test their models against the various test cases proposed during the first three phases of RAMI. This has proven extremely useful to earlier RAMI participants who could more easily identify coding errors and other sources of errors on the basis of much simpler scenes than those proposed in the actual canopy scenarios of RAMI-IV. These model verifications can nowadays be carried out autonomously and in quasi-real time with the RAMI Online Model Checker (ROMC) facility that contains a reference dataset from six credible Monte Carlo RT models developed in the wake of the third phase of RAMI (see http://romc.jrc.ec.europa.eu/). The ROMC is continually expanded, among others with a reference dataset (of flux quantities in the visible and near-infrared) that was used to evaluate the quality of RT formulations embedded in the land surface schemes of soil–vegetation–atmosphere transfer (SVAT), numerical weather prediction (NWP), and global circulation models (Widlowsk Pinty, Clerici, Dai, De Kauwe, de Ridder, et al., 2011.).

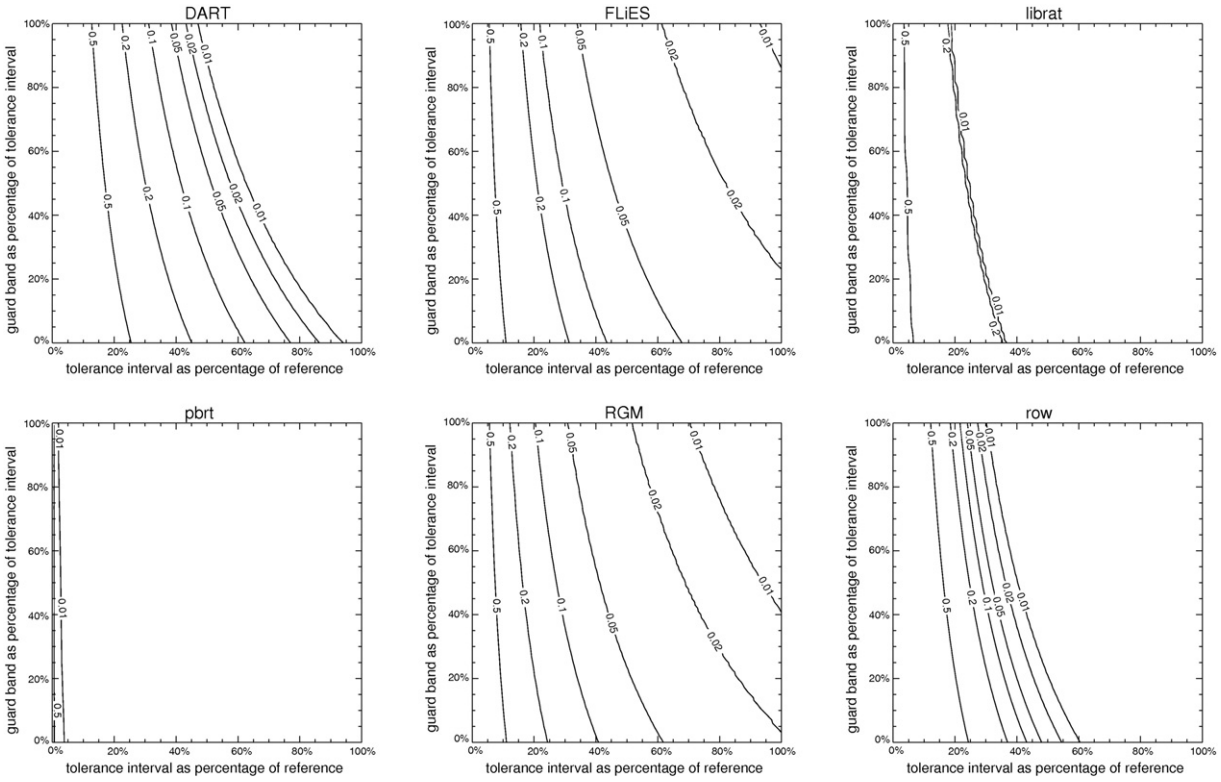

Fig. 10. Contour plots of the rejection rate of individual RT simulation methods for BRF simulations (in the orthogonal and principal plane) over the Wellington Citrus Orchard test case (HET14). Compliance is determined using a guarded rejection approach, that is, if $|R-C| > \Delta + k \cdot u\varepsilon$, where k is a coverage factor, $u\varepsilon$ is the combined standard uncertainty of the reference (R) and candidate (C) methods, and Δ is the tolerance interval (expressed as a percentage of the reference, i.e., $\Delta = (f/100) \cdot R$ with the percentage f featuring as the abscissa of the graphs). Note that $k \cdot u\varepsilon$ is also the guard band width (expressed as a fraction of the tolerance interval, i.e., $k \cdot u\varepsilon = (g/100) \cdot \Delta$ with the percentage g featuring as the ordinate of the graphs). The reference is provided by the rayspread simulation method.

5. Concluding remarks

In this phase of RAMI, RT simulation methods were executed on some of the most complex 3D canopy reconstructions currently available. These virtual plant environments – while based on field inventory data from actual test sites – do not constitute identical copies of their real world originals, however. This is because even the most intensive field measurement campaigns are not yet in a position to deliver descriptions of the spectro-directional and architectural properties for every element along the vertical and horizontal in the target canopy volume. A large number of data gaps thus had to be filled in order to arrive at comprehensive descriptions of the RAMI-IV virtual plant environments capable of meeting the input data requirements of the most versatile of RT simulation models. While this allowed for very realistic-looking canopy scenarios, it also implied that some participants had to make choices as to how this wealth of information could be adapted to more statistical canopy representation schemes contained within certain types of RT models. The latter aspect was intentional though, since it mimics the process that happens every time RT models are used to simulate the radiative properties of existing test sites; and thus the dispersion among RAMI-IV simulation results was expected to become indicative of the type of bias typical of such scientific studies.

Although the actual canopy scenes in RAMI-IV are not exact representations of the corresponding real-world targets, they allow to work under fully controlled experimental conditions in complex virtual environments that are among the most realistic that current state-of-the-art 3D canopy RT schemes are capable of handling. Outside of the scope of RAMI, these virtual plant environments can be used, for example, to assess the quality of quantitative retrieval algorithms on the basis of simulated satellite and in situ measurements, to improve spatial sampling techniques of radiative quantities in virtual field campaigns, or, to perform closure experiments that aim at matching actual remote sensing observations within their inherent uncertainty ranges. The latter process in particular, may identify potential gaps in our abilities to characterise complex forest sites in the field, as well as, in the manner in which in situ information is used in RT models. As better field measurement techniques become available (e.g., terrestrial and/or airborne laser scanning, as well as UAV-borne radiometers and/or portable leaf goniometers), and the necessary parameterisation schemes are developed to transfer these data into RT model relevant formats, then the gap between simulated sensor responses and actual remote sensing measurements over real test sites should gradually diminish. Hopefully, the availability of the RAMI-IV actual canopy scenarios, where every architectural and spectral property is known (and can be controlled), will prove helpful in these efforts.

A total of 12 different RT modelling groups provided simulation results for the actual canopy scenarios of RAMI-IV. The number of test cases that were submitted, however, varied widely between these groups. Priority was generally given to the simulation of measurable BRF and flux quantities, whereas their uncollided, single and multiple-collided subcomponents were often missing. As a consequence, it was rarely possible in those cases to identify the causes of any observed discrepancies. This is unfortunate, because the dispersion of simulation results was found to be rather large both for flux simulations and for BRFs. In addition, there were rarely any obvious clusters of simulation results that might help to guide the analysis. One exception was the HET14 Wellington Citrus Orchard scene where both rayspread and pbrt showed quasi-identical results across several thousand BRF simulations. In all other cases, a weighted ensemble averaging technique (robust mean) was used to establish a reference dataset for the subsequent performance analysis.

Possible causes for the substantial dispersion among the simulation results could be 1) *operator choices* when transferring the prescribed test site properties to the needs and capabilities of a given RT model

(e.g., representing the realistic-looking trees from the RAMI website by cylindrical trunks and ellipsoidal hulls filled with scatterers), 2) *operator errors* during this scene creation process, and/or during the setup of the actual model run, and/or any eventual post-processing steps (e.g., specifying a wrong illumination azimuth angle or sensor location due to different angular conventions with respect to those on the RAMI website), and 3) software bugs and conceptual errors in the RT formulation of the model. The difference between operator choices and errors is that the former will leave the tree and/or canopy scale statistics unchanged, while the latter will not. Given that several of the participating RT models had been found to deliver accurate simulation results in the recent analysis of the RAMI-IV abstract canopy scenarios (Widlowski et al., 2013) one is left to conclude that – in absence of any subsequent changes to these model and/or processing schemes – it must be primarily the operator induced aspects that have caused the observed variability. For simulation methods that have not yet been evaluated against ensembles of third party models the possibility exists of course that their RT formulation itself may be faulty. A rerun of the various actual canopy scenes should help in clarifying these questions (as it did for the RAMI phases 1 to 3).

The dispersion among the various BRF, flux and THP simulations was already foretelling the outcome of any subsequent performance analysis. As such, the emphasis was not so much placed on a detailed analysis of model quality, but rather on presenting methods to test conformity with respect to permissible deviations from a reference. This topic is increasingly relevant due to the recurring need to demonstrate the fitness-for-purpose of quantitative data (e.g., whether obtained from field measurements, model simulations, retrieval algorithms, processing chains) in application-specific contexts. More specifically, both the shared and guarded risk approaches were introduced to highlight the role of uncertainties associated with both the reference and candidate methods. Truly, a dataset can only be eligible to act as reference solution in compliance testing if its associated uncertainty is but a small fraction of the width of the tolerance interval within which the candidate method must fall. Thus, whenever the width of the permissible tolerance interval is changed, one must evaluate the suitability of the reference solution anew.

The examples provided in this study can be used to place confidence on compliance declarations (as well as on the rejection) of items. Crucial to the outcome of any conformity testing, however, are the choices made a priori for the width of the permissible tolerance interval, the minimum required compliancy rate (or alternatively the maximum permissible non-compliance rate), as well as the confidence level required for making (non-)conformity declarations. Once these items have been defined, the uncertainty of potential reference datasets must be established in order to assess the eligibility of that dataset to act as reference in a subsequent conformity testing effort. Assessing the uncertainty of in situ measurement efforts for the validation of medium spatial resolution satellite products, for example, is not only costly, but also rather difficult. However, given the ability to simulate the radiative signatures of increasingly realistic terrestrial environments at both the local and domain-level scales, state-of-the-art three-dimensional RT models may be used to assess the magnitude of many of the contributing uncertainty components of current in situ efforts. This is then where intercomparison activities like RAMI come in again: To make sure that the quality of these simulation tools is well characterised and suitable for the requirements of the task at hand.

Acknowledgements

This study would not have been possible without the continuing support of A. Belward, head of the Land Resources Monitoring unit, at the Institute for Environment and Sustainability of the European Commission's Joint Research Centre in Ispra, Italy. A special thank you must also go to M. Buzica and M. Gerboles who clarified the first authors' understanding of both equivalence and compliance testing in the context of European ambient air quality directives. The technical support of M. Robustelli, as well as the contributions of J. Písek and T. Lükk for the test site characterizations, are also greatly appreciated. The field work at Järvselja was funded by the Estonian Ministry of Education and Research (target financing) as well as by Estonian Science Foundation grants.

References

Atzberger, C. (2000). Development of an invertible forest reflectance model: The INFOR-model. In M. Buchroithner (Ed.), *A decade of Trans-European remote sensing cooperation. Proceedings of the 20th EARSeL Symposium Dresden, Germany, 14–16 June 2000* (pp. 39–44). Leiden, The Netherlands: CRC Press/Balkema.

Biliouris, D., Verstraeten, W. W., Dutré, P., van Aardt, J. A. N., Muys, B., & Coppin, P. (2007). A Compact Laboratory Spectro-Goniometer (CLabSpeG) to assess the BRDF of materials. Presentation, calibration and implementation on *Fagus sylvatica* L. leaves. *Sensors*, *7*, 1846–1870.

Bojinski, S., Verstraete, M. M., Peterson, T. C., Richter, C., Simmons, A., & Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, *95*http://dx.doi.org/10.1175/BAMS-D-13-00047.1.

Bousquet, L., Lachérade, S., Jacquemoud, S., & Moya, I. (2005). Leaf BRDF measurements and model for specular and diffuse components differentiation. *Remote Sensing of Environment*, *98*, 201–211.

Brakke, T. W., Smith, J. A., & Harnden, J. M. (1989). Bidirectional scattering of light from tree leaves. *Remote Sensing of Environment*, *29*, 175–183.

Breece, H. T., & Holmes, R. A. (1971). Bidirectional scattering characteristics of healthy green soybean and corn leaves in vivo. *Applied Optics*, *10*, 119–127.

Bruegge, C. J., Chrien, N. L., Ando, R. R., Diner, D. J., Abdou, W. A., Helmlinger, M. C., et al. (2002). Early validation of the multi-angle imaging spectroradiometer (MISR) radiometric scale. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1477–1492.

Chelle, M. (2006). Could plant leaves be treated as Lambertian surfaces in dense crop canopies to estimate light absorption? *Ecological Modelling*, *198*(1–2), 219–228. http://dx.doi.org/10.1016/j.ecolmodel.2006.04.010.

Claverie, M., Vermote, E., Weiss, M., Baret, F., Hagolle, O., & Demarrez, V. (2013). Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. *Remote Sensing of Environment*, *139*, 216–230.

Côté, J. -F., Fournier, R. A., & Verstraete, M. M. (2015). Canopy architectural models in support of methods using hemispherical photography. In R. A. Fournier, & R. Hall (Eds.), *Hemispherical photography for forest science: Theory, methods and applications*. Springer (in press).

Disney, M. I., Lewis, P., Bouvet, M., Prieto-Blanco, A., & Hancock, S. (2009). Quantifying surface reflectivity for spaceborne lidar via two independent methods. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(10), 3262–3271. http://dx.doi.org/10.1109/TGRS.2009.2019268.

Essery, R., Bunting, P., Hardy, J., Link, T., Marks, D., Melloh, R., et al. (2008). Radiative transfer modeling of a coniferous canopy characterized by airborne remote sensing. *Journal of Hydrometeorology*, *9*(2), 228–241.

Gastellu-Etchegorry, J. -P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sensing of Environment*, *58*, 131–156.

Gastellu-Etchegorry, J. -P., Martin, E., & Gascon, F. (2004). Dart: A 3D model for simulating satellite images and studying surface radiation budget. *International Journal of Remote Sensing*, *25*, 73–96.

Gastellu-Etchegorry, J. P., Yin, T., Lauret, N., Cajgfinger, T., Gregoire, T., Lopes, M., et al. (2015). Discrete anisotropic radiative transfer (DART 5) for modelling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes. *Remote Sensing*, *7*(2), 1667–1701.

GCOS (2011). *Systematic observation requirements for satellite-based data products for climate (2011 update), supplemental details to the satellitebased component of the implementation plan for the global observing system for climate in support of the UNFCCC (2011 update). GCOS- 154*, Global Climate Observing System (GCOS), World Meteorological Organisation.

Govaerts, Y. (1995). *A model of light scattering in three-dimensional plant canopies: A Monte Carlo ray tracing approach.* (Ph. D. thesis) 2, Chemin du Cyclotron, B–1348. Louvain-la-Neuve, Belgique: Université Catholique de Louvain-la-Neuve, Département de Physique.

Govaerts, Y. M., Clerici, M., & Clerbaux, N. (2004). Operational calibration of the meteosat radiometer VIS band. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(9), 1900–1914.

Haverd, V., Lovell, J. L., Cuntz, M., Jupp, D. L. B., Newnham, G. J., & Sea, W. (2012). The Canopy Semi-analytic Pgap And Radiative Transfer (CanSPART) model: Formulation and application. *Agricultural and Forest Meteorology*, *160*, 14–35.

Helder, D., Thome, K. J., Mishra, N., Chander, G., Xiong, X., Angal, A., et al. (2013). Absolute radiometric calibration of landsat using a pseudo invariant calibration site. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(3), 1360–1369.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., et al. (1998). AERONET — A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1–16.

Huang, H., Chen, M., & Liu, Q. (2009). A realistic structure model for large-scale surface leaving radiance simulation of forest canopy and accuracy assessment. *International Journal of Remote Sensing*, *30*(20), 5421–5439.

ISO 13528 (2005). *Statistical methods for use in proficiency testing by interlaboratory comparisons.* International Standard, ISO 13528:2005(E)*ISO/TC6-SC6*. Geneva, Switzerland: International Organization for Standardization (66 pp.).

ISO-10576-1 (2003). *Statistical methods — Guidelines for the evaluation of conformity with specified requirements - Part 1: General principles.* ISO 10576–1:2003(E). Geneva, Switzerland: International Organisation for Standards.

ISO-14253-1 (1998). *Geometrical Product Specification (GPS) — Inspection by measurement of workpieces and measuring equipment.* Part 1: Decision rules for providing conformance or non-conformance with specifications, ISO 14253–1:1998(E). Geneva, Switzerland: International Organisation for Standards.

ISO-5725 (1994). *Accuracy (trueness and precision) of measurement methods and results (part 1–6 together with corrections).* Geneva, Switzerland: International Organisation for Standards.

JCGM-100 (2008). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM 1995 with minor corrections).* Joint Committee for Guides in Metrology, JCGM 100 (available from http://www.bipm.org/).

JCGM-101 (2008). *Evaluation of measurement data — Supplement 1 to the Guide to the expression of uncertainty in measurement.* Propagation of distributions using a Monte Carlo method, JCGM 101 (available from http://www.bipm.org/).

JCGM-106 (2012). *Evaluation of measurement data — The role of measurement uncertainty in conformity assessment, JCGM 106.* (available from http://www.bipm.org/).

Jonckheere, I., Nackaerts, K., Muys, B., & Coppin, P. (2005). Assessment of automatic gap fraction estimation of forests from digital hemispherical photography. *Agricultural and Forest Meteorology*, *132*, 96–114.

Jonckheere, I., Nackaerts, K., Muys, B., van Aardt, J., & Coppin, P. (2006). A fractal dimension-based modelling approach for studying the effect of leaf distribution on LAI retrieval in forest canopies. *Ecological Modelling*, *197*, 179–195.

Kaasalainen, S., & Rautiainen, M. (2007). Backscattering measurements from individual Scots pine needles. *Applied Optics*, *46*(22), 4916–4922.

Kobayashi, H., & Iwabuchi, H. (2008). A coupled 1-D atmosphere and 3-D canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. *Remote Sensing of Environment*, *112*, 173–185.

Kötz, B., Schaepman, M., Morsdorf, F., Bowyer, P., Itten, K., & Allgöwer, B. (2004). Radiative transfer modeling within a heterogeneous canopy for estimation of forest fire fuel properties. *Remote Sensing of Environment*, *92*, 332–344.

Kuusk, A., Kuusk, J., & Lang, M. (2009). A dataset for the validation of reflectance models. *Remote Sensing of Environment*, *113*, 889–892.

Kuusk, A., Kuusk, J., & Lang, M. (2014). Modeling directional forest reflectance with the hybrid type forest reflectance model FRT. *Remote Sensing of Environment*, *149*, 196–204.

Kuusk, A., Lang, M., & Kuusk, J. (2013). Database of optical and structural data for the validation of forest radiative transfer models. In Alexander A. Kokhanovsky (Ed.), *Radiative transfer and optical properties of atmosphere and underlying surface. Light Scattering Reviews*, *7*. (pp. 109–148). Springer.

Kuusk, A., & Nilson, T. (2000). A directional multispectral forest reflectance model. *Remote Sensing of Environment*, *72*, 244–252.

Kuusk, A., Nilson, T., Kuusk, J., & Lang, M. (2010). Reflectance spectra of RAMI forest stands in Estonia: Simulations and measurements. *Remote Sensing of Environment*, *114*(12), 2962–2969.

Kuusk, A., Nilson, T., Paas, M., Lang, M., & Kuusk, J. (2008). Validation of the forest radiative transfer model FRT. *Remote Sensing of Environment*, *112*(1), 51–58.

Lavergne, T., Kaminski, T., Pinty, B., Taberner, M., Gobron, N., Verstraete, M. M., et al. (2007). Application to MISR land products of an RPV model inversion package using adjoint and Hessian codes. *Remote Sensing of Environment*, *107*(1–2), 362–375.

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the botanical plant modelling system. *Agron. Agric. Environment*, *19*, 185–210.

Lintermann, B., & Deussen, O. (1999). Interactive modelling of plants. *IEEE Computer Graphics and Applications*, *19*(1), 2–11.

Liu, Q., Huang, H., Qin, W., Fu, K., & Li, X. (2007). An extended 3-D radiosity graphics combined model for studying thermal-emission directionality of crop canopy. *IEEE Transactions on Geoscience and Remote Sensing*, *45*, 2900–2918.

Lolli, L., Pisani, M., Rajteri, M., Widlowski, J. -L., Bialek, A., Greenwell, C., et al. (2014). PHYTOS: A portable goniometer for in situ spectro-directional measurements of leaves. *Metrologia*, *51*, 309.

Lovell, J. L., Haverd, V., Jupp, D. L. B., & Newnham, G. J. (2012). The Canopy Semi-analytic Pgap And Radiative Transfer (CanSPART) model: Validation using ground based lidar. *Agricultural and Forest Meteorology*, *158–159*, 1–12.

McCorkel, J., Thome, K., & Lockwood, R. B. (2013). Absolute radiometric calibration of narrow-swath imaging sensors with reference to non-coincident wide-swath sensors. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(3), 1309–1318.

Milton, E. J., Schaepman, M. E., Anderson, K., Kneubühler, M., & Fox, N. (2007). Progress in field spectroscopy. *Remote Sensing of Environment*, *113*, 92–109.

Morsdorf, F., Kötz, B., Meier, E., Itten, K. I., & Allgöwer, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, *104*(1), 50–61.

Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., & Allgöwer, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, *92*(3), 353–362.

Mõttus, M., Sulev, M., & Lang, M. (2006). Estimation of crown volume for a geometric radiation model from detailed measurements of tree structure. *Ecological Modelling*, *198*(4–5), 506–514.

Ni-Meister, W., Yang, W., & Kiang, N. Y. (2010). A clumped-foliage canopy radiative transfer model for a global dynamic terrestrial ecosystem model I: Theory. *Agricultural Forest Meteorology*, *150*, 881–894.

Pendrill, L. R. (2014). Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*, *51*, 206–218.

Pharr, M., & Humphreys, G. (2010). *Physically based rendering: From theory to implementation.* San Francisco: Morgan Kaufmann (1167 pp.).

Pinty, B., Gobron, N., Widlowski, J. -L., Gerstl, S. A. W., Verstraete, M. M., Antunes, M., et al. (2001). The RAdiation transfer Model Intercomparison (RAMI) exercise. *Journal of Geophysical Research*, *106*, 11937–11956.

Pinty, B., Gobron, N., Widlowski, J. -L., Lavergne, T., & Verstraete, M. M. (2004). Synergy between 1-D and 3-D radiation transfer models to retrieve vegetation canopy properties from remote sensing data. *Journal of Geophysical Research*, *109*, D21205. http://dx.doi.org/10.1029/2004JD005214.

Pinty, B., Widlowski, J. -L., Taberner, M., Gobron, N., Verstraete, M. M., Disney, M., et al. (2004). RAdiation transfer Model Intercomparison (RAMI) exercise: Results from the second phase. *Journal of Geophysical Research*, *109*, D06210.

Pisek, J., Lang, M., Nilson, T., Korhonen, L., & Karu, H. (2011). Comparison of methods for measuring gap size distribution and canopy nonrandomness at Järvselja RAMI (RAdiation transfer Model Intercomparison) test sites. *Agricultural and Forest Meteorology*, *151*(3), 365–377.

Rahman, H., Pinty, B., & Verstraete, M. M. (1993). Coupled surface–atmosphere reflectance (CSAR) model. 2. Semiempirical surface model usable with NOAA Advanced Very High Resolution Radiometer data. *Journal of Geophysical Research*, *98*, 20,791–20,801.

Rahman, H., Verstraete, M. M., & Pinty, B. (1993). Coupled surface–atmosphere reflectance (CSAR) model. 1. Model description and inversion on synthetic data. *Journal of Geophysical Research*, *98*, 20,779–20,789.

Sandmeier, S. R. (2000). Acquisition of bidirectional reflectance factor data with field goniometers. *Remote Sensing of Environment*, *73*, 257–269.

Schlerf, M., & Atzberger, M. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, *100*, 281–294.

Schneider, F. D., Leiterer, R., Morsdorf, F., Gastellu-Etchegorry, J. -P., Lauret, N., Pfeifer, N., et al. (2014). Simulating imaging spectrometer data: 3D forest modeling based on LiDAR and in situ data. *Remote Sensing of Environment*, *152*, 235–250.

Smolander, S., & Stenberg, P. (2003). A method to account for shoot scale clumping in coniferous canopy reflectance models. *Remote Sensing of Environment*, *88*, 363–373.

Somers, B., Delalieux, S., Verstraeten, W. W., & Coppin, P. (2009). A conceptual framework for the simultaneous extraction of sub-pixel spatial extent and spectral characteristics of crops. *Photogrammetric Engineering and Remote Sensing*, *75*, 57–68.

Sommer, K. D., & Kochsiek, M. (2002). *Role of measurement uncertainty in deciding conformance in legal metrology. OIML Bulletin, vol. XLIII (2).* (pp. 19–24), 19–24.

Stenberg, P., Palmroth, S., Bond, B. J., Sprugel, D. G., & Smolander, H. (2001). Shoot structure and photosynthetic efficiency along the light gradient in a Scots pine canopy. *Tree Physiology*, *21*, 805–814.

Stuckens, J., Somers, B., Delalieux, S., Verstraeten, W. W., & Coppin, P. (2009). The impact of common assumptions on canopy radiative transfer simulations: A casestudy in citrus orchards. *Journal of Quantitative Spectroscopy and Radiative Transfer*http://dx.doi.org/10.1016/j.jqsrt.2008.09.001.

Thome, K. J., Arai, K., Tsuchida, S., & Biggar, S. F. (2008). Vicarious calibration of ASTER via the reflectance-based approach. *IEEE Transactions on Geoscience and Remote Sensing*, *46*, 3285–3295.

Wang, Y., Czapla-Myers, J., Lyapustin, A., Thome, K., & Dutton, E. (2011). Aeronet-based surface reflectance validation network (ASRVN) data evaluation: Case study for railroad valley calibration site. *Remote Sensing Of Environment*, *115*, 2710–2717. http://dx.doi.org/10.1016/j.rse.2011.06.011.

Weber, J., & Penn, J. (1995). Creation and rendering of realistic trees. *SIGGRAPH '95, Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (pp. 119–128). New York, NY, USA: ACM0-89791-701-4 (©1995).

Widlowsk, J. -L., Pinty, B., Clerici, M., Dai, Y., De Kauwe, M., de Ridder, K., et al. (2011). RAMI4PILPS: An intercomparison of formulations for the partitioning of solar radiation in land surface models. *Journal of Geophysical Research*, *116*(G02019), 25. http://dx.doi.org/10.1029/2010JG001511.

Widlowski, J. -L. (2010). On the bias of instantaneous FAPAR estimates in open-canopy forests. *Agricultural and Forest Meteorology*, *150*, 1501–1522.

Widlowski, J. -L. (2015). Conformity testing in view of regulatory uses of satellite-derived quantitative surface variables. *Environmental Science and Policy*, *51*, 149–169.

Widlowski, J. -L., Lavergne, T., Pinty, B., Verstraete, M. M., & Gobron, N. (2006). Rayspread: A virtual laboratory for rapid BRF simulations over 3-D plant canopies. In G. Frank (Ed.), *Computational Methods in Transport. Lecture Notes in Computational Science and Engineering Series*, *48*. (pp. 122–123). Berlin: Springer Verlag (10 3–540–28).

Widlowski, J. -L., Pinty, B., Lopatka, M., Atzberger, C., Buzica, D., Chelle, M., et al. (2013). The 4th Radiative Transfer Model Intercomparison (RAMI-IV): Proficiency testing of canopy reflectance models with ISO-13528. *Journal of Geophysical Research — Atmospheres*, *118*http://dx.doi.org/10.1002/jgrd.5049.

Widlowski, J. -L., Robustelli, M., Disney, M., Gastellu-Etchegorry, J. -P., Lavergne, T., Lewis, P., et al. (2007). The RAMI On-line Model Checker (ROMC): A web-based benchmarking facility for canopy reflectance models. *Remote Sensing of Environment*, *112*(3), 1144–1150.

Widlowski, J. -L., Taberner, M., Pinty, B., Bruniquel-Pinel, V., Disney, M., Fernandes, R., et al. (2007). The third RAdiation transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance modeling. *Journal of Geophysical Research*, *112*, D0911. http://dx.doi.org/10.1029/2006JD007821.

Yang, W., Ni-Meister, W., Kiang, N. Y., Moorcraft, P., Strahler, A. H., & Oliphant, A. (2010). A clumped-foliage canopy radiative transfer model for a global dynamic terrestrial ecosystem model II: Validation. *Agricultural Forest Meteorology*, *150*, 895–907.

Zenone, T., Migliavacca, M., Montagnani, L., Seufert, G., & Valentini, R. (2008). Carbon sequestration in short rotation forestry and traditional poplar plantation. *Proceedings of the Short Rotation Crops International Conference, Minneapolis, USA, August 18–22* (http://www.cinram.umn.edu/srwc/cs1es.html).

Zhao, F., Gu, X., Verhoef, W., Wang, Q., Yu, T., Liu, Q., et al. (2010). A spectral directional reflectance model of row crops. *Remote Sensing of Environment*, *114*(2), 265–285.