The RAMI On-line Model Checker (ROMC): A web-based benchmarking facility for canopy reflectance models

J.-L. Widlowski [a,*], M. Robustelli [a], M. Disney [b,c], J.-P. Gastellu-Etchegorry [d], T. Lavergne [e], P. Lewis [b,c], P.R.J. North [f], B. Pinty [a], R. Thompson [g], M.M. Verstraete [a]

[a] Joint Research Center, European Commission, Via E. Fermi, 1, 21020 Ispra (VA), Italy
[b] Department of Geography, University College London, 26 Bedford way, London, WC1H 0AP, UK
[c] NERC Centre for Terrestrial Carbon Dynamics, UK
[d] Centre d'Etudes Spatiales de la Biosphère (CESBIO), 18 avenue Edouard Belin, bpi 2801, 31401 Toulouse cedex 9, France
[e] Section for Remote Sensing, Research and Development Department, Norwegian Meteorological Institut, P.O. Box 43, Blindern, N-0313 Oslo, Norway
[f] Climate and Land-Surface Systems Interaction Centre, Department of Geography, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, UK
[g] Alachua Research Institute, Alachua, Florida, USA

Abstract

The exploitation of global Earth Observation data hinges increasingly on physically-based radiative transfer (RT) models. These models simulate the interactions of solar radiation within a given medium (e.g., clouds, plant canopies) and are used to generate look-up-tables that are embedded into quantitative retrieval algorithms, such as those delivering the operational surface products for MODIS, MISR and MERIS. An assessment of the quality of canopy RT models thus appears essential if accurate and reliable information is to be derived from them. Until recently such an undertaking was a time consuming and labour intensive process that was made even more challenging by the general lack of absolute reference standards. Several years of benchmarking activities in the frame of the RAdiation transfer Model Intercomparison (RAMI) exercise have now led to the development of the RAMI On-line Model Checker (ROMC). The ROMC is a web-based tool allowing model developers and users to autonomously assess the performance of canopy RT models (http://romc.jrc.ec.europa.eu/). Access to the ROMC is free and enables users to obtain both statistical and graphical indications as to the performance of their canopy RT model. In addition to providing an overall indication of the skill of a given model to correctly match the reference data, the ROMC allows also for interactive comparison/evaluations of different model versions/submissions of a given user. All ROMC graphs can be downloaded in PostScript format and come with a reference number for easy usage in presentations and publications. It is hoped that the ROMC will prove useful for the RT modeling community as a whole, not only by providing a convenient means to evaluate models outside the triennial phases of RAMI but also to attract participation in future RAMI activities.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Benchmarking; Radiative transfer; Model intercomparison; BRDF; Plant canopy

1. Introduction

Indicators of the quality of physically-based radiative transfer (RT) models are of relevance to their developers and users, as well as the scientists, space agencies and policy makers that use or support the products and information derived on the basis of such model simulations/inversions. The RAdiation transfer Model Intercomparison (RAMI) initiative was launched in an attempt to shed light on the reliability and accuracy of existing canopy radiation transfer models (http://rami-benchmark.jrc.ec.europa.eu/). RAMI is a triennial community exercise that encourages the systematic evaluation of canopy reflectance models under well-controlled experimental conditions and on a voluntary basis (Pinty et al., 2001, 2004; Widlowski et al., 2007). Significant efforts were being made during the first (1999) and second (2002) phases of RAMI in order to document and reduce the dispersion between increasing numbers of participating models, but it was not until the completion of the third phase (2005) that a sufficiently strong consensus emerged among RT models capable of simulating the entire palette of RAMI test cases: from simple plane-parallel turbid medium scenarios to complex heterogeneous

---

vegetation canopies with and without underlying topography. Section 2 will document the generation of a reference data set from six state-of-the-art 3-D Monte Carlo models identified during RAMI-3. Section 3 then describes the capabilities of the RAMI On-Line Model Checker (ROMC) to compare RT model simulations against this "surrogate truth" via the Internet.

2. The ROMC reference data set

The self-consistency (e.g., energy conservation) together with the absolute and relative performance of RT models were evaluated in great detail during RAMI-3 (Widlowski et al., 2007). Substantial improvements in model agreement were observed, in particular, for the 3-D Monte Carlo (MC) models that participated. These models allow for explicit 3-D representations of complex canopy architectures and avoid unnecessary assumptions and approximations in the solving of the radiative transfer equation due to their stochastic sampling of the surface leaving radiation field (Disney et al., 2000). During RAMI-3, for example, it was shown that – in the few cases where analytical solutions were available – the bi-directional reflectance factor (BRF) simulations of MC models were capable of matching the predicted values to within 10^{-4} (Widlowski et al., 2007). The general lack of absolute reference criteria, however, prevents such an evaluation strategy for most of the conceivable canopy architectures. Instead, Pinty et al. (2001, 2004) proposed to compare the output from individual RT models to ensemble averages computed from simulation results of other RT models. Relative model comparison, when repeated over large ensembles of test cases, may thus discover systematic trends and biases in the performance of a model and thereby may help to diagnose the underlying cause(s) of a model's divergence. RAMI thus follows the argumentation of Oreskes (1994) who maintains that the complete validation of computer simulation models is quite impossible and that any such endeavour should focus instead on the finding of flaws in a model's behaviour. This is because conformity with the anticipated outcome is not proof of a model's physical correctness since different sources of errors may compensate each other and lead to apparently correct results.

Among the various relative model comparison metrics proposed during RAMI the average model-to-ensemble dispersion $\bar{\delta}_m$ (in percent) may be of most relevance here. It can be computed for any given model (m) by comparing its simulations of some radiative quantity (X) to those generated by a set of other models (c) for any given number of spectral (λ), structural (ζ), viewing (v), and illumination (i) conditions:

$$\bar{\delta}_m = \frac{200}{\bar{N}} \sum_{\lambda=1}^{N_\lambda} \sum_{\zeta=1}^{N_\zeta} \sum_{v=1}^{N_{\Omega v}} \sum_{i=1}^{N_{\Omega i}} \sum_{c=1; c \neq m}^{N_c} \left| \frac{X_m(\lambda, \zeta, v, i) - X_c(\lambda, \zeta, v, i)}{X_m(\lambda, \zeta, v, i) + X_c(\lambda, \zeta, v, i)} \right|$$

where N_c is the number of models with which the output of model m is to be compared, and $\bar{N}$ is the total number of simulations of model m that are included in the computation of $\bar{\delta}_m$.

Table 1 shows the average model-to-ensemble dispersion $\bar{\delta}_m$ [%] obtained from total BRF simulations made by six of the 3-D Monte Carlo models participating in RAMI-2 and/or RAMI-3. These six models are labelled: DART (Gastellu-Etchegorry et al.,

Table 1
Model-to-ensemble dispersion statistics $\bar{\delta}_m$ [%] for total BRF simulations submitted by six 3-D Monte Carlo models in RAMI-2 and/or RAMI-3

Model name	Discrete scenes		Turbid scenes	
	RAMI-2	RAMI-3	RAMI-2	RAMI-3
DART	–	–	1.42	1.46
drat	1.92	0.55	–	–
FLIGHT	1.26	0.97	9.63	1.06
Rayspread	–	0.55	–	0.64
raytran	1.31	0.60	1.06	0.69
Sprint3	1.29	1.01	9.66	0.69

$\bar{\delta}_m$ was computed from all available structural, spectral, illumination and viewing conditions for a given model m.

2004), drat (Lewis, 1999), FLIGHT (North, 1996), Rayspread (Widlowski et al., 2006), raytran (Govaerts & Verstraete, 1998) and Sprint3 (Goel & Thompson, 2000). Whereas the $\bar{\delta}_m$ values of DART remained very similar between the second (RAMI-2) and third (RAMI-3) phases of RAMI, the $\bar{\delta}_m$ values of the remaining five models all improved during that same period, meaning that a smaller dispersion now exists between the latest version of these models. The weighted average dispersion between all six 3-D MC models in Table 1 was found to have almost halved from RAMI-2 (1.37%) to RAMI-3 (0.72%) for test cases with finite sized scatterers (discrete scenes), and in the case of infinitesimally small scatterers (turbid scenes) it improved by a factor of ~7 from RAMI-2 (6.36%) to RAMI-3 (0.91%). As such the mutual agreement between these six 3-D MC models is significantly lower than anything previously achieved and definitely below the absolute accuracy of current space borne measurements. Together with the performance of the MC models in the self-consistency and absolute model evaluation tests of RAMI-3 (Widlowski et al., 2007) the figures in Table 1 thus support the usage of these RT tools to establish a "surrogate truth" dataset against which other/future radiative transfer models can be compared.

Due to variable model participation and performance (see Fig. 17 in Widlowski et al., 2007) the precise number and names of the available 3-D MC models could, however, change from one experiment and measurement to the next. In order to obtain a "surrogate truth" estimate under these conditions the simulation (X) obtained from a set of $N_{3\text{-D}}^{\text{credible}}$ "credible" 3-D Monte Carlo models were simply averaged for any given spectral (λ), structural (ζ), viewing (v), and illumination (i) conditions:

$$X_{\text{ref}}(\lambda, \zeta, v, i) = \frac{1}{N_{3-D}^{\text{credible}}} \sum_{n=1}^{N_{3-D}^{\text{credible}}} X_{3-D}^{\text{credible}}(\lambda, \zeta, v, i; n)$$

where the precise number and names of the 3-D Monte Carlo models that feature within $N_{3\text{-D}}^{\text{credible}}$ are selected from among the DART, drat, FLIGHT, Rayspread, raytran, and Sprint3 models for every RAMI experiment and measurement type individually. To perform this model selection the following list of criteria was applied:

1. For every RAMI BRF (flux) measurement, identify at least two (one) 3-D Monte Carlo models that do not belong to the same RT modelling school/family,

2. If two (or more) models from the same RT modelling school/family are available, e.g., Rayspread and raytran, choose the one with the least amount of apparent Monte Carlo noise,

3. Remove all those 3-D Monte Carlo models from the reference set that are noticeably different from the main cluster of 3-D MC simulations,

4. If sufficient models are contained in the main cluster of 3-D MC simulations then remove those models that would introduce noticeable levels of "MC noise" into the reference set,

5. If there are two distinct clusters of 3-D Monte Carlo models, or, no obvious cluster at all, then use all available 3-D RT models to define a reference solution.

A drawback of this selection procedure lies in the fact that the ROMC reference dataset may not be fully compliant with energy conservation. This is a direct consequence of 1) the fact that not all "credible" models performed the complete set of RAMI test cases and measurement types, and 2) the varying quality with which the six 3-D Monte Carlo models simulated the 11 RAMI measurements
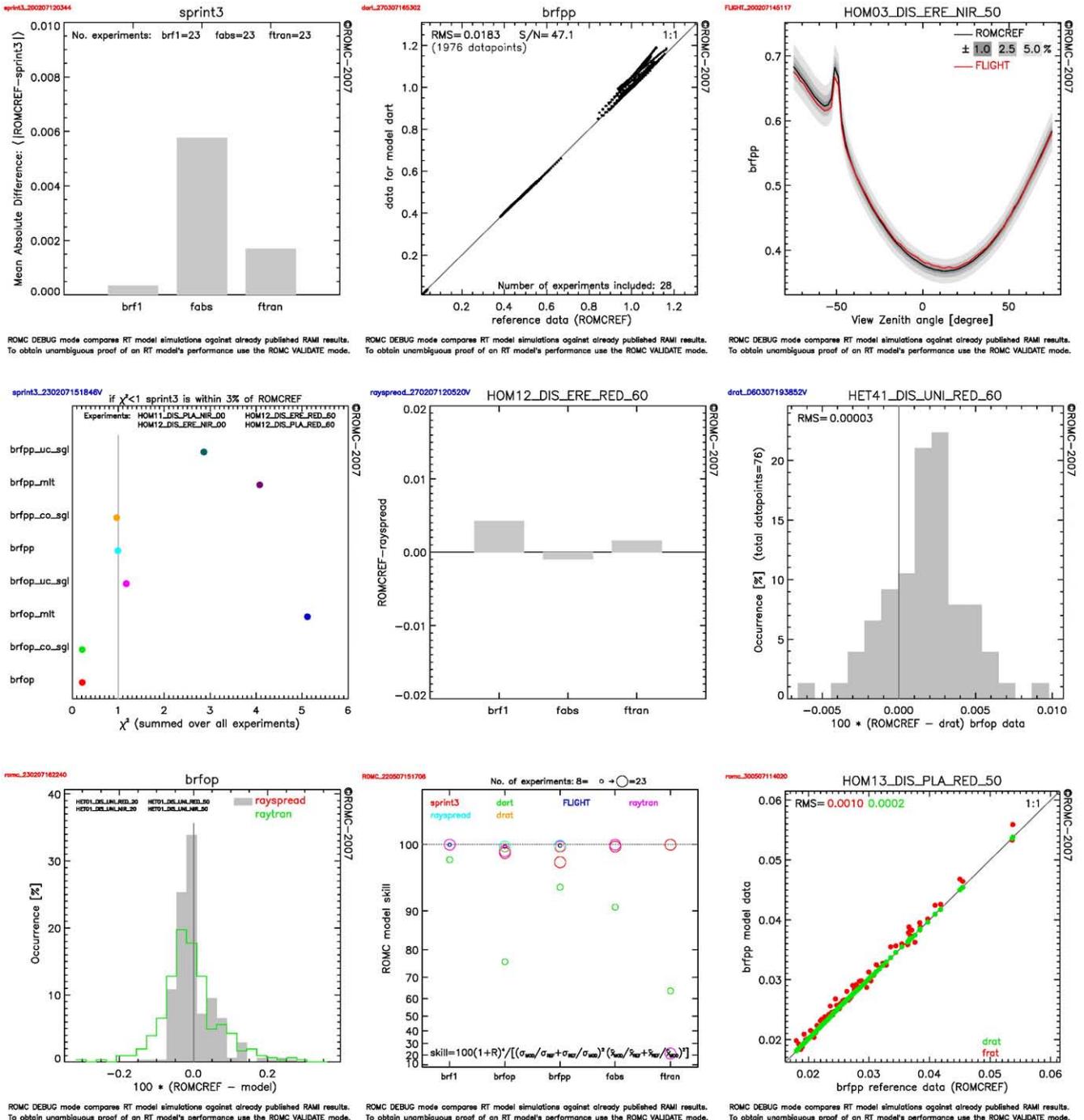

Fig. 1. Examples of downloadable ROMC graphs showing model performance for individual experiments or ensembles of experiments carried out in debug mode (top panels), validate mode (middle panels), and, using the interactive capabilities of the ROMC (bottom panels).

types under different structural, spectral, viewing and illumination conditions. Nevertheless, the average absolute deviation of the ROMC reference dataset from energy conservation was found to be 0.00025 in the red, 0.00007 in the near-infrared, and $9.5 \cdot 10^{-7}$ for conservative scattering conditions (purist corner). A synoptic table featuring the names and performances of the 3-D Monte Carlo models that contributed toward the computation of the "surrogate truth" is available for each RAMI-3 experiment and measurement type on the following webpage: http://romc.jrc.ec.europa.eu/WWW/PAGES/ROMC_Home/RAMIREF.php.

3. The Rami On-line Model Checker

With this valuable reference dataset at hand, it is possible to allow canopy RT model owners, developers and customers to evaluate the performance of a given model even outside the frame of a RAMI phase. To facilitate such an undertaking the RAMI On-line Model Checker (ROMC) was developed at the Joint Research Centre of the European Commission in Ispra, Italy. The ROMC is a web-based interface allowing for the on-line evaluation of RT models using as reference the "surrogate truth" derived from among the six credible 3-D Monte Carlo models identified during RAMI-3. Access to the ROMC can be obtained either via the RAMI website or directly using the URL http://romc.jrc.ec.europa.eu/. After providing a username and valid email address, authenticated users can register up to three different canopy reflectance model names. To evaluate the quality of these models the ROMC can be utilized in two different ways: 1) in "debug mode", which allows to repeatedly compare the output of a RT model to the "surrogate truth" of one or more experiments and/or measurements from RAMI-3, i.e., the simulation results are known since they are available on the RAMI website, and 2) in "validate mode", which enables the once-only testing of the RT model against a randomly selected set of test cases that are similar but not quite equivalent to those from RAMI, i.e., the solutions are not known *a priori* and the model runs cannot be repeated.

- In *debug mode* users may choose to execute one particular experiment and/or measurement from the set of RAMI-3 test cases *ad infinitum*, or, at least until they are satisfied with the performance of their model. Detailed descriptions of the structural, spectral, illumination and measurement conditions are available for each test case. Once the model simulation results are generated, users can upload these via the ROMC web-interface, and – provided that they adhere to the amply described RAMI file naming and formatting conventions – this process will result in a series of graphical files being made available for all test cases. In debug mode users may not only download these ROMC graphs but also ASCII files containing the actual "surrogate truth" data for the submitted test cases.
- In *validate mode* users may choose between structurally homogeneous and/or heterogeneous "floating spheres" canopies to verify the performance of their RT model. These test cases are similar but not quite identical to those featuring in RAMI-3. The actual test cases will be drawn at random from a large list of possible ones, such that it is unlikely to obtain the same test case twice, i.e., in all likelihood one will not "know"

the solution *a priori*. Here too, the "surrogate truth" was derived from simulations generated by models belonging to the set of "credible" 3-D MC models. In validate mode, however, the reference data will not be available for downloading. The procedure for data submission, on the other hand, is identical to that of the debug mode, and – provided that all RAMI formatting and file naming requirements were applied – will also lead to a results page featuring a variety of intercomparison graphics.

Users may download their ROMC results either as jpeg formatted images from the ROMC website, or else, opt for receiving them via email in PostScript form. Both the debug and validate mode ROMC graphs feature a reference number and many ROMC graphs are available in colour as well as in black and white. In general, ROMC graphs have different appearances depending on the context in which they have been generated. In debug mode, for example, the ROMC reference number has a smaller font size than in validate mode. Furthermore, all debug mode graphs feature a two-line footer informing readers that: *ROMC DEBUG mode compares RT model simulations against already published RAMI results. To obtain unambiguous proof of an RT model's performance use the ROMC VALIDATE mode.* ROMC users that have submitted model simulation results for different models or multiple versions of the same model may explore the ROMC's "interactive" graph plotting capabilities. Fig. 1 provides a series of examples of ROMC graphs generated in debug mode (top row), validate mode (middle row), and using the "interactive" ROMC capabilities when analysing simulation results generated in debug mode (bottom row). In these graphs the "surrogate truth" data set is referred to as 'ROMCREF' where applicable. The list of currently available ROMC graph types include:

- *Data plots*, featuring both the model and the reference BRFs along the principal or orthogonal plane. Three grey-coloured envelopes corresponding to 1, 2.5 and 5% of the BRF of the "surrogate truth" solution surround the reference solution. For an example of such a data plot featuring BRFs in the principal plane see the right panel in the top row of Fig. 1.
- *1 to 1 plots*, where the model (X_{usr}) data is plotted against the reference (X_{ref}) data. These graphs also feature the number of data points N, the root-mean-square (RMS) error:

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left[X_{ref}(n) - X_{usr}(n) \right]^2}$$

and the signal-to-noise ratio (S/R), defined as:

$$\text{S/R} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^{N} \left[X_{ref}(n) \right]^2}}{\text{RMS}}$$

In the case of BRF quantities these graphs are available both for individual test cases and for all submitted test cases combined. In the case of flux quantities the '1 to 1' graphs always include simulation results from multiple test cases. The middle panel of the top row in Fig. 1 provides an example of a '1 to 1' graph with

data points from a large ensemble of ROMC test cases, whereas the bottom right panel of Fig. 1 displays a '1 to 1' graph for BRF simulations generated by two models (belonging to the same user) when applied to a single test case.

- *Histograms*, showing the distribution of the differences between reference and model BRFs. These graphs are generated for individual test cases as well as for all submitted test cases together, and include information on the RMS error and the number of contributing experiments, or, data points. For examples see the right panel in the middle row of Fig. 1 (where a histogram of BRF differences for a single test case is shown), and the left panel in the bottom row of Fig. 1 (where the histograms of two RT models are shown that performed four different test cases).

- χ^2 *graphs*, quantifying the deviations of a model from the "surrogate truth" using the following metric:

$$\chi^2 = \frac{1}{N-1} \sum_{\lambda=1}^{N_\lambda} \sum_{\zeta=1}^{N_\zeta} \sum_{v=1}^{N_{\Omega v}} \sum_{i=1}^{N_{\Omega i}} \frac{[X_{\mathrm{usr}}(\lambda, \zeta, v, i) - X_{\mathrm{ref}}(\lambda, \zeta, v, i)]^2}{\sigma^2(\lambda, \zeta, v, i)}$$

where $N = N_\lambda + N_\zeta + N_{\Omega v} + N_{\Omega i}$ is the total number of simulations covering a variety of spectral λ, structural ζ, viewing v and illumination i conditions (if applicable), and, where the associated uncertainty $\sigma(\lambda, \zeta, v, i) = f \cdot X_{\mathrm{ref}}(\lambda, \zeta, v, i)$ corresponds to a fraction f of the "surrogate truth" obtained from the "credible" 3-D Monte Carlo models. By default $f = 0.03$ which reflects the absolute calibration accuracy of current space borne instruments like MISR (Bruegge et al., 2002) and MERIS (Kneubühler et al., 2002), among others (see the left panel in the middle row of Fig. 1 for an example of a χ^2 graph).

- *Bar charts*, depicting the deviation of the model and reference fluxes. For individual test cases absolute differences $(X_{\mathrm{ref}} - X_{\mathrm{usr}})$ are documented, whereas, when multiple test cases are dealt with together then the average absolute deviation, i.e., $(1/N)\Sigma_1^N |X_{\mathrm{ref}}(n) - X_{\mathrm{usr}}(n)|$, is reported. Examples of this graph type can be seen in the central (individual test case) and top left (multiple test cases) panels of Fig. 1.

- *Skill plots*, displaying the ability of the model to match the pattern (expressed by the correlation R), the variability (expressed by the ratio of standard deviations: $\sigma_{\mathrm{usr}}/\sigma_{\mathrm{ref}}$) and mean amplitude of the reference data (expressed by the ratio of dataset means: $\overline{X}_{\mathrm{usr}}/\overline{X}_{\mathrm{ref}}$). Within ROMC the skill of a model is defined as:

$$\mathrm{Skill} = 100 \frac{(1+R)^4}{\left(\frac{\overline{x}_{\mathrm{usr}}}{\overline{x}_{\mathrm{ref}}} + \frac{\overline{x}_{\mathrm{ref}}}{\overline{x}_{\mathrm{usr}}}\right)^2 \left(\frac{\sigma_{\mathrm{usr}}}{\sigma_{\mathrm{ref}}} + \frac{\sigma_{\mathrm{ref}}}{\sigma_{\mathrm{usr}}}\right)^2}$$

Here σ is the standard deviation and $\overline{X}$ is the mean of the user's ($_{\mathrm{usr}}$) or "surrogate truth"' ($_{\mathrm{ref}}$) data. R is the correlation between the two data sets having N data points each:

$$R = \frac{\frac{1}{N} \sum_{n=1}^{N} [X_{\mathrm{usr}}(n) - \overline{X}_{\mathrm{usr}}] \cdot [X_{\mathrm{ref}}(n) - \overline{X}_{\mathrm{ref}}]}{\sigma_{\mathrm{usr}} \cdot \sigma_{\mathrm{ref}}}$$

The ROMC skill scores are defined from zero (least skillful) to 100 (most skillful) assuming that a perfect match is possible. A perfectly negative correlation ($R = -1$) will lead to the minimum

Skill value (Skill=0) reflecting the importance of accurate 'shape', or pattern generation in the context of BRF simulations. Data sets that are completely 'flat' (i.e., Lambertian surfaces) or have a mean of zero (i.e., absorption in conservative scattering cases) are not supported. The middle panel in the bottom row of Fig. 1 shows the skills of the six 3-D Monte Carlo models (colours) when performing various measurement types over structurally homogeneous leaf canopies.

- *Taylor diagrams*, providing a concise statistical summary of how well patterns match each other in terms of their correlation (R), their root-mean-square difference (ΔRMS), and the ratio of their variances $\sigma_{\mathrm{usr}}^2/\sigma_{\mathrm{ref}}^2$ (Taylor, 2006). This is achieved by separating the overall RMS error (see above) into a contribution arising from the mean bias $\overline{RMS} = \overline{X}_{\mathrm{usr}} - \overline{X}_{\mathrm{ref}}$ and a contribution due to the root-mean-square differences between the data sets (ΔRMS):

$$\mathrm{RMS}^2 = \overline{RMS}^2 + \Delta\mathrm{RMS}^2$$

such that the root-mean-square difference can be written as:

$$\Delta\mathrm{RMS}^2 = \sigma_{\mathrm{usr}}^2 + \sigma_{\mathrm{ref}}^2 - 2\sigma_{\mathrm{usr}}\sigma_{\mathrm{ref}}R$$

where the relation of ΔRMS to the correlation coefficient (R), and the standard deviations of both the user (σ_{usr}) and reference (σ_{ref}) data sets is similar to that expressed by the law of cosines for the relationship between two adjacent sides of a triangle together with the angle between them, and the third (opposite) side of the triangle. This relationship then allows to construct a two-dimensional diagram where three different information items can be read from any point (Taylor, 2006).

Fig. 2 provides two examples of ROMC generated Taylor diagrams in debug mode. The left panel displays the performance of the raytran model when simulating a variety of BRF measurement types for a single test case, whereas the right panel shows how the DART model simulates the single-collided BRF component for a variety of structurally homogeneous test cases. Rather than using the actual variances of the data sets, Fig. 2 normalises ΔRMS in the above equation by the variance of the reference data set σ_{ref}. As such the reference solution will always be located at unit distance along the abscissa (big diamond symbol). The closer the simulations of a model are to the diamond symbol the better their matching with the "surrogate truth" data set. More specifically, the radial distances from the origin to the center of the symbols identifying the user's model (i.e., discs in the left panel, circles in the right panel) are proportional to the normalised-standard-deviation (that is, the ratio of the standard deviation of the user's data to that of the reference data). Model simulations that are located outside the dotted circular arc – indicating unit normalised-standard-deviation – feature more variability than the ROMC's "surrogate truth" data set. An example of this are the multiple scattering BRF components (blue/mauve data points) in the left panel of Fig. 2 that are due to the inherent Monte Carlo noise in raytran simulations ($\sim 0.8 \cdot 10^{-4}$) and the small magnitude of the reference BRFs ($6 \cdot 10^{-4}$–$8 \cdot 10^{-4}$). Conversely, if the data points are located within the unit circular arc then the model simulations exhibit less
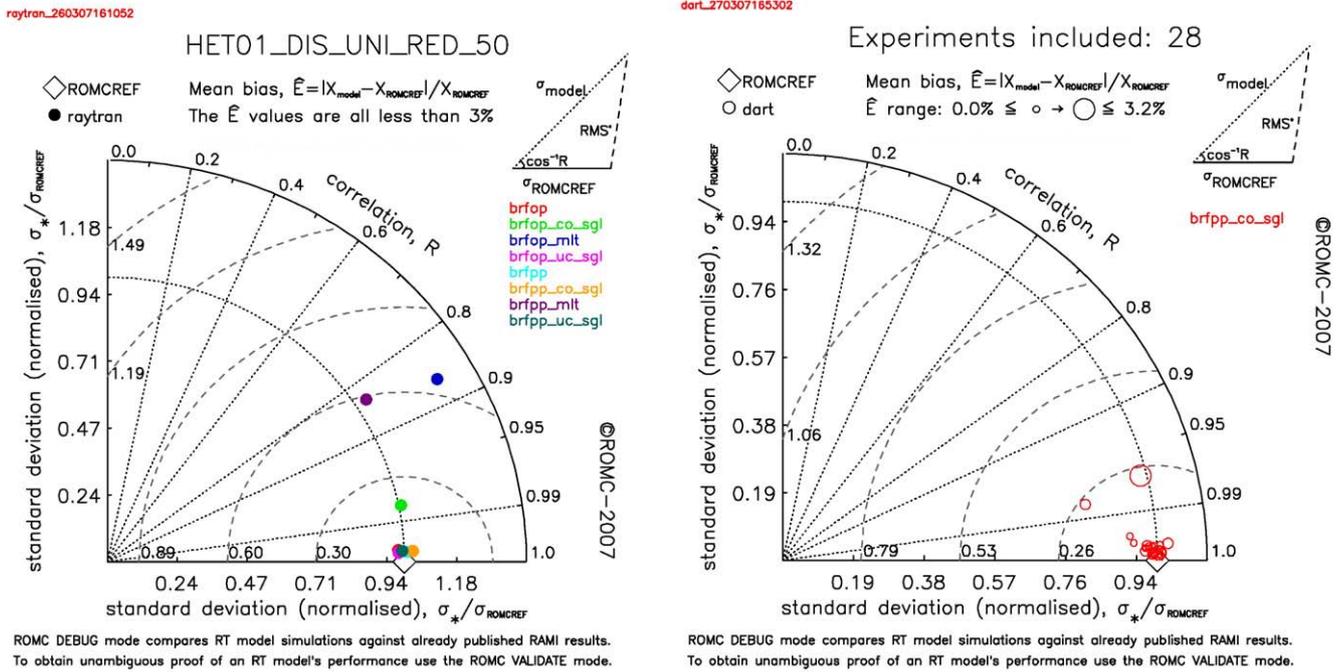
Fig. 2. Examples of Taylor diagrams in debug mode depicting model performance in terms of the ability to match the pattern (correlation R) and variability (normalised standard deviation) of the reference solution. The normalised RMS difference (dashed circular arcs) and the mean bias (symbol size) are also indicated. The left panel features results for different measurement types and one ROMC test case, whereas the right panel contains results for one single measurement type and several ROMC test cases.

variability than the reference data set (an example of this is the leftmost red circle in the right panel of Fig. 2). The cosine of the elevation angle of the line connecting the origin to the data point indicates the correlation, R between the model simulated and the "surrogate truth" data sets (the labels of the correlation coefficient are given along the outer circular rim of the Taylor diagram). One can see that the azimuthal distance of the ROMC reference data (big diamond) corresponds to a perfect correlation ($R=1$). Finally, the dashed circular lines that originate from the (big diamond) symbol of the reference data correspond to the normalised RMS difference between the two data sets. For the test case displayed in the left panel of Fig. 2, one can thus see that, with the exception of the multiple collided BRF component, most of the simulations of the raytran model fall very close to the (big diamond) of the "surrogate truth" data both in terms of the correlation and normalised RMS difference. Since Taylor diagrams do not differentiate between two data sets that differ only by a constant offset, however, the ROMC varies the size of the plotting symbol as soon as $\overline{RMS}$ exceeds 3% of the reference solution (and no negative correlations occur). This can be seen in the right panel of Fig. 2 where the DART generated single-collided BRF data exhibits on one occasion a mean bias of 3.2% with respect to the "surrogate truth" solution (largest red circle).

Users of ROMC are encouraged to utilise only ROMC results that were obtained in validate mode for publication purposes. Those obtained in debug mode, obviously, can not qualify as unambiguous proof of model performance since 1) all simulation results may readily be viewed on the RAMI website, and 2) the "surrogate truth" data itself is freely available. It should also be noted that it is not permissible to modify, change or edit the results

provided by the ROMC. This applies to ROMC graphs, statistics, as well as "surrogate truth" data which must all be used 'as provided' and duly referenced as described on the "Frequently Asked Questions" (FAQ) section of the ROMC website.

It is important to realize that the ROMC is primarily intended to provide rapid and indicative evaluations of the performance of RT models, in particular in validate mode. RAMI, on the other hand, approaches model evaluation in a more comprehensive manner covering many different aspects of model performance under a large variety of structural, spectral and illumination conditions. Use of the ROMC should thus be seen as a first step toward the full participation in a next phase of RAMI.

4. Conclusion

The RAMI On-line Model Checker (ROMC), a web-based tool for the autonomous benchmarking of canopy radiation transfer models was presented (http://romc.jrc.ec.europa.eu/). The ROMC was made possible due to substantial improvements in model agreement observed during the third phase of the RAdiation transfer Model Intercomparison (RAMI) exercise. The ROMC utilizes a reference data set established after several years of efforts of the international modeling community to identify a series of "credible" 3-D Monte Carlo models. Usage of the ROMC is simple and allows to distinguish between quality assessments intended to 1) repeatedly "debug" a model against test cases from previous phases of RAMI, and 2) enable a "validation" of a model under randomly chosen spectral, structural and illumination conditions. Correctly formatted RT model simulations of the selected/assigned test cases have to be

uploaded via the ROMC web-interface and result in a series of graphs that document the closeness of the model simulations to the "surrogate truth" data set of the ROMC. All ROMC graphs can be received in PostScript format for easy inclusion in presentations and publications. It is expected that the ROMC will evolve into a yardstick for scientists, funding agencies, and policy makers alike when it comes to appreciating the quality of a given model and ultimately also the information that it helps retrieving from Earth Observation data.

Acknowledgements

References

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 Multiangle Imaging Spectro-Radiometer (MISR) radiometric scale. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1477−1492.

Disney, M. I., Lewis, P., & North, P. R. J. (2000). Monte Carlo ray-tracing in optical canopy reflectance modeling. *Remote Sensing Reviews*, *18*, 163−196.

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.

Goel, N. S., & Thompson, R. L. (2000). A snapshot of canopy reflectance models and a universal model for the radiation regime. *Remote Sensing Reviews*, *18*, 197−206.

Govaerts, Y., & Verstraete, M. M. (1998). Raytran: A Monte Carlo ray tracing model to compute light scattering in threedimensional heterogeneous media. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 493−505.

Kneubühler, M., Schaepman, M., Thome, K., Baret, F., & Müller, A. (2002). Calibration and validation of Envisat MERIS. Part 1: vicarious calibration at Rail Road valley Playa (NV). *Proceedings of MERIS level 2 validation workshop, ESRIN, Frascati, Italy, December 9–13*.

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the botanical plant modelling system. *Agronomie — Agriculture and Environment*, *19*, 185−210.

North, P. R. J. (1996). Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Transactions on Geoscience and Remote Sensing*, *34*, 946−956.

Oreskes, N. K. (1994). Verification, validation and confirmation of numerical models in the earth sciences. *Science*, *263*, 641−646.

Pinty, B., Gobron, N., Widlowski, J. -L., Gerstl, S. A. W., Verstraete, M. M., Antunes, M., et al. (2001). Radiation Transfer Model Intercomparison (RAMI) exercise. *Journal of Geophysical Research*, *106*, 11,937−11,956.

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. doi:10.1029/2003JD004252

Taylor, K. E. (2006). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183−7192.

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 F. Graziani (Ed.), *Computational methods in transport. Lecture notes in computational science and engineering series*, *Vol. 48* (pp. 211−231). Berlin: Springer Verlag. ISBN.10 3.540.28.122.3.

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 modelling. *Journal of Geophysical Research*, *112*. doi:10.1029/2006JD007821