University of Wollongong

Research Online

Faculty of Science, Medicine and Health -
Papers: part A

Faculty of Science, Medicine and Health

1-1-2014

Minimizing measurement uncertainties of coniferous needle-leaf optical properties. Part II: experimental setup and error analysis

Lucia Yanez-Rausell
Wageningen University

Zbynek Malenovsky
University of Tasmania, zbynek@uow.edu.au

Jan G. P. W Clevers
Wageningen University

Michael E. Schaepman
University of Zurich

Follow this and additional works at: <https://ro.uow.edu.au/smhpapers>

Part of the [Medicine and Health Sciences Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Yanez-Rausell, Lucia; Malenovsky, Zbynek; Clevers, Jan G. P. W; and Schaepman, Michael E., "Minimizing measurement uncertainties of coniferous needle-leaf optical properties. Part II: experimental setup and error analysis" (2014). *Faculty of Science, Medicine and Health - Papers: part A*. 2498.
<https://ro.uow.edu.au/smhpapers/2498>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Minimizing measurement uncertainties of coniferous needle-leaf optical properties. Part II: experimental setup and error analysis

Abstract

We present uncertainties associated with the measurement of coniferous needle-leaf optical properties (OPs) with an integrating sphere using an optimized gap-fraction (GF) correction method, where GF refers to the air gaps appearing between the needles of a measured sample. We used an optically stable artificial material simulating needle leaves to investigate the potential effects of: 1) the sample holder carrying the needles during measurements and 2) multiple scattering in between the measured needles. Our optimization of integrating sphere port configurations using the sample holder showed an underestimation of the needle transmittance signal of at least 2% in flat needles and 4% in nonflat needles. If the needles have a nonflat cross section, multiple scattering of the photons during the GF measurement led to a GF overestimation. In addition, the multiple scattering of photons during the optical measurements caused less accurate performance of the GF-correction algorithms, which are based on the assumption of linear relationship between the nonGF-corrected signal and increasing GF, resulting in transmittance overestimation of nonflat needle samples. Overall, the final deviation achieved after optimizing the method is about 1% in reflectance and 6% in transmittance if the needles are flat, and if they are nonflat, the error increases to 4%-6% in reflectance and 10%-12% in transmittance. These results suggest that formulae for measurements and computation of coniferous needle OPs require modification that includes also the phenomenon of multiple scattering between the measured needles.

Keywords

Conifers, gap fraction (GF), integrating sphere, leaf, needles, optical properties (OPs), reflectance, transmittance

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Yanez-Rausell, L., Malenovsky, Z., Clevers, J. G. P. W. & Schaepman, M. E. (2014). Minimizing measurement uncertainties of coniferous needle-leaf optical properties. Part II: experimental setup and error analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7 (2), 406-420.

Minimizing Measurement Uncertainties of Coniferous Needle-Leaf Optical Properties, Part II: Experimental set-up and error analysis

Lucia Yáñez-Rausell, Zbyněk Malenovský *Member, IEEE*, Jan G. P. W. Clevers, and Michael E. Schaepman, *Senior Member, IEEE*

Abstract—We present uncertainties associated with measurement of coniferous needle-leaf optical properties (*OPs*) with an integrating sphere using an optimized gap-fraction (*GF*) correction method, where *GF* refers to the air gaps appearing between needles of a measured sample. We used an optically stable artificial material simulating needle leaves to investigate the potential effects of: (a) the sample holder carrying the needles during measurements, and (b) multiple scattering in between the measured needles. Our optimization of integrating sphere port configurations using the sample holder showed an underestimation of the needle transmittance signal of at least 2% in flat needles and 4% in non-flat needles. If the needles have a non-flat cross-section, multiple scattering of the photons during the *GF* measurement led to a *GF* overestimation. In addition, multiple scattering of the photons during the optical measurements induced an exponential increase of the recorded signal with increasing *GF*, instead of a linear increase supported by current *GF*-correction algorithms. Thus, non-flat needle samples are affected by overestimating needle transmittance, which cannot be corrected through standard algorithms. Overall, the final deviation achieved after optimizing the method is about 1% in reflectance and 6% in transmittance if the needles are flat and, if they are non-flat, the error increases to 4-6% in reflectance and 10-12% in transmittance. These results suggest that formulas for measurements and computation of coniferous needle *OPs* require modification that includes also the phenomenon of multiple scattering between measured needles.

Index Terms—Needles, optical properties, reflectance, transmittance, integrating sphere, leaf, conifers, gap fraction

I. INTRODUCTION

RECENT methods for measuring narrow leaf optical properties (*OPs*), with special attention on non-flat nonbifacial coniferous needle-leaves (e.g., Norway spruce needles), have been reviewed (this issue). Based on the outcomes of this review, we propose an experimental set-up optimizing the ‘Mesarch *et al.*’ needle-leaf *OPs* measurement approach [1]. The proposed experiment is addressing the following objectives: (1) to investigate a potential influence of the sample holder presence on the measured leaf reflectance (*R*) and transmittance (*T*); and (2) to investigate the effect of varying gap fraction and multiple scattering between neighbor needles, focusing on: (a) the influence of the needle cross-section shape and (b) the distance between the needles in the sample. In case of a needle cross-section influence we are hypothesising that higher occurrence of small illumination incident angles, caused by a circular or rhomboidal needle cross-section shape, increases multiple scattering between the measured needles. In other words, the photons hitting the needle surface in a direction different from the normal to the needle surface have higher probability to interact with needles in their near neighbourhood [2]. Simultaneously, we hypothesize that an increasing distance between needle sample elements (larger air gaps) decreases the probability of multiple scattering in between them [1, 3].

After analyzing the experimental results we outline the recommendations for the best practice ensuring reliable measurements of coniferous needle *OPs*.

Manuscript received in February, 06, 2013. This work was supported by the European Community’s Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER- I-NET) and by the UZH Research Priority Programme on ‘Global Change

L. Yáñez-Rausell is with the Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, PO BOX 47, 6700 AA Wageningen, The Netherlands (e-mail: l.yanez@wur.nl) and the Department of Geography, Univ. of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland (e-mail: lucia.yanezrausell@geo.uzh.ch).

Z. Malenovsky was with Department of Geography, Univ. of Zurich, Switzerland. He is now with the School of Geography & Environmental Studies, University of Tasmania, Tasmania, Australia (e-mail: zbynek.malenovsky@gmail.com).

J. G. P. W. Clevers is with the Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Wageningen, The Netherlands (e-mail: jan.clevers@wur.nl).

M. E. Schaepman is with the Department of Geography, Univ. of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland (e-mail: michael.schaepman@geo.uzh.ch).

II. MATERIALS AND METHODS

A. Artificial Needle Leaves

To carry out our analyses we used artificial needle leaves of 1 mm width that were cut-off from two types of material of known R and T . Contrary to real leaves, both materials were optically stable over time, i.e. temporally nondegrading (at least during the experiments) ensuring that measured R and T of the same material would result in similar material uncertainty. Also, assuming that the R and T were inherent properties of the material itself, the OPs obtained from the artificial needle samples were fully comparable with OPs measured on uncut ‘broad-leaf-like’ pieces of the same material. Since our goal was to estimate the errors of the measuring technique itself, not reproducing exactly the spectral signatures of real needles is not affecting the conclusions of this study.

The first selected material was a green-colored plastic (0.1 mm thick) with OPs similar to the photographic film used by ‘Mesarch *et al.*’ [1]. This material simulated what we call ‘flat narrow leaves’ (e.g. geometrical shape similar to grass, mesquite leaflets, etc.). The second material was a green silicon mat (1.0 mm thick) that simulated what we call ‘non-flat narrow needle leaves’ (i.e. leaves of many coniferous species). The silicon was chosen due to a suitable transmittance (up to 50% below 800 nm) and a thickness comparable to the real coniferous needles, e.g. Norway spruce (*Picea abies* (L.) Karst.) needles. ‘Broad-leaf-like’ pieces of each material (*uncut*) were measured and used as a reference.

B. Measurement Protocol for Narrow-Leaf OPs

For our analysis we followed the five steps as summarized in the ‘Mesarch *et al.*’ [1] approach for measuring narrow-leaf R and T . The OPs were measured using a spectroradiometer (ASD FieldSpec 3) coupled to a portable single-beam ASD leaf-integrating sphere (ASD 190 RTS-3ZC; Fig. 1(b)). During the measurements, the flat plastic and silicon needles were placed in a sample holder similar to the one described in [4] that was specially machined to fit the integrating sphere. The holder consists of two 1-mm-thick optically flat (black-painted) metallic plates and a central aperture larger than the sample port (16.5 mm in diameter) (Fig. 1(a)). The holder shape fits firmly to the sphere sample ports, ensuring consistent OPs measurements from both sides (Fig. 1(a), and 1(b)). A laser pointer located above the sphere light source assembly ensured no misplacement of the lamp assembly during measurements. To prevent residual light leaks, the integrating sphere was covered with a black, low reflecting cloth during all optical measurements. A light tunnel of the length equal to the diameter of the ASD integrating sphere was introduced for T measurements to ensure that the same sample area of comparable size is being illuminated and measured during both R and T readings. Masks mimicking the shape and size of the sample illumination area were built from a black-painted paper. Digital images of the masked sample holder aperture (i.e., the area presenting the needles during the optical measurements) were acquired with a double-lamp scanner (EPSON Perfection TM 4490 PHOTO) and stored in 8-bit gray-scale format. During the sample scanning, masks were precisely aligned and fixed to the sample holder plates, and these were positioned using references previously marked on the scanner window in order to minimize misplacements. All these steps contributed to the optimization of the optical measurements and scanning protocol and improved repeatability. We used the images to estimate the sample GF , defined as the ratio of the total gap area between needles to the total measurement area. The total number of gap pixels in the masked image was calculated by applying a ‘white-pixel-threshold’ to discriminate gap pixels from needle pixels [1] using the image processing software GIMP 2.6, GNU. We will refer to this computed GF as $Digital_{GF}$. The measurement area, i.e. size of the illuminating beam, was slightly different in R and T modes (9 and 7 mm in diameter, respectively), which required one R and one T mask and resulted in two $Digital_{GF}$ values per sample.

The measured spectra and $Digital_{GF}$ were introduced in the GF -correction formulae [1] to compute the individual-needle-leaf directional-hemispherical R_{needle} (1) and T_{needle} (2) per sample per spectral waveband as in

$$R_{needle} = \frac{R_{TOTAL}}{1 - GF} \quad (1)$$

and

$$T_{needle} = \frac{(T_{TOTAL} - R_w \cdot GF)}{1 - GF}, \quad (2)$$

where R_{needle} is the R of individual needles, T_{needle} is the T through individual needles, and R_w is the R of the integrating sphere wall (assumed to be close to 1, i.e. 100%). The R_{TOTAL} and T_{TOTAL} variables are sample R and T , respectively, computed as the recorded total reflected and transmitted radiation, including the photons lost or added by the air gaps, but corrected for the stray light and normalized to the reflectance of a white reference panel (cf., Appendix I).

A summary of the experimental set-up followed to achieve our objectives is presented in Fig. 2. For simplicity, from now on we will refer to ‘Mesarch *et al.*’ [1] as *Mesarch*, to avoid repeating continuously the same reference.

C. Effect of the Sample Holder

Prior to the needle-leaf *OPs* measurements, we measured the reference R ($R_{reference}$) and T ($T_{reference}$) signal of 10 samples of uncut pieces from both artificial materials: flat plastic and silicon. The sample sizes were bigger than the measurement area (i.e. size of the illuminating beam; Fig. 3(a), and 3(d)) and for simplicity we will refer to them as to *broad-leaf* samples. Their *OPs* were measured without the special sample holder, following the standard leaf measurement protocol recommended by the integrating sphere manufacturer.

To test the effect of the sample holder, each *broad-leaf* sample was subsequently placed between the sample holder plates (*SH*) and then its R ($R_{uncutSH}$) and T ($T_{uncutSH}$) were measured using four sample holder scenarios. These scenarios consisted of modified configurations of the integrating sphere ports (Table I). In the first scenario the sample holder was used only for holding the sample at the sample port. In the three remaining scenarios, however, the stray light (*STR*) or ‘white reference’ (*REF*) measurements also involved placing the sample holder at the corresponding port, i.e. an empty sample holder was placed at the sample port or in front of the white reference while acquiring *STR* or *REF* measurements (Table I and Appendix I).

The root mean square errors per scenario from the resulting averaged $R_{uncutSH}$ and $T_{uncutSH}$ were then computed for each material by using (3):

$$RMSEsp_{SH}(i) = \sqrt{\frac{\sum_{\lambda=1}^m (sp_{uncutSH}(\lambda, i) - sp_{reference}(\lambda))^2}{m}}, \quad (3)$$

where $sp_{uncutSH}(\lambda, i)$ is the mean $R_{uncutSH}$ or $T_{uncutSH}$ of 10 samples per scenario at wavelength λ for one of the two materials; i refers to the sample holder scenario number ($i=1, \dots, 4$); and $sp_{reference}(\lambda)$ is the corresponding mean $R_{reference}$ or $T_{reference}$ of 10 samples at the same wavelength λ and for the same material. Wavelength λ varies from 450-1700 nm. The spectral range below 450 and above 1700 nm was removed due to an insufficient signal-to-noise ratio caused by the spectroradiometer and integrating sphere.

Finally, for each material, the scenario corresponding to the minimal error was selected as the optimal measuring set-up and used for *OPs* measurements of needle-samples.

D. Effect of the Multiple Scattering Between Neighbor Needles

To test the two hypotheses concerning the impact of multiple scattering of light, we built nine needle-sample scenarios with artificial needles. A needle-sample is composed of several needles built from a specific needle cross-section type, which are placed parallel to each other at a specific distance inside the sample holder plates. The scenarios, called *F1*, *F2*, *F3*, *S1*, *S2*, *S3*, *Rh1*, *Rh2*, and *Rh3*, were built by combining three cross-section types (flat-*F*, squared-*S* or rhomboidal-*Rh*) and three air-gap distances, (index 1 = 0.5 mm (e.g. *F1*); 2 = 1.0 mm; and 3 = 1.5 mm). The flat cross section (*F*) corresponds to flat plastic needles (Figs. 3(b), 3(e), and 3(h)); the squared cross-section (*S*) corresponds to silicon needles positioned inside the sample holder with 2 needle-sides laying on the sample holder plates and parallel to them (Figs. 3(b), 3(e), and 3(h)); and the rhomboidal cross-section (*Rh*) corresponds to silicon needles positioned inside the sample holder with no needle-sides parallel to the sample holder plates (Figs. 3(c), 3(f), and 3(i)). Two sewing needles (commercial steel dressmaker pins of size No. 12, 19 mm in length, and 0.5 mm in diameter) were pushed through the upper and lower end of each artificial needle to ensure the desired alignment of silicon needles in the sample. Achieving the aimed gap distances between the sample needles required positioning them very carefully using previously marked references on the sample holder plates. However, manual handling of the small needle elements is extremely difficult and small misplacements are practically unavoidable. Consequently, the real achieved gap distances between needles in a sample can differ slightly from the theoretical values aimed for each scenario (0.5, 1.0 or 1.5 mm). We will refer to this error as the *best-effort-handling deviation* (*BEHD*). Also, the selection of the gap distances was driven by the *BEHD*, as it was not feasible to place needles at a distance smaller than half their width with a sufficient accuracy. After several trials, the smallest gap distance considered was 0.5 mm, i.e., half the needle width (the half-width of *Rh* cross-section needle is slightly larger, being = 0.7 mm). Our hypothesis regarding the *BEHD* is that it increases for the same needle cross-section scenario (*F*, *S* or *Rh*) with decreasing gap-distance (starting from 1.5 to 1.0 to 0.5 mm) and due to the handling difficulty it is bigger for *Rh* than for *F* and *S* for the same gap-distance scenarios. We opted for building our analysis on the gap-distance between needles instead of on the *GF* size used by *Mesarch* because computing the sample *GF* (i.e. the ratio of the total gap area between needles to the total measurement area) is, first, not visually straightforward (i.e. requires computing the area of polygon-shaped air gaps intersected by a circular light beam (Figs. 3(b) and 3(c)) and second, sample-dependent (each needle size is different when using *in-vivo* coniferous needles). This means that, once the needles are detached from the shoot and placed inside the sample holder, several trial and error realignments and *GF*-computations per sample are needed to approximate the desired *GF*. For real needles, where the foliar tissue degradation starts several minutes after detachment from the shoot, this adjustment procedure might result in degraded biochemistry and structure [5].

We prepared 10 samples for each of the nine scenarios, and measured their *OPs* reproducing the best sample-holder scenario

resulting from the sample holder effect analysis. In all cases the spectral range below 450 and above 1700 nm was again removed, because of large noise.

After the *OPs* measurements, three *GF* values were computed for each *R* and *T* mode: *Ideal_{GF}*, *Digital_{GF}* and *True_{GF}*. The *Ideal_{GF}R* and *Ideal_{GF}T* are theoretical *GFs* computed per scenario by using simple trigonometry based on the known size of the illuminated area (represented by the *R* or *T* mask), the needles and the air-gaps corresponding to each scenario. Due to the *BEHD*, the real sample *GF* differs slightly from the *Ideal_{GF}*. To compute the *Digital_{GF}R* and *Digital_{GF}T*, each sample was masked, scanned, and digitally processed (section B). We identified the optimal scan-settings (resolution (*r*), brightness (*b*) and contrast (*c*)) and the ‘white-pixel-threshold’ value (*t*), required to discriminate the air gap and needle pixels in the digital image (cf., Appendix II).

Finally, based on *Mesarch’s* definition of the ‘true’ *GF* we computed the *True_{GF}T* per sample of each scenario. This computation consists in substituting the *GF*-corrected *T* spectrum of an individual needle, T_{needle} (2), by its corresponding *broad-leaf* ‘true’ (‘nongap’) *T* spectrum, and in extracting the *GF* value from the equation. The T_{needle} is assumed to be equal to $T_{uncutSH}$ since, after the *GF* correction, both quantities should represent the inherent optical property of the measured material [1]. This way the *True_{GF}T* can be extracted through (4):

$$True_{GF}T(\lambda) = \frac{T_{TOTAL}(\lambda) - T_{uncutSH}(\lambda)}{1 - T_{uncutSH}(\lambda)}. \quad (4)$$

The same strategy is applied in case of *R* to compute the *True_{GF}R*. The *GF* is extracted from *Mesarch’s* R_{needle} formula (1) as (5):

$$True_{GF}R(\lambda) = \frac{R_{TOTAL}(\lambda) - R_{uncutSH}(\lambda)}{R_{uncutSH}(\lambda)}. \quad (5)$$

To neutralize the sample holder effect affecting T_{needle} , in (4) we used the $T_{uncutSH}$ resulting from the best sample holder scenario (section II.C) instead of the *broad-leaf* $T_{reference}$. Using *T* spectra measured under the same sample holder scenario ensures that the reference is equally affected by the same holder effect. The same applies for $R_{uncutSH}$ in (5). Since thickness of a rhomboidal needle is not exactly the same as thickness of a flat silicon *broad-leaf* (‘nongap’) sample, the concept of *True_{GF}* applied to the rhomboidal cross-section samples might be slightly biased. Mean volumes of a squared cross-section needle sample and a rhomboidal one are, however, equivalent, thus we assume that absorption of a rhomboidal needle (A_{needle}) is comparable to the absorption of a flat silicon *broad-leaf* ($A_{uncutSH}$) one, especially at wavelengths with prevailing light scattering and low absorbance. According to [1, 3], the three-dimensional profile of non-flat needle cross-section (e.g. circular, semicircular, or rhomboidal) increases the probability of photon multiple scattering between measured elements especially if the needles are close to each other (i.e. at small gap-distance). The scattered light can escape from or be introduced into the integrating sphere during measurements, subtracting or adding a certain amount of photons to the recorded optical signals. This effect is not taken into account in *Mesarch’s* formulae (cf., (1) and (2)), since the fraction of incoming light passing through the sample air gaps, for which the signal has to be corrected, is calculated based on a two dimensional solution (i.e. the gap size was computed by subtracting the sample needle-projected area from the total measured (illuminated) area). Therefore, one can expect the multiple scattering effects to influence the *True_{GF}* values computed from *Mesarch’s* formulae. If the needle transmittance computed using the *GF*-correction is overestimated, then by using (2) and (4) the relationship $T_{needle} > T_{uncutSH}$ can be expressed as:

$$\frac{T_{TOTAL}(\lambda) - GF}{1 - GF} > \frac{T_{TOTAL}(\lambda) - True_{GF}T(\lambda)}{1 - True_{GF}T(\lambda)}, \quad (6)$$

where *GF* refers to the ‘real’ sample *GF*. Thus the ‘real’ sample *GF* is smaller than *True_{GF}T*. The opposite occurs if T_{needle} is underestimated, i.e. the ‘real’ sample *GF* is larger than the computed *True_{GF}T*. Following the same rationale, if the needle reflectance computed using the *GF*-correction is overestimated, i.e. $R_{needle} > R_{uncutSH}$, the ‘real’ sample *GF* is based on (1) and (5) larger than *True_{GF}R*. The contrary applies if R_{needle} is underestimated.

Our hypotheses related to the effects of multiple scattering on *Mesarch’s* method focused on the influence of two factors: (a) the needle cross-section shape, and (b) the distance between needles in the sample. Regarding (a), we expect that for the same gap-distance (e.g. *F2*, *S2*, and *Rh2*), the deviation caused by the multiple scattering effects will increase from flat (*F*) to squared (*S*) to rhomboidal (*Rh*) cross-section types, due to differences in the light incident angles and in the volume scattering occurring in non-flat cross-section scenarios *S* and *Rh* (Figs. 3h) and 3(i)), i.e. the subsurface scattering inside the needles [6]. Concerning (b), we hypothesize that for the same cross-section scenarios (e.g. *S1*, *S2*, and *S3*), the deviation caused by the multiple scattering effects will increase with decreasing gap-distance (from 1.5 to 1.0 to 0.5 mm). This is based on the assumption that photons hitting a needle at angles different than the normal to the needle surface are more likely to re-interact with neighbor needles (e.g.

Rh cross-section, Fig. 3(i) compared to *F* cross-section, Fig. 3(h)), especially if the needles are closer to each other as in small gap-distance scenarios. To test these hypotheses we computed the deviation of the $Digital_{GF}$ and the $True_{GF}$ from the theoretical $Ideal_{GF}$ using (7) and (8):

$$RMSE_{gf_{DI}}(i) = \sqrt{\frac{\sum_{s=1}^n (Digital_{GF}(s,i) - Ideal_{GF}(i))^2}{n}} \quad (7)$$

and

$$RMSE_{gf_{TI}}(i) = \sqrt{\frac{\sum_{\lambda=1}^m \sum_{s=1}^n (True_{GF}(\lambda,s,i) - Ideal_{GF}(i))^2}{m \cdot n}}, \quad (8)$$

where s refers to the sample number ($s=1, \dots, 10$), i to the scenario number ($i=1, \dots, 9$) and λ to the particular wavelength (varying from 450 -1700 nm). The differences in (a) and (b) are expected to affect the sharpness of the needle edges in the scanned digital image and subsequently the output $Digital_{GF}$ computed from this image. Therefore, (7) deals with the effect of the light scattered during the sample scanning. The $Digital_{GF}$ values used in (7) corresponded to the optimized scanning and processing settings resulting from the sensitivity analysis (Appendix II). Equation (8) focuses on the effect of the needles multiple scattering during the sample spectral measurements. Ideally, the $True_{GF}$ value should be equal to the ‘real’ sample GF . However, as explained above, factors (a) and (b) are expected to influence the scattering behavior of the incoming photons and cause over-/under-estimation of the $True_{GF}$ values extracted from *Mesarch’s* formulae (cf., (1) and (2)). Both $RMSE_{gf_{DI}}$ and $RMSE_{gf_{TI}}$ are affected by the *BEHD*. Despite the *BEHD*, the theoretical $Ideal_{GF}$, which is computed from the fixed dimension of the artificial needle element and air-gaps established per scenario, is the closest reference to the ‘real’ sample GF available. Additionally to the $RMSE$ computations, a paired Student t-test on the probability level $\alpha = 0.05$ was applied to test significant difference between $Ideal_{GF}$, $Digital_{GF}$, and $True_{GF}$ per scenario (i.e., difference between *R* and *T* mode) and also between scenarios.

Since dimensions of ‘real’ needle leaves vary, GF correction of ‘real’ narrow leaves measured with *Mesarch’s* method can rely only on the $Digital_{GF}$ values. Thus, to test our hypotheses we also computed the $True_{GF}$ deviation from the corresponding $Digital_{GF}$ for both *R* and *T* (9):

$$RMSE_{gf_{TD}}(i) = \sqrt{\frac{\sum_{\lambda=1}^m \sum_{s=1}^n (True_{GF}(\lambda,s,i) - Digital_{GF}(s,i))^2}{m \cdot n}} \quad (9)$$

Equation (9) gathers both effects considered in (7) and (8) and neutralizes the *BEHD*, as the *BEHD* of the same sample does not change.

Finally, we expect that the amount of photons affected by the multiple scattering during the *R* measurements is for a given sample the same as during the *T* measurements, because the needle elements are not re-aligned between both measurements. Flipping the sample holder when switching from the *R* to the *T* measurement mode ensures that the same side of the sample is always facing the light source. This, however, does not mean that the multiple scattering has the same over-/under-estimating effect on the final *R* and *T* spectra [1].

E. Validation of OPs after GF Correction

The individual needle-leaf directional-hemispherical *R* (R_{needle}) and *T* (T_{needle}) (for terminology see [7]) per sample per spectral waveband was computed using *Mesarch’s* formulae for GF -correction ((1) and (2)). The $Digital_{GF}$ used for the correction are the values corresponding to the optimized scanning and processing settings. The resulting R_{needle} and T_{needle} spectra were compared to the corresponding reference through (10):

$$RMSE_{sp_{needle}}(i, ref) = \sqrt{\frac{\sum_{\lambda=1}^m \sum_{s=1}^n (sp_{needle}(\lambda,s,i) - sp_{broadleaf}(\lambda,i,ref))^2}{m \cdot n}}, \quad (10)$$

where sp_{needle} is the R_{needle} or T_{needle} ; $sp_{broadleaf}$ is the average *broad-leaf* reference; i is the scenario number ($i=1, \dots, 9$); s is the sample number ($s=1, \dots, 10$); and λ is the wavelength in the range 450-1700 nm. The $sp_{broadleaf}$ depends on the *broad-leaf* reference type ('*ref*'): (a) the *broad-leaf* R or T for the best sample holder scenario ($R_{uncutSH}$ or $T_{uncutSH}$); or (b) the *broad-leaf* R or T measured without sample holder ($R_{reference}$ and $T_{reference}$). In (b) the RMSE in (10) comprised all potential error sources in the method including the sample holder effect, the $Digital_{GF}$ estimation effect after optimizing the scanning and image processing settings, and the multiple scattering effect between neighbor needles. In (a), the reference $sp_{broadleaf}$ and needle spectra sp_{needle} suffer from the same potential sample-holder effects and thus this effect is not included in the output RMSE. Statistical difference between R_{needle} or T_{needle} and the corresponding $sp_{broadleaf}$ was tested through a paired Student t-test on the probability level $\alpha = 0.05$.

In addition, we computed sp_{needle} ignoring the GF correction, i.e., using the standard formulae suggested by the sphere manufacturer for broad (uncut) leaves. The sp_{needle} is, therefore, the signal before the GF correction, i.e., the sample R_{TOTAL} in (1) and T_{TOTAL} in (2) formed by both the needle elements and the air gaps and normalized to the white reference panel. The relationship between these non- GF -corrected reflectance and transmittance signals and the sample GF is expected to be linear, given no error in estimation of GF and no interactions between sample elements and incident beam [1]. However, our multiple scattering hypothesis assumes that light interactions between the needle elements should affect this relationship in a non-linear way. To verify this expectation, a function fitting analysis was applied per scenario to assess the nature of the relationship.

III. RESULTS AND DISCUSSION

A. Effect of the Sample Holder

First, the effect of introducing a special sample holder for narrow leaves is presented in terms of RMSE based on (3). Results showed that, although the sample holder was never hit by direct light, it caused a signal underestimation at almost all the wavelengths along the selected range, especially for T as illustrated in Fig. 4. Scenario 2 produced the minimum deviation from the corresponding reference signal (Table II). Thus, if compared to the standard sample holder set-up (scenario 1), adding a sample holder in front of the white reference while acquiring the REF measurements (Table I, scenario 2) decreases the T error to a value of 2% in flat material and 4% in silicon. Error differences between both materials can be attributed to their different OP s and thickness. Scenario 3 and 4 produced an error per material equal to the one in scenario 1 and 2, respectively (results not shown), revealing that an empty sample holder added at the corresponding empty port during the stray light measurements (Table I, scenario 3 and 4) has neither effect on R nor on T . This shows that no light leaks in the measuring system were introduced by the use of a sample holder.

The fact that there is no direct reflection from the sample holder contaminating the signal suggests that the driving force of this error is probably the distance of the sample to the integrating sphere's inner surface caused by the sample holder use. According to [2], due to the external integrating sphere ports, the outer wall of the sphere is a few millimeters away from the reflecting inner wall and thus a fraction of the transmitted light fails to strike the integrating surface due to absorption around the port edge producing a systematic underestimation of T . When using a sample holder, this effect is likely to increase due to the increased distance. These absorbed photons might explain the underestimation of both R and T in our results, especially in scenario 1 (Fig. 4). In scenario 2, the same absorption affects the measured reference signal to which the recorded R and T signal is normalized, compensating this effect to some extent. Exceptionally, T of the flat plastic material is overestimated up to 10% above 1100 nm.

In our experiment, the 1-mm thickness of the sample holder was selected after several tests done with different thicknesses of the same metallic plates. The sample holder thickness of 1 mm was the minimum thickness possible to prevent the plates bending slightly when adjusting them to firmly hold real needles of Norway spruce (*Picea abies* (L.) Karst.). With thinner plates we experienced that when screwing the plates to trap the needles in between them (Fig. 1(a)) and prevent misplacements during the measurements, the plates were slightly bending and therefore affecting the position of the illuminated area of the sample and increasing the distance to the sphere's inner wall. Thus, we do not recommend to use thinner plates for needles as thick as 1 mm; however, if the purpose is to measure thinner narrow leaves (e.g. grass) it might be possible to reduce the error by decreasing the thickness of the plates.

We did not test corrections for the absorption effects suggested by [2], which are based on the assumption of negligible absorption of real broad leaves in the NIR. In addition, it was shown that the effect is not systematic [8]. In needle samples, the multiple scattering caused by the non-flat nature of the needles might cause a portion of the light to be scattered directly onto the edge of the sample port, producing a different response in the absorption than the broad leaves. This is especially interesting when the cross-sectional shape of the needles would result in low scattering angles, which is reported to increase the mentioned effect [9]. In [9], a diffuser between the sample and the integrating sphere was used during T measurements to minimize this apparent absorption problem in low-angle scattering samples. The results showed an improvement on the accuracy of transmittance in glass samples but to the best of our knowledge the technique has never been applied to real leaves.

For the objective of this paper we recommend to use the sample holder set-up of scenario 2 in order to compensate for the

sample holder effect when using a single-beam integrating sphere with external sample ports as the one used in our experiment.

B. Effect of the Multiple Scattering Between Neighbor Needles

The multiple-scattering hypotheses regarding the influence of the needle cross-section shape and the distance between needles are testing two steps of our measuring method: the first is the sample scanning required for the $Digital_{GF}$ estimation ($RMSE_{gf_{DI}}$, (7)), and the second is the sample optical measurements ($RMSE_{gf_{TI}}$, (8)).

Results from $RMSE_{gf_{DI}}$ (Fig. 5(a), left graph) showed that samples with rhomboidal cross-section needles (Rh) have higher errors in their $Digital_{GF}$ than flat (F) and squared scenarios (S) regardless of the gap distance: 10% average error in Rh versus 2-3% in F and S . The highest error among the rhomboidal cross-section scenarios appeared at the shortest distances ($Rh1 = 0.5$ mm), whereas no pattern on the error variation with gap-distance is found in the flat or squared cross-section types. The $Digital_{GF}$ estimation from the digital scanned images is based on applying a threshold to discriminate the needle pixels from the air gap pixels. Thus, the accuracy at which the needle-air edges are estimated, determines the quality of the $Digital_{GF}$ output. This accuracy is determined by the selected combination of scanning and processing settings. The optimization efforts in Appendix II showed that the optimal settings can improve yet not eliminate the difference between the $Digital_{GF}$ and the theoretical reference ($Ideal_{GF}$). Majority of per-scenario comparisons between $Digital_{GF}$ and $Ideal_{GF}$ values did not show any statistically significant differences ($P < 0.05$). The only exceptions were the rhomboidal cross-section scenarios: $Rh1$ and $Rh2$ in reflectance, and $Rh1$ and $Rh3$ in transmittance measurements. These results are in line with $RMSE_{gf_{DI}}$, where $Digital_{GF}$ of Rh needle samples significantly differs from $Ideal_{GF}$. Comparison of the scenarios in Fig. 5(b) revealed that $Ideal_{GF}$ of the rhomboidal cross-section scenarios (Rh) were on average about 8% lower than in the other cross-section scenarios (F and S), where the values were equal. This was expected, since the flat (F) and squared (S) cross-section needles have for the same measurement area the same needle projected area and logically the same $Ideal_{GF}$, in the rhomboidal needle scenarios the needle projected area is larger, and therefore the fraction of gaps and the $Ideal_{GF}$ are smaller. The $Digital_{GF}$ variation did not follow the same trend.

For the flat (F) and squared (S) cross-section scenarios, $Digital_{GF}$ values remained almost similar to the $Ideal_{GF}$ with a 1% difference, attributed to measuring errors (e.g., $BEHD$). $Digital_{GF}$ of the rhomboidal cross-section scenarios were 4% lower than in the other cross-section scenarios (F and S), except for the smallest gap distance scenario ($Rh1 = 0.5$ mm) where it was 4% higher. This explains the higher $RMSE_{gf_{DI}}$ (7) occurring at the rhomboidal cross-section needle scenarios and suggests that the cross-section of the needle elements modulates the scanner light in a way that cannot be compensated by optimizing the scanning or image processing settings. The small differences shown between flat and squared cross-section scenarios imply that differences in the OPs of the material used to simulate the needles (flat plastic vs. the silicon) are not as important as the cross-section. The incident light direction during the scanning of rhomboidal cross-section needles is different from the normal to the needle surface producing a longer photon path and inducing more interactions of the photons between neighboring needles. Consequently, the rhomboidal needles appear in the scanned image optically thinner than in reality, which results in a less accurate estimation of their projected needle area. The second error source contributing to the higher $RMSE_{gf_{DI}}$ of the rhomboidal cross-section needle scenarios is the $BEHD$, which we expected to be more pronounced in the smallest gap distance scenario ($Rh1 = 0.5$ mm). In summary, we observed that: (a) the hypothesis regarding the needle cross-section influence is true for the rhomboidal cases, and (b) the hypothesis regarding the needle air gaps is true only if combined with the rhomboidal cross-section. For the other scenarios, effective scanning and processing settings were found through the optimization (Appendix II). Best results were achieved for the smallest distance, except for rhomboidal cross-section, where 0.5 mm represents less than half-the-needle width. This indicates that distance smaller than half-the-needle width potentially reinforces the multiple scattering effects between needles. Finally, difference between $Digital_{GF}$ in reflectance and transmittance per scenario showed no statistically significant difference ($P < 0.05$), except for the $F1$ scenario, where the bias is attributed to measuring errors.

$RMSE_{gf_{TI}}$ (8) provides results related to the multiple scattering during the spectral measurements performed in the integrating sphere. Different cross-section shapes and distances between needles induce different scattering behavior of the interacting photons, which is expected to result in discrepancies between the $True_{GF}$ and the $Ideal_{GF}$ values. Additionally, the $BEHD$ is expected to contribute to the overall error, especially in the non-flat cross-section and/or small gap distance scenarios. Likely, the applicability of $True_{GF}$ to the rhomboidal cross-section samples is limited by the fact that thickness of their cross-section is not constant. Compared to the needle with 1 mm thick squared cross-section, thickness in the rhomboidal needle is 1.41 mm for the central part and decreasing towards the edges. The volume determining the optical thickness of both silicon needle types, is, however, equivalent. We assume that a higher absorption rate in the central part of a rhomboidal needle is compensated by a lower absorption at thinner edges. Moreover, for wavelengths where absorption is low and scattering dominates the measured signal, difference in thickness is less important than difference in cross-section shape ruling the scattering. Scattering (i.e. albedo W_{needle}) between 500 and 1700 nm is on average higher than needle absorption (A_{needle}). The concept of $True_{GF}$ is, therefore, considered as applicable in this wavelength range also for the rhomboidal cross-section scenarios.

Comparison of $True_{GF}$ against $Ideal_{GF}$ values showed a statistically significant difference for all scenarios ($P < 0.05$) except the transmittance of rhomboidal cross-section $Rh1$. When comparing the cross-section scenarios of the same gap-distance (e.g. $F2$, $S2$, and $Rh2$), the error generally increased from flat to squared to rhomboidal cross-section (Fig. 5(a), middle graph), which

supports our hypothesis (a). This trend is produced by the $True_{GF}$ values, which, especially in transmittance, tend to increase from F to S to Rh in most cases (Fig. 5(b)). The $S2$ error in reflectance, which is higher than $Rh2$, and also equal errors of $F2$ and $S2$ in transmittance are not following this general trend. These exceptions are caused by the $Ideal_{GF}$ value, which, being by definition equal in F and S , is distorting the trend. The $True_{GF}$ variation indicates explicitly existence of the multiple scattering effect due to the cross-section shape, especially when comparing F and S cases. If there is no influence from the multiple scattering, $True_{GF}$ values of F and S cases should be theoretically similar, as it occurs with the $Ideal_{GF}$. Results are however showing that $True_{GF}$ is not defined only by the geometry of the sample, but also by the scattering processes triggered by this geometry during the optical measurements. Concerning the second hypothesis about the gap-distance, error tendency in the F cross-section scenarios differs from the non-flat (S and Rh) ones. In the flat scenarios the error does not follow the expected trend (i.e., it increases with decreasing gap distance). The highest deviation (9%) occurs at gap distances of 1 mm ($F2$), while more similar values occur for $F1$ and $F3$ cases. As expected, the highest errors of the non-flat scenarios appear at the narrowest gap distances ($S1$ and $Rh1$; 0.5 mm), except in reflectance of the S cases where error behavior is similar to the one found for F scenarios. We can, therefore, conclude that the best results are achieved with the smallest distance of 0.5 mm in flat cross-section scenarios, while the opposite occurs in the non-flat cases, where the error is generally higher for Rh needles, especially in transmittance. Since 0.5 mm in Rh cross-section scenarios is less than half-the-needle width, the multiple scattering of photons seems to be reinforced by a too small distance between the needles. In addition, the error variation between reflectance and transmittance is of 1% for the same flat cross-section scenarios, whereas in the non-flat cases the transmittance error values are much higher than the reflectance errors: 5% higher for the squared cross-section scenarios (S) and 10% for rhomboidal (Rh). These results imply that the cross-section effect appeared to have a greater influence than the gap distances tested in this study, especially for transmittance measurements (average variation of the error due to the cross-section is about 14% compared to the 3% variation due to the gap distance).

If we look at the difference between the $True_{GF}$ and $Ideal_{GF}$ values per scenario (Table III), we observe a systematic overestimation of the needle transmittance happening in the non-flat cross-section scenarios. The fact that the GF -corrected T_{needle} , computed using the most accurate GF available (i.e. $Ideal_{GF} \pm BEHD$), did not reproduce the ‘nongap’ reference ($T_{uncutSH}$) indicates that the multiple scattering between non-flat needles contributes to the transmittance signal recorded in the integrating sphere.

Results of $RMSE_{gf_{TD}}$ (9) in Fig. 5(a) (right graph) show a pattern that is similar to $RMSE_{gf_{TI}}$, especially in the flat (F) and squared cross-section scenarios (S). This suggests that the $BEHD$, which is not present in $RMSE_{gf_{TD}}$, does not have a crucial effect on our $Ideal_{GF}$ values. A higher influence of the $BEHD$ and also of the light scattering effect during the sample scanning can be seen in the rhomboidal cross-section (Rh) scenarios, where $RMSE_{gf_{TD}}$ was reduced on average by 2 % in reflectance and by 7 % in transmittance, especially for the smallest gap distances ($Rh1$). When comparing $True_{GF}$ and $Digital_{GF}$ values per scenario, a statistically significant difference ($P < 0.05$) was found for most cases, except $F1$, $Rh1$ and $Rh3$ in reflectance and $Rh1$ in transmittance. High overlap of variation ranges (i.e., mean $\pm$ standard deviation; Fig. 5(b)) can explain the similarities found in these cases between $Digital_{GF}$ and $True_{GF}$. Finally, Fig. 5(b) illustrated a good agreement between the $Digital_{GF}$ and the $True_{GF}$ reflectance values, whereas the transmittance $True_{GF}$ values are on average 12% higher than the $Digital_{GF}$ estimations, which is caused mainly by the above discussed contribution from the multiple scattering during the optical measurements.

C. Validation of OPs after GF Correction

Mean individual needle-leaf directional-hemispherical R_{needle} and T_{needle} per spectral waveband computed using Mesarch’s formulae for GF -correction (1) and (2), were compared to the two available corresponding references (average of 10 samples): (a) the broad-leaf R or T for the best sample holder scenario ($R_{uncutSH}$ or $T_{uncutSH}$); and (b) the broad-leaf R or T measured without sample holder ($R_{reference}$ and $T_{reference}$). Results from case (a) (Fig. 6 (a)) show that the error in transmittance was on average 3% higher than in reflectance for the flat cross-section scenarios (F) and 7-10% for the non-flat (S and Rh) scenarios. Statistical analysis revealed no significant difference ($P < 0.05$) between R_{needle} and $R_{uncutSH}$ (except for $F2$, $F3$, $S2$ and $Rh2$), while the difference between T_{needle} and $T_{uncutSH}$ was significant (except for $F1$, $F2$ and $F3$). For the same gap-distance, error tended to increase in the direction flat-squared-rhomboidal cross-sections, except for the scenario with gap distance 1 mm ($F2$ - $S2$ - $Rh2$), where the error behaves differently depending on the cross-section case. Among the same cross-sections, the error tended to increase from the smallest gap-distance (scenario 1=0.5 mm) to the widest (scenario 3=1.5 mm), especially for the non-flat (S and Rh) cases. However, values at gap-distance scenario 2 (1 mm) did not follow this pattern and differed per cross-section case. Finally, the error variation associated with increasing gap tended to be smaller than the error variation associated with cross-section (about 1 % smaller in reflectance and 5% in transmittance).

These results can be explained by analyzing the GF -corrected needle signal shown in Fig. 7. The GF -corrected R_{needle} signal tends to be overestimated in the flat and squared cross-section but underestimated in the rhomboidal, and conversely, T_{needle} is underestimated in the flat whereas overestimated in the non-flat cases. These results per scenario are in line with the ones shown in Table III. In addition, the higher deviation from the reference occurring in the non-flat cross-section scenarios, especially in transmittance, is obvious. The underestimation of T_{needle} in the flat needles was expected since it has been reported by [1] after their analysis performed on flat ‘film-strips’ needles. This suggests that even though we optimized the GF estimation for this

type of needles, the method is still producing an inherent measurement error. The overestimation of T_{needle} in non-flat needles indicates that the increase in recorded signal due to the multiple scattering is such that even the overcorrection originating from the *Mesarch*'s formulae is not able to compensate for it.

In case (b), the error values in reflectance are the same as in case (a) (except 1% decrease on the flat cross-section cases at 1 mm ($F2$) and 1.5 mm gap distances ($F3$)). In transmittance, a systematic 2% error increase occurring in the flat cross-section scenarios contrasted with an approximate 2% error decrease in the non-flat cross-section scenarios (Fig. 6a). The error increase in the flat cross-section scenarios can be attributed to the effect of the sample holder (2%; Table II), since this effect is included in case (b). However, in the non-flat cross-section scenarios the sample holder effect (4%; Table II) does not induce an increase of the error compared to case (a) but a decrease, therefore improving the total error.

Overall, the final error achieved after optimizing *Mesarch*'s method is about 1% in reflectance and 5-7% in transmittance if the needles are flat. The error increases up to 3-5% in reflectance and 9-12% in transmittance for squared cross-section needles (S) and up to 4-6% in reflectance and 9-17% in transmittance for rhomboidal cross-section needles (Rh).

In addition to $RMSE_{pneedle}$ (10) computed for R_{needle} and T_{needle} , we also computed $RMSE_{pneedle}$ for R_{TOTAL} and T_{TOTAL} , i.e. for signal before the GF correction. Results from (10) in case (b) (i.e. reference measured with no sample holder) showed that, in reflectance the error values tended to be higher for non-flat cross-section scenarios ($RMSE_{pneedle}$ equals 8% in F , 12% in S , 14 % in Rh). Error values increased with increasing gap distance and they were more similar between cross-section cases of the same gap distance scenario in transmittance. The error increase was more abrupt in transmittance than in reflectance, which can be explained by the reflectance and transmittance measurement set-ups. Reflectance signal recorded by the spectroradiometer is dominated by light reflected from the sample. Contrary to this, the light passing through the gaps dominates the signal recorded during a transmittance measurement, because it is stronger than the signal transmitted through needles (especially when samples are mounted with large gap fractions [1]). The fraction of direct transmitted light is so strong, that it masks the influence of the needles cross-sections shape on the transmittance-recorded signal. Conversely, it has no influence on the reflectance measurements.

Transmittance RMSE variation with increasing gap distance is linear for the flat cross-section scenarios (F ; $R^2 = 1.00$) and non-linear for the non-flat cases ($R^2 = 1.00$ for a polynomial of 2nd degree fit; 0.95-for S and 0.96 for Rh scenarios for linear fit) (Fig. 6 (b)).

Function fitting analysis of R_{TOTAL} and T_{TOTAL} variation with increasing gap distances, showed that contrary to the expected linear relationship between non- GF -corrected reflectance and transmittance signals and the sample GF , the relationship is closer to a polynomial function of 2nd degree. In reflectance and transmittance R^2 equals 1 for polynomial fit, while it is 0.99 in F , 0.97 (R) and 0.94 (T) in S , and 0.75 in Rh cross-sections for a linear fit. These results suggest that while the F scenarios do follow a linear variation with increasing gap [1], the rhomboidal ones have a clear non-linear trend, which supports our hypotheses about the interaction between needles in the non-flat cases. Consequently, irrespective from the various method errors (i.e., optical measurements and GF estimation through scanning and digital image processing), *Mesarch*'s algorithms have a lower accuracy for non-flat cross-section needles, because the multiple scattering effects are not taken into account. Nevertheless, more measurements, testing other gap distances, and perhaps ray tracing computer simulations are needed to approve this finding.

Finally, the non GF corrected signal showed the influence of the sample holder presented previously. In all scenarios, the sample holder induced higher errors in reflectance and lower in transmittance (Fig. 6 (b)). This is caused by $T_{reference}$ signal being higher than the $T_{uncutSH}$ (Fig. 4) signal, and therefore closer to the noncorrected needle transmittance signal. This situation is propagated to the GF -corrected T_{needle} signal (Fig. 6 (a)) only in the non-flat cross-section scenarios, where a lower performance of the correction algorithms causes the T_{needle} signal to remain high. However, T_{needle} of the flat cross-section cases decreased after the GF correction. It is, therefore, closer to the $T_{uncutSH}$ than to $T_{reference}$, which explains the results obtained from (10).

IV. CONCLUSION

In this study we performed comprehensive uncertainty analyses of the method developed by *Mesarch* for measuring coniferous needle-leaf OPs . Our study focused on the following measurement aspects: 1) the effect of a sample holder used to support the needles during the measurements, and 2) the effect of the multiple scattering in between the measured artificial needle leaves. Analysis of the sample holder effect showed an average underestimation of the needle transmittance signal of 2% in flat needles and 4% in non-flat needles (based on the RMSE according to (3)). Results on the sensitivity of the digitally estimated GF to the image acquisition and image processing settings showed that optimization of these settings reduced the deviation considerably, producing a negligible error. However, in spite of using the most optimal settings, multiple scattering between artificial needles was still affecting the digital GF estimation, resulting in average errors of only 2-3% in samples with flat and squared cross-section needles, but of about 10% in samples with rhomboidal cross-section needles. The last case showed a clear overestimation of the digital GF , especially when the needles were as close to each other as 0.5 mm, which is for needles of rhomboidal cross-section less than half-the-needle width. This indicates that half-the-needle width is a threshold at which the multiple scattering between needles is reinforced and biases the measurement. The results of our sensitivity analysis scenarios suggest that the needle cross-section might have a stronger negative effect on the needle-leaf OPs than the needle gap distance. The multiple scattering between artificial needles also affected the signal recorded during the optical measurements, causing

higher deviations of non-flat needle-samples, particularly in transmittance. The needle transmittance corrected for the GF using *Mesarch's* formulae was about 10-20% higher than the 'nongap' (*broad-leaf*) reference. The interaction of the incoming light with rhomboidal cross-section needles leads to a non-linear variation of the signal with increasing needle gap distance. For this reason, *Mesarch's* formulae, which are based on a linear GF correction, are not suitable for computing the OPs of non-flat needles. Overall, the final error achieved after optimizing the image scanning and processing settings was about 1% in reflectance and 5-7% in transmittance for flat needles. The error increased up to 3-5% in reflectance and 9-12% in transmittance for squared cross-section needles, and even up to 4-6% in reflectance and 9-17% in transmittance for rhomboidal cross-section needles. In general, more accurate OPs can be achieved when the distance between measured needles is about half the needle width (i.e., 0.5 or 0.7 mm in our cases). The results of this study pointed out that approaches designed to measure more comprehensively OPs of non-flat coniferous needle samples should take into account multiple scattering between the measured leaves as currently done in radiative transfer modeling [10].

APPENDIX I

Technical details about the sample holder configurations are presented in Fig. 8. Computation of R_{TOTAL} in (1) and T_{TOTAL} in (2) was done through the algorithms (11) and (12) recommended by the sphere manufacturers:

$$R_{TOTAL}(\lambda) = \left(\frac{R_{sample}(\lambda) - STR_{Re.fl.mode}(\lambda)}{REF_{Re.fl.mode}(\lambda) - STR_{Re.fl.mode}(\lambda)} \right) \cdot R_r(\lambda) \quad (11)$$

and

$$T_{TOTAL}(\lambda) = \left(\frac{T_{sample}(\lambda) - STR_{Trans.mode}(\lambda)}{REF_{Trans.mode}(\lambda) - STR_{Trans.mode}(\lambda)} \right) \cdot R_r(\lambda), \quad (12)$$

where $R_r(\lambda)$ is the reflectance of the calibrated reference standard at wavelength λ and the other inputs are explained in Table I. Each measured input represented an average of 100 spectral scans.

APPENDIX II

A. Optimization of scanning and processing settings for digital GF estimation

The $Digital_{GF}$ is defined as the ratio of the number of air gap pixels inside the measured (i.e. illuminated) area of a needle sample to the total number of pixels inside the measurement area (number of needle pixels + gap pixels, represented by the empty scanned mask). Computation of the $Digital_{GF}$ required masking and scanning each needle sample, and subsequently processing the digital output image. The total number of air-gap pixels in the digital image was calculated by applying a 'white-pixel-threshold' to discriminate gap pixels from needle pixels [1]. The optimal scanning settings (resolution (r), brightness (b) and contrast (c)) and the 'white-pixel-threshold' values (t) were identified by conducting two sensitivity analyses. For the first one, we built three *1-needle-sample scenarios*, one per needle cross-section shape available in our study (i.e., flat- F , squared- S and rhomboidal- Rh). Each sample (5 per scenario) was composed of only 1 needle element, which was carefully placed inside the sample holder at a known distance from the center of the holder aperture. We scanned each masked (1-needle) sample applying 300 *scan-settings scenarios* defined by the varying scan r - b - c combination (Table IV, 1st sensitivity analysis). All scans were saved as 8-bit-grey scale digital images and each of them subsequently processed to estimate the $Digital_{GF}$. The processing was performed for each scan, according to 49 *image-processing scenarios*, where the 'white-pixel-threshold' (t) was varying between 5.1-249.9 (corresponding to values within the range of an 8-bit-grey scale digital image -0 to 256- selected in 2% steps). $Ideal_{GF}$ and $Digital_{GF}$ were computed per *1-needle-sample scenario* from each scan. The sensitivity of the $Digital_{GF}$ to the scanning and image-processing settings was analyzed with (13) for all r - b - c - t combinations by varying one of the four parameters at a time and fixing the others at their minimum, median and maximum value:

$$\varepsilon T(s, i, r, b, c, t) = \left| \frac{Digital_{GF}T(s, i, r, b, c, t) - Ideal_{GF}T(i)}{Ideal_{GF}T(i)} \right|, \quad (13)$$

where $Digital_{GF}T(s, i, r, b, c, t)$, is the GF estimated for the standard-needle-sample scenario i ($i=1, \dots, 9$), from the digital image of sample s ($s=1, \dots, 5$), which was scanned and processed with the T mask for scan-settings scenario $r-b-c$ ($r-b-c=1, \dots, 300$) and image-processing scenario t ($t=1, \dots, 49$); the $Ideal_{GF}T(i)$ is the corresponding $Ideal_{GF}T$ for the same i . No significant differences are expected between the resulting $Digital_{GF}T$ and $Digital_{GF}R$ for a sample of the same $r-b-c-t$ combination, because there is no repositioning of the needles inside the sample holder when measuring and scanning a sample in T and R mode. Only the size of the illuminated area differs slightly during the OPs measurements, which translates into the use of T and R specific masks during the scanning and digital image processing. Following this rationale, the sensitivity analysis was performed only for T scans (i.e. samples with T mask). The $BEHD$ in the 1-needle samples is expected to be almost negligible (it is feasible to position a single needle in the sample holder with the desirable precision). Therefore, this εT is used as the best indicator of the error inherited from the GF estimation via digital image processing.

Based on the results from the first sensitivity analysis, we performed a *second sensitivity analysis* applied on all samples corresponding to the nine needle sample scenarios ($F1, F2, F3, S1, S2, S3, Rh1, Rh2$, and $Rh3$). Here, we used the optimal 3 *scan-settings scenarios* identified within the previous extensive sensitivity analysis (Table IV, 1st sensitivity analysis) and all scans were digitally processed according to the same 49 *image-processing scenarios* used in the first sensitivity analysis. The effect of the GF estimation through scanning and image processing for each $r-b-c-t$ combinations was analyzed through (13) as in the first analysis. Following the same logic used in the *first sensitivity analysis*, this *second sensitivity analysis* was carried out only for T scans (using T mask) and subsequently the resulting optimal $r-b-c-t$ per scenario were used to compute the final $Digital_{GF}T$ and $Digital_{GF}R$ per sample per scenario.

As explained in section II.D, potential differences are expected in the optimal $r-b-c-t$ settings per scenario due to variations in the scattering behavior of the scanner light caused by the different needle cross-section shapes and distance between needles. These scattering effects might result in different sharpness of the needle element edges in the scanned digital images forcing the need for specific optimal settings per scenario.

B. Effect of the GF estimation through digital image processing

Results from the *first sensitivity analysis* (i.e., 1-needle-sample scenarios, Table IV) showed that the deviation of the $Digital_{GF}$ from the $Ideal_{GF}$ expressed through (13), followed a similar pattern among the three available cross-section scenarios, with higher values in the rhomboidal cross-section (Rh) than in the flat (F) and squared (S) ones, where values are similar. Regardless of the scanning settings, the maximum error appeared always when the image-processing threshold (t) was fixed at the minimum value, with error values being: 19(3) % for flat (F) and squared (S) cross-section scenarios; and 28(4) % in rhomboidal (Rh) ones (numbers between brackets refer to the standard deviation). Conversely, if the threshold was not fixed at its minimum, error decreases to 1(0.3) % (4(1) % in rhomboidal scenario), irrespective of the other parameters. If the scanning settings were fixed at their minimum or median values, the error variation with increasing threshold showed an inflexion point (exact value differs per scenario) from which lower threshold values trigger the maximum error. Threshold values above the inflexion point caused error drops of 20% in the flat (F) and squared (S) scenarios and of almost 30% in the rhomboidal (Rh) ones, resulting in a stable error value (<4%). The optimal threshold value (i.e., for minimum error) differed per scenario and produced minimum errors of 0.1%. If the scanning settings were fixed at their maximum, error values stayed stable for all thresholds but increased among the cross-section scenarios: 1% for the flat scenario (F); 2% for the squared cross-section (S); and 7% for the rhomboidal (Rh) one. Thus, the main findings from 1st sensitivity analysis are: (1) the reliability of the $Digital_{GF}$ is mainly driven by the threshold value; (2) needles cross-section affects the image processing and thus optimal threshold differs per scenario, as expected; and (3) selection of the proper threshold can ensure a negligible effect of the GF estimation through scanning and image processing (0.1 %).

Based on the higher sensitivity of the $Digital_{GF}$ to the threshold value, the *scan-setting scenarios* were optimized to the best three combinations, whereas the 49 *image-processing scenarios* (threshold values) remained unchanged during the *second sensitivity analysis* (Table IV). Results here showed that, similarly to the *first sensitivity analysis*, variation of the error according to the scanning settings followed a similar pattern in all the nine needle-sample scenarios, with a bigger deviation in the rhomboidal cross-section scenarios than in the flat (F) and squared (S) ones. The error here appeared to be also driven by the threshold value. Thus, the scanning settings ($r-b-c$) combination corresponding to a resolution of 800 ppi, and 0% contrast and brightness (800-0-0) was selected and fixed in further analysis, aiming to standardize the technique involving the lowest requirements possible, i.e., 800-0-0 settings are easily available in common and low-cost scanners.

For this scanning settings (800-0-0), the error variation with increasing threshold (values from 5.1-249.9) showed also an inflexion point, below which lower threshold values triggered errors above 100%. This point differs per scenario covering 30% of the lowest threshold values among the available range (5.1-249.9), i.e., from 0-77 (absolute t values per scenario in Table V). For thresholds above the critical value, the following patterns were observed: (1) for the same gap-distance scenarios (e.g., $F1$,

SI, and *Rh1*) the effect of the needle cross-section triggered higher errors in the rhomboidal cross-section scenarios (*Rh*) and similar in the flat (*F*) and squared (*S*) ones; (2) for the same cross-section scenarios (e.g. *F1*, *F2*, and *F3*), the effect of the gap-distance between needles showed that errors tended to be higher at gap-distance scenario 1 (0.5 mm, e.g. *F1*) and lowest at gap-distance scenario 2 (1 mm, e.g. *F2*), closely followed by values at scenario 3; (3) for the same threshold value, the error corresponding to a certain needle sample scenarios was higher than at the corresponding 1-needle-sample scenario (e.g. *Rh1*, *Rh2*, or *Rh3* versus *Rh* 1-needle-scenario). This difference was higher in the rhomboidal cross-section scenarios (*Rh*). These three error patterns support our hypotheses about the *BEHD* increase, except for the slightly lower error in gap scenario 2 (1 mm) instead of in 3 (1.5 mm).

As expected, the optimal threshold, i.e. the value giving the lowest error per needle-sample scenario, differed per scenario (Table V). Similarly, if we compared results from a needle-sample scenario with its corresponding 1-needle-sample scenario (e.g. *SI*, *S2*, and *S3* standard-needle sample scenarios versus *S* 1-needle sample), we found not only an increase of the error in the first ones (attributed to the *BEHD*), but also a change in the optimal threshold value. This supports the hypothesis concerning the effect of the needle cross-section shape and gap-distance on the multiple scattering of the scanner light. In 1-needle-sample scenarios, where *BEHD* is assumed to be negligible, for a different material (flat plastic vs. silicon), the optimal t value is different because of differences in the material properties. Also, for the same material but different cross-section (*S* and *Rh* 1-needle-sample scenarios), photons from the scanner light will be hitting the needle surface at different incident angles (Fig. 3(h) vs. Fig. 3(i)), and scattered differently, which results again in different optimal t value, due to differences in the needle-edge sharpness. This is also supported by the broader critical threshold value-range observed in rhomboidal 1-needle-sample scenario, where higher thresholds are needed to sharpen the image and reduce the error in the *Digital_{GF}*. However, once the critical threshold value range is passed, the error stays stable for the three 1-needle-sample scenarios no matter the cross-section shape, and the error can be negligible if the proper threshold value is selected in each scenario. When we increase the number of needles in the scanned area, as in the nine needle-sample scenarios, on top of the effect caused by material and cross-section shape differences, an extra scattering phenomenon is likely to occur. There, the photons hitting the needle surface might be scattered towards neighbor needles, especially if they hit in a direction different than the normal to the needle surface (*Rh* cross-section; Fig. 3(i)) and if the needles are close to each other (i.e. smaller gap-distance scenarios). Thus, we also expect the optimal settings to be different per gap-distance scenario when the material and cross-section is the same (e.g. *Rh1*, *Rh2*, and *Rh3*). The multiple-scattering phenomenon however is not expected to be driving the error once the best optimal threshold values corresponding to each scenario are applied, because as we observed in the 1-needle-sample scenarios, it is possible to reduce the error to a negligible value with the proper image processing settings. Thus, a major portion of error observed in the nine needle-sample scenarios is most likely caused by the *BEHD*, whereas a major portion of the error observed in the 1-needle-sample scenarios is most likely caused by the inherent digital *GF* estimation technique. Since the latter is negligible for the optimal settings, we can conclude that the effect of the *GF* estimation through digital image processing can be optimized to insignificant levels.

C. Recommendations

We would like to finalize this section by giving some recommendations to potential users of this methodology. Differences per scenario on the optimal threshold make the standardization of the *Digital_{GF}* estimation a complicated task, even enhanced in real needle samples due to the irregular shape and size of the needles. Based on our results, we recommend keeping the scanning settings (r - b - c) combination on 800-0-0 as a standard, since the error is driven by the threshold value. We also recommend selecting a threshold value higher than the first 40% values of the available t range (5.1-249.9). In Table V we provide a selection of threshold value ranges corresponding to the minimum $\varepsilon T \pm$ a deviation of 1%. We recommend applying the threshold values corresponding to the closest cross-section shape available in Table IV and leaving a gap-distance between the needles similar to the needles width, since the 1 mm gap distance scenario appears to give the lowest error. As an example, for Norway spruce (*Picea abies* (L.) Karst.; rhomboidal cross-section) or Scots pine needles (*Pinus sylvestris* L.; semicircular cross-section), we would select the t values from the scenario *Rh* at 1 mm distance. Nevertheless, since manual manipulation of small needle implies a *BEHD*, avoiding big gaps is preferred [1], inclining the choice more towards ranges between 0.5-1 mm gap distance (half to same distance as the needle width) than 1-1.5 mm gap distance (same or wider distance than needle width).

Finally, the selection of the threshold value range in this analysis is very specific and chosen to cover all possible values available in our 8-bit grey-scale digital images. Scanned images saved in formats other than 8-bit grey-scale might require an adaptation of the threshold range according to the same logic. Nevertheless, aiming to standardize the technique as much as possible we provide all results concerning the threshold values as a % for the available range, 5.1-249.9 (i.e. instead of selecting the t value 249.9 select the value corresponding to 98% of the available range in the specific digital image, e.g. 16 bit grey scale). However error propagation in a different range of t values has to be tested.

ACKNOWLEDGMENT

The authors thank V. Kaplan, J. Hanus and P. Lukes (Global Change Research Centre AS CR, v.v.i, Brno, Czech Republic) for helping with the experimental set-up, logistics and data processing. The work was supported by the UZH Research Priority Programme on 'Global Change and Biodiversity'.

REFERENCES

- [1] M. A. Mesarch, E. A. Walter-Shea, G. P. Asner, E. M. Middleton, and S. S. Chan, "A revised measurement methodology for conifer needles spectral optical properties: Evaluating the influence of gaps between elements," *Remote Sensing of Environment*, vol. 68, pp. 177-192, 1999
- [2] M. N. Merzlyak, O. B. Chivkunova, T. B. Melø, and K. R. Naqvi, "Does a leaf absorb radiation in the near infrared (780-900 nm) region? A new approach to quantifying optical reflection, absorption and transmission of leaves," *Photosynthesis Research*, vol. 72, pp. 263-270, 2002.
- [3] C. S. T. Daughtry, L. L. Biehl, and K. J. Ranson, "A new technique to measure the spectral properties of conifer needles," *Remote Sensing of Environment*, vol. 27, pp. 81-91, 1989.
- [4] Z. Malenovsky, J. Albrechtová, Z. Lhotáková, R. Zurita-Milla, J. G. P. W. Clevers, M. E. Schaepman, and P. Cudlin, "Applicability of the PROSPECT model for Norway spruce needles," *International Journal of Remote Sensing*, vol. 27, pp. 5315-5340, 2006.
- [5] H. K. Lichtenthaler, "[34] Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," in *Methods in Enzymology*, vol. 148, pp. 350-382, 1987.
- [6] G. V. G. Baranoski and J. G. Rokne, "Biological Issues," in *Light Interaction With Plants: A Computer Graphics Perspective*, first ed. Chichester, U.K.: Horwood, 2004, ch. 4, pp. 46-59.
- [7] G. Schaepman-Strub, M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik, "Reflectance quantities in optical remote sensing-definitions and case studies," *Remote Sensing of Environment*, vol. 103, pp. 27-42, 2006
- [8] H. L. Gorton, C. R. Brodersen, W. E. Williams, and T. C. Vogelmann, "Measurement of the optical properties of leaves under diffuse light," *Photochemistry and Photobiology*, vol. 86, pp. 1076-1083, 2010
- [9] A. M. Nilsson, A. Jonsson, J. C. Jonsson, and A. Roos, "Method for more accurate transmittance measurements of low-angle scattering samples using an integrating sphere with an entry port beam diffuser," *Appl. Opt.*, vol. 50, pp. 999-1006, 2011
- [10] Y. Knyazikhin, M. A. Schull, P. Stenberg, M. Mänttös, M. Rautiainen, Y. Yang, A. Marshak, P. Latorre Carmona, R. K. Kaufmann, P. Lewis, M. I. Disney, V. Vanderbilt, A. B. Davis, F. d. r. Baret, S. p. Jacquemoud, A. Lyapustin, and R. B. Myneni, "Hyperspectral remote sensing of foliar nitrogen content," *Proceedings of the National Academy of Sciences*, December 4, 2012.

Lucia Yáñez-Rausell received the degree in forestry engineering from the Universidad Politécnica de Madrid, Spain, in 2005 and the M.Sc. degree in geo-information science from Wageningen University, The Netherlands, in 2006. In 2007 she worked as research assistant at CSIRO Land and Water, in Canberra, ACT, Australia. She is currently working toward the Ph.D. degree at the in the Laboratory of Geo-information Science and Remote Sensing at Wageningen University and working at the Department of Geography, Remote Sensing Laboratories, of the University of Zurich, Switzerland.

Since 2008, she has been working on narrow leaves optical properties with a special focus on coniferous needle leaves and upscaling issues. Her recent interests include remote sensing of bio- and geophysical parameters for vegetation monitoring studies using radiative transfer models and imaging spectroscopy..

Zbyněk Malenovský (M'06) received the M.Sc. degree in terrestrial ecology from the Palacký University, Olomouc, Czech Republic, in 1998, and the Ph.D. degree in production ecology & resource conservation from Wageningen University, the Netherlands, in 2006. He was previously conducting research at Remote Sensing Laboratories, Department of Geography, University of Zurich, Switzerland. From 2012 he works as a Research Associate for the School of Geography & Environmental Studies, University of Tasmania, Australia. His main scientific interest is development of optical remote sensing approaches assessing quantitatively the physiological state and stress responses of vegetation using models of radiative transfer at both leaf and canopy levels.

Jan G. P. W. Clevers received the M.Sc. degree in agronomy and the Ph.D. degree in remote sensing from Wageningen University, Wageningen, The Netherlands, in 1981 and 1986, respectively. The subject of his dissertation was on the application of remote sensing to agricultural field trials and he developed a practically applicable reflectance model for estimating crop characteristics.

His present activities concern the continuation of the developments of optical reflectance models (including bidirectional reflectance and hyperspectral measurements), the linking to crop growth models, the synergy hypothesis with the purpose of the combined use of optical and microwave observations as well as of prior knowledge, and land cover mapping using remote sensing data at different scales. He has been a Project Manager of several projects within the Dutch National Remote Sensing Program and of a JRC study contract. He has contributed to more than 70 peer-reviewed journal papers. He is currently an Associate Professor and Lecturer in remote sensing at Wageningen University..

Michael E. Schaepman (M'05–SM'07) received the M.Sc. degree and the Ph.D. degree in geography from the University of Zurich (UZH), Zurich, Switzerland, in 1993 and 1998, respectively. In 1999, he spent his postdoctoral time at the Optical Sciences Center, The University of Arizona, Tucson. In 2000, he was appointed Project Manager of the European Space Agency Airborne Prism Experiment spectrometer. In 2003, he accepted a position of Full Chair of geoinformation science and remote sensing at Wageningen University, Wageningen, The Netherlands. In 2009, he was appointed Full Chair of remote sensing at UZH, where he is currently heading the Remote Sensing Laboratories, Department of Geography. His interests are in computational Earth sciences using remote sensing and physical models, with particular focus on the land–atmosphere interface using imaging spectroscopy.

Fig. 1. (a) Sample holder with a sample; (b) Integrating sphere (ASD 190 RTS-3ZC) used for the measurements with the sample holder machined to perfectly fit the sample ports.

Fig. 2. Experimental set-up: (*) These scenarios refer to the 9 *needle-sample scenarios* (*F1*, *F2*, *F3*, *S1*, *S2*, *S3*, *Rh1*, *Rh2*, and *Rh3*) built to analyze the 'effect of multiple scattering' (Section II.D). The scenarios were built by combining three cross-section types (flat-F, squared-S or rhomboidal-Rh) and three air-gap distances, (index *l* = 0.5 mm (e.g. *F1*); 2 = 1.0 mm; and 3 = 1.5 mm). Best outputs from the 'Effect of sample holder' analysis (Section II.C), *R_{uncutSH}* and *T_{uncutSH}*, are used as references for the 'OPs validation' of the GF-corrected-needle-OPs computed per scenario (*R_{needle}* and *T_{needle}*; Section II.E). Best outputs from the 'Scan-and image-processing sensitivity analysis' (Appendix II) are used as final scanning and processing settings for the 'Sample scanning' and 'GF computation through image processing' for all samples of the nine *needle-sample scenarios*.

Fig. 3. Frontal and cross-sectional view of each sample type when placed at the integrating sphere port: the ‘broad-leaf’ samples from both materials ((a), (d), (g)) and the nine standard-needle-sample scenarios: *F1, F2* and *F3* (flat cross-section needles; *S1, S2, S3* (squared cross-section needles; and *Rh1, Rh2, and Rh3* (rhomboidal cross-section needles). Distance between needles is illustrative. Thickness ‘a’ (h) in *F1, F2, F3* is much smaller than the length ‘b’ while it is equal to b in *S1, S2, S3*. If the sample is placed at the integrating sphere in *R* mode the light beam is hitting one side of the sample while the light trap is at the other side of the sample; if the sample is placed in *T* mode the light beam is hitting one side of the sample while the sphere is at the other side of the sample. In (a), (b) and (c) the sample holder is not visible, since the central aperture is bigger than the sphere port. The illuminated area is always smaller than the area of the sample. The arrows represent the incoming light rays and their multiple scattering at the surfaces. Volume scattering (inside the sample) is not shown.

Fig. 4. Average reflectance and transmittance measurements (fractional quantities 0-1) from *broad-leaf* flat plastic samples (left graphs) and silicon (right graphs), measured by positioning of the sample holders corresponding to scenario 1, scenario 2, and without sample holder used ('Reference').

(a)

(b)

Fig. 5. (a) RMSE computed between: 1) $Digital_{GF}$ of the best scanning and image processing settings and corresponding $Ideal_{GF}$ ($RMSE_{gf_{DI}}$, (6); left graph); 2) $True_{GF}$ and corresponding $Ideal_{GF}$ ($RMSE_{gf_{TI}}$, (7); middle graph); and 3) $True_{GF}$ and corresponding $Digital_{GF}$ of the best scanning and image processing settings ($RMSE_{gf_{DT}}$, (8); right graph). RMSE is computed for the wavelength range 450-1700 nm. Labels $F1$, $F2$, $F3$, $S1$, $S2$, $S3$, $Rh1$, $Rh2$, and $Rh3$ on x-axis correspond to the standard-needle-sample scenarios. Part (b) presents the average GF values ($\pm$ standard deviation) per needle-sample-scenario for reflectance (left graph) and transmittance (right graph).

Fig. 6. (a) RMSE in (10) computed for GF corrected reflectance (R_{needle} ; left graph) and transmittance of individual needles (T_{needle} ; right graph) towards two corresponding reference spectra: 1) reference $R_{reference}$ or $T_{reference}$, i.e., R or T spectral signal of the *broad-leaf* measured with no sample holder (' $ref=sp_{reference}$ '); and 2) reference $R_{uncutSH}$ or $T_{uncutSH}$, i.e., R or T spectral signal of the *broad-leaf* (uncut) measured in the best sample holder scenario (' $ref=sp_{uncutSH}$ '). (b) RMSE in (10) computed towards the same references but for R_{needle} and T_{needle} computed before the GF correction is applied. The labels $F1$, $F2$, $F3$, $S1$, $S2$, $S3$, $Rh1$, $Rh2$, and $Rh3$ in the x-axis, represent the nine standard-needle-sample scenarios.

(a)

(b)

Fig. 7. Average directional-hemispherical reflectance (a) and transmittance (b) spectra from: I) individual needle-leaves, computed after the *GF*-correction through *Mesarch*'s formulae (R_{needle} (1) and T_{needle} (2)); II) from the *broad-leaf* measured according to the best sample holder scenario ($R_{uncutSH}$ or $T_{uncutSH}$); and III) from the *broad-leaf* measured without sample holder ($R_{reference}$ and $T_{reference}$).

Fig. 8. Examples of sample holder configurations. (a) top view of 'S+SH' described in Table I (two sample holder plates and the needles in between); (b) top-view of 'W+SH' (Table I); (c)-(f) show a top-view of the ASD integrating sphere for Reflectance mode configurations: (c) R_{SAMPLE} (Table I) for scenario 1; (d) R_{SAMPLE} (Table I) for scenario 2; (e) $REF_{Refl.mode}$ (Table I) for scenario 1; (f) $REF_{Refl.mode}$ (Table I) for scenario 2. (g)-(i) show a top-view of the ASD integrating sphere for Transmittance mode configurations: (g) T_{SAMPLE} (Table I) for scenario 1 and 2; (h) $REF_{Trans.mode}$ (Table I) for scenario 1; (i) $REF_{Trans.mode}$ (Table I) for scenario 2. Port E has the white plug in all configurations (not shown).

TABLE I
SPHERE CONFIGURATIONS PER SAMPLE HOLDER SCENARIO

Scenario	Quantity	Sphere Plugs Configuration per Port				
		Port A	Port B	Port C	Port D	Port E
1	R_{SAMPLE}	L	W	S+S H	P	P
	$STR_{Refl. mode}$	L	W	O	P	P
	$REF_{Refl. mode}$	L	S+S H	W	P	P
	T_{SAMPLE}	P	W	O	L+S+S H	P
	$STR_{Trans. mode}$	P	O	W	L	P
	$REF_{Trans. mode}$	P	W	S+S H	L	P
2	R_{SAMPLE}	L	W+S H	S+S H	P	P
	$STR_{Refl. mode}$	L	W+S H	O	P	P
	$REF_{Refl. mode}$	L	S+S H	W+S H	P	P
	T_{SAMPLE}	P	W	O	L+S+S H	P
	$STR_{Trans. mode}$	P	O	W	L	P
	$REF_{Trans. mode}$	P	W+S H	S+S H	L	P

L = Integrating sphere external light source; W = White reference; S = Sample; O = Empty port (with light trap); P = White plug; +SH=Sample holder is used (e.g. in 'S+SH' the sample holder is holding the sample; in 'W+SH' the sample holder plate is between the port and the white reference. Port A to E corresponds to the ASD integrating sphere ports (ASD 190 RTS-3ZC). Scenario 1 is the starting scenario and the one used by [1]. In Scenario 2 we only add a sample holder plate to the W (i.e. sample holder plate between the sphere and the W). Scenario 3 and 4 (not shown) correspond to the same set up as Scenario 1 and 2 respectively but adding an empty sample holder at the corresponding empty port during the STR signal measurements in both R ($Refl. mode$) and in T ($Trans. mode$). Please refer to Appendix I for details about these configurations.

TABLE II
ROOT MEAN SQUARED ERROR OF *BROAD-LEAF* MEASURED SIGNAL PER SAMPLE HOLDER SCENARIO

Symbol	Flat plastic uncut <i>broad-leaf</i> samples		Silicon uncut <i>broad-leaf</i> samples	
	Scenario 1	Scenario 2	Scenario 1	Scenario 2
RMSE R_{uncutSH}	0.01	0.01	0.02	0.01
RMSE T_{uncutSH}	0.03	0.02	0.07	0.04

These are results from (3). Results from scenario 3 and 4 are equal to scenario 1 and 2, respectively.

TABLE III
INFLUENCE OF NEEDLE CROSS-SECTION SHAPE ON THE *GF*-CORRECTION FORMULAE PERFORMANCE

Cross-section scenario	Relationships (Fig. 5(b))				GF-correction output
	Reflectance		Transmittance		
Flat (<i>F</i>)	$True_{GF} < Ideal_{GF}$	$R_{needle} > R_{uncutSH}$	$True_{GF} <= Ideal_{GF}$	$T_{needle} <= T_{uncutSH}$	Reflectance is overestimated; transmittance underestimated ⁽¹⁾
Squared (<i>S</i>)	$True_{GF} <= Ideal_{GF}$	$R_{needle} > R_{uncutSH}$	$True_{GF} > Ideal_{GF}$	$T_{needle} > T_{uncutSH}$	Both overestimated (reflectance ⁽¹⁾)
Rhomboidal (<i>Rh</i>)	$True_{GF} >= Ideal_{GF}$	$R_{needle} <= R_{uncutSH}$	$True_{GF} > Ideal_{GF}$	$T_{needle} > T_{uncutSH}$	Reflectance is underestimated ⁽¹⁾ ; transmittance overestimated

⁽¹⁾ The standard deviation range (errorbars in Fig. 5(b)) overlaps with $Ideal_{GF}$ value.

TABLE IV
SCANNING AND IMAGE PROCESSING SCENARIOS FOR *GF* DIGITAL COMPUTATION

1 st Sensitivity analysis: <i>1-needle-sample scenarios</i> (<i>F</i> , <i>S</i> , and <i>Rh</i> cross-section scenarios)			
SCENARIO	QUANTITY	UNITS	Values
Scan-settings	Resolution (<i>r</i>)	p.p.i.	800, 1200, 2400
	Brightness (<i>b</i>)	%	0-100 in steps of 10%
	Contrast (<i>c</i>)	%	0-100 in steps of 10%
Image processing	White-threshold (<i>t</i>)	n.a	0-256 in steps of 2%
2 nd Sensitivity analysis: <i>Standard-needle-sample scenarios</i> (<i>F1</i> , <i>F2</i> , <i>F3</i> , <i>S1</i> , <i>S2</i> , <i>S3</i> , <i>Rh1</i> , <i>Rh2</i> , and <i>Rh3</i>)			
Scenario	Quantity	Units	Values
Scan-settings	Resolution-Brightness-Contrast (<i>r-b-c</i>)	p.p.i.; % and %	(1) 800-0-0
			(2) 2400-100-0
			(3) 2400-100-100
Image processing	White-threshold (<i>t</i>)	n.a	0-256 in steps of 2%

p.p.i = pixels per inch, n.a. = not applicable.

TABLE V
RANGE OF OPTIMAL THRESHOLD VALUES PER STANDARD-NEEDLE-SAMPLE SCENARIO

Scenario	Range of optimal t values (relative t in % for range 0-256)	Best threshold
$F1$	249.9 (98%)	
$F2$	137.7 to 234.6 (54 to 92%)	204 (80%)
$F3$	96.9 to 193.8 (38 to 76 %)	147.9 (58%)
$S1$	249.9 (98%)	
$S2$	91.8 to 137.7 (36 to 54%)	107.1 (42 %)
$S3$	173.4 to 244.8 (68 to 96%)	224.4 (88 %)
$Rh1$	249.9 (98%)	
$Rh2$	204 to 239.7 (80 to 94 %)	219.3 (86%)
$Rh3$	249.9 (98%)	

Resolution (r) = 800 ppi, brightness (b) and contrast (c) = 0% and the t values above correspond to the optimal r - b - c - t scan and image processing settings per scenario, used to compute the final $Digital_{GF}R$ and $Digital_{GF}T$ used to compute the OPs per scenario.