Paper:
http://dx.doi.org/10.1016/j.geoderma.2017.11.012
Compatibility of methods used for soil water repellency determination for organic and organo-mineral soils

Ewa Papierowska*, Wojciech Matysiakb, Jan Szatyłowicza, Guillaume Debaene, Emilia Urbanekd, Barbara Kalisz, Andrzej Łachacz

a Warsaw University of Life Sciences - SGGW, Faculty of Civil and Environmental Engineering, ul. Nowoursynowska 166, 02-787 Warszawa, Poland; email addresses: ewa_papierowska@sggw.pl (E. Papierowska); jan_szatylowicz@sggw.pl (J. Szatyłowicz)
b Warsaw University of Technology, Faculty of Mathematics and Information Science, Department of Probability and Mathematical Statistics, ul. Koszykowa 75, 00-662 Warszawa, Poland; email address: matysiak@mini.pw.edu.pl (W. Matysiak)
c Institute of Soil Science and Plant Cultivation - State Research Institute, ul. Czartoryskich 8, 24-100 Puławy, Poland; email address: gdebaene@iung.pulawy.pl (G. Debaene)
d Department of Geography, College of Science, Swansea University, Swansea, UK; email address: e.urbanek@swansea.ac.uk (E. Urbanek)
e University of Warmia and Mazury in Olsztyn, Faculty of Environmental Management and Agriculture, Department of Soil Science and Land Reclamation, Plac Łódzki 3, 10-727 Olsztyn, Poland; email addresses: andrzej.lachacz@uwm.edu.pl (A. Łachacz), barbara.kalisz@uwm.edu.pl (B. Kalisz)
* Corresponding author - email: ewa_papierowska@sggw.pl; tel.: +48 22 593 53 79; fax: + 48 22 593 53 55
**Highlights**

1. Agreement between observers allows for statistical comparison of soil water repellency results
2. Best average CA compatibility was observed between the WDPT/MED and sessile drop methods
3. WDPT below 5 s relates to an average CA below 40° for hydrophilic samples
4. Good relationship between MED and CA were obtained for a range of hydrophobic samples
5. CA ranged from 110 to 130° for strongly hydrophobic samples (600 s < WDPT < 3600 s)

**Abstract**

Soil water repellency (i.e. hydrophobicity, SWR) is a common soil phenomenon inhibiting water infiltration and water movement in the soil. SWR has significant hydrological implications for enhanced overland and preferential water flows and erosion. Several methods are used to determine the degree of SWR. The methods are typically chosen based on their suitability for field or laboratory work, as well as time and resources availability. Unfortunately, each measurement method has a different analytical approach, hence the direct comparison between results from different methods is not possible. A faster and statistically sound technique for converting results is needed, especially to convert results from field applicable techniques to contact angle (CA) value, which is a valuable parameter for soil hydraulic modelling. The aim of this paper is to define a reliable compliance between methods defined on a statistical approach basis (weighted kappa coefficient \( \kappa_w \)), which will allow to determine the CA value based on straightforward tests, such as water drop penetration time (WDPT) and molarity of an ethanol droplet (MED). For this purpose, we measured SWR in 106 organic and organo-mineral soils collected from different locations in North East Poland using four common methods. The sessile drop and Wilhelmy plate laboratory-based methods were used to determine the CA between water and the solid phase. The other two tests are common field methods for assessing SWR by measuring water infiltration time (WDPT) and the highest surface tension of ethanol-water droplet infiltration into the soil (MED). The results revealed that the weighted kappa coefficient, when assumed as a measurement of an observer’s compliance, indicates a strong relationship \( \kappa_w = 0.84 \) between the average CA \( CA_{av} \), measured with the sessile drop method, and the median value of the WDPT \( WDPT_{med} \). Based on the results, we can conclude that hydrophilic samples with WDPT less than 5 s have the average CA values below 40°, while extremely hydrophobic samples with WDPT above 3600 s have CA values higher than 130°. This is a proof that these tests can be a good estimator of CA value for SWR determination in the laboratory or the field.

**Key words**

Soil hydrophobicity; compatibility assessment; agreement between observers; weighted kappa coefficient, MED, WDPT, contact angle

**Abbreviations**
$CA_{in}$ - initial contact angle ($t = 1$ s) obtained by the sessile drop method (°); $CA_{fin}$ - final contact angle ($t = 15$ s) obtained by the sessile drop method (°); $CA_{av}$ - average contact angle obtained by the sessile drop method (°); $CA_{A}$ - advancing contact angle obtained by the Wilhelmy plate method (°); $CA_{R}$ - receding contact angle obtained by the Wilhelmy plate method (°); $CA_{M}$ - average contact angle obtained by the Wilhelmy plate method (°); $WDPT$ - water drop penetration time (s); $WDPT_{av}$ - average value of water drop penetration time (s); $WDPT_{me}$ - median value of water drop penetration time (s); $MED$ - molarity of an ethanol droplet (%)
1. Introduction

The correct estimation of SWR is essential to anticipate and prevent its negative environmental effects. SWR persistence and severity can also be taken into account (Chau et al., 2014). These parameters can be estimated by measuring the CA (e.g. Doerr, 1998), by MED (Letey et al., 2000) and WDPT (Doerr, 1998) tests or by water repellency index (RI) (Tillman et al., 1989). RI being the estimated ratio of soil-water to soil-ethanol sorptivities. Water repellency is a very important soil property, as it has crucial implications for environmental processes related to water management in the soil profile (Doerr et al., 2000). Water repellency of soils limits their water sorptivity (Carrick et al., 2011) and results in uneven moisture distribution, forming preferential water flow in the soil profile. Water moves from zones of less repellent soil, leaving other areas completely dry for long periods (Ritsema et al., 1993; Dekker and Ritsema, 2000) or along pathways resulting from cracks, root channels and other types of macropores (Urbanek and Shakesby, 2009; Urbanek et al., 2015). Therefore, SWR has a significant impact on the phenomenon of water penetration into the soil (DeBano, 1981; Feng et al., 2001). In case of ponded infiltration in hydrophobic soils, the infiltration rate increases with time, contrary to wettable soils, in which the infiltration rate declines over time. In the absence of ponding conditions, a layer of water quickly forms on the surface of hydrophobic soils, which, in the case of heavy rainfall and a steep slope, flows from the ground surface, resulting in erosion (Imeson et al., 1992; Doerr et al., 2000; Schnabel et al., 2013; Butzen et al., 2015). The phenomenon of SWR can also reduce the height of the capillary rise (Scott, 2000) and limit the evaporation (Shokri et al., 2008, Kim et al., 2015), which leads to negative effects on germination and plant growth (Gupta et al., 2015). A highly hydrophobic soil delays the germination process and reduces the germination rate, which may lead to a decrease in crop yields (York and Canaway, 2000; Müller et al., 2014). SWR also affects the soil moisture retention curve (Liu et al., 2012) and soil water conductivity (Lamparter et al., 2010). SWR has typically been associated with dry environments, but research in the last two decades has shown the occurrence of SWR in many different soils under various climatic conditions and vegetation types (DeBano, 2000). Furthermore, the development of SWR in organic rich soils is still far less understood and investigated in comparison to mineral soil, with only a few studies concentrating on peat soil hydrophobicity (Hewelke et al., 2016).

In order to evaluate SWR, fast, simple and inexpensive methods are preferred, such as WDPT test (Bisdom et al., 1993, Doerr et al., 1996; Doerr, 1998; Letey et al., 2000; Jaramillo et al., 2000) or the molarity of an ethanol droplet (MED) test (Letey et al., 2000; Roy and McGill, 2002). Methods involving the determination of the CA value are less frequently selected (Bachmann et al., 2003; Ellies et al., 2005; Ramírez-Flores et al., 2008). Knowledge of the CA allows for the surface free energy of soils to be determined (Hajnos et al., 2013) and the impact of water repellency on soil water sorptivity (Cosentino et al., 2010) or the soil water retention curve to be estimated (Czachor et al., 2010). Moreover, it is also important for geotechnical engineering because it can offer novel solutions to the design of systems in
order to cover overlying municipal or mine waste storage facilities or for other applications (Beckett et al., 2016).

The statistically robust conversion of data from different methods is urgently needed therefore in this paper we test the use of the pedotransfer functions (PTF) to transform SWR results obtained from the field test to mathematically meaningful CA values. PTFs are often defined as predictive functions of important soil properties from easily, routinely or cheaply measured ones (McBratney et al., 2002). The majority of PTFs have been developed to predict soil water retention and soil hydraulic properties (Schaap et al., 2001; Wösten et al., 2001; Manyame et al., 2007; Hewelke et al., 2015; Ghanbarian et al., 2017). Some PTFs have also been advanced to estimate soil physical (Martín et al., 2017, Schjønning et al. 2017), chemical (Valadares et al., 2017, Fernández-Ugalde and Tóth, 2017) and biological (Ebrahimi et al. 2017) properties. A few studies have already applied PTFs to predict SWR (Harper and Gilkes, 1994; Regalado et al., 2008; Lachacz et al., 2009). The aim of this study is to test whether it is possible to predict CA values based on the simple measurements of SWR using WDPT and MED tests.

Many authors have dealt with the comparison of methods to assess soil hydrophobicity (Buczko and Bens, 2006; Leelamanie et al., 2008a and b; Cosentino et al., 2010; Deurer et al., 2011). However, linear and non-linear regression equations proposed in the literature are not universal for all types of soils, nor widely used. In this paper, we introduce an original approach to determine the CA value based on two simple tests (WDPT and MED tests). For this purpose, we propose the use of a statistical technique called rater agreement analysis for estimating the compatibility degree between experts evaluating the same objects (popularly known as agreement between observers). As a measure of agreement between the analysed methods, the weighted kappa coefficient is applied. This statistical technique has not been previously used in SWR studies. We hypothesize that, with a high value of kappa coefficient, which means reliable compatibility between methods, it is possible to estimate the CA value on the basis of simple test (WDPT or MED) results. These tests, contrary to methods of CA measurements, do not require expensive equipment, and can be easily and quickly performed under both field and laboratory conditions.
2. Materials and methods

The study was conducted on 106 soil samples collected from 15 locations and 41 soil profiles located in North East Poland. The examined soils were classified according to the following five reference soil groups (IUSS Working Group, 2014 (updated 2015)): Histosols, Gleysols, Fluvisols, Arenosols and Podzols. Soil samples were collected from organic rich soils, mainly from surface horizons (0-30 cm), but samples from Histosol subsurface horizons (up to 100 cm) were also included. The soils from which samples were collected were formed from fen peats of various botanical origins (sedge, reed, moss, woody/alder) and represented various degrees of decomposition. Some surface horizons of peat soils had undergone secondary transformation and were therefore classified as mursh formations. Similar to murshes, but containing less soil organic matter (SOM) and substantial admixtures of sand fractions, were the semimursh formations (10-20% SOM) and postmursh formations (3-10% SOM). Examined gyttja, which represented bottom lake deposits, were mainly detritus (organic) and calcareous, while silty telmatic muds occurring in river valleys contained over 20% SOM (similar to muds, but containing less than 20% SOM, are muddy formations). Ectohumus formations and A horizons of forest soils, composed of coniferous trees (*Pinus sylvestris* and *Picea abies*), were also included in the study (Arenosols and Podzols).

Soil samples with a defined mineral part, which were included in the study, had the lowest SOM contents and their texture was classified as sandy. As a result of various types of origin and composition, the studied soil formations varied greatly in respect of pH (H$_2$O), ranging from 3.32 to 8.41 (Table 1).

2.1. Samples preparation

Bulk soil samples, after being collected with a shovel at different depths of the soil profile (from ground surface to depth of 1.1 m), were transported to the laboratory and air-dried. Organic soil samples were ground into a fine powder using a ball mill to ensure material homogeneity, while samples from mineral soils were sieved through a 2.0 mm mesh sieve, with visible plant remnants manually removed from the samples. A total of 106 air-dried and homogenized soil samples were separated into subsamples for the determination of SWR, using four different tests of CA measurements (the sessile drop method, the *WDPT* test and the *MED* test), and other basic soil properties: SOM content, organic carbon content (OC), total nitrogen (N) and pH. SWR measurements were conducted at constant room temperature (20°C) and with a relative humidity ranging from 35 to 45%.
Table 1. Soil groups selected physical and chemical properties

<table>
<thead>
<tr>
<th>Reference soil groups</th>
<th>Number of samples</th>
<th>Value</th>
<th>SOM content (%)</th>
<th>OC (%)</th>
<th>N (%)</th>
<th>C:N (-)</th>
<th>pH in H$_2$O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arenosols</td>
<td>14</td>
<td>Min</td>
<td>2.1</td>
<td>0.97</td>
<td>0.47</td>
<td>2.25</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>68.03</td>
<td>44.13</td>
<td>0.04</td>
<td>37.4</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>15.55</td>
<td>9.32</td>
<td>1.55</td>
<td>19.6</td>
<td>4.38</td>
</tr>
<tr>
<td>Fluvisols</td>
<td>11</td>
<td>Min</td>
<td>3.99</td>
<td>2.11</td>
<td>0.14</td>
<td>8.23</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>39.23</td>
<td>14.3</td>
<td>0.99</td>
<td>15.20</td>
<td>7.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>14.79</td>
<td>6.70</td>
<td>0.57</td>
<td>11.50</td>
<td>6.25</td>
</tr>
<tr>
<td>Gleysols</td>
<td>15</td>
<td>Min</td>
<td>3.88</td>
<td>1.55</td>
<td>0.17</td>
<td>9.24</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>89.60</td>
<td>42.75</td>
<td>2.67</td>
<td>20.52</td>
<td>6.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>35.07</td>
<td>15.82</td>
<td>1.11</td>
<td>13.72</td>
<td>5.24</td>
</tr>
<tr>
<td>Histosols</td>
<td>61</td>
<td>Min</td>
<td>4.15</td>
<td>2.27</td>
<td>0.1</td>
<td>9.75</td>
<td>4.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>94.06</td>
<td>50.60</td>
<td>4.75</td>
<td>93.71</td>
<td>8.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>59.85</td>
<td>30.07</td>
<td>1.64</td>
<td>25.85</td>
<td>6.19</td>
</tr>
<tr>
<td>Podzols</td>
<td>5</td>
<td>Min</td>
<td>6.86</td>
<td>3.83</td>
<td>0.16</td>
<td>15.05</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>73.83</td>
<td>36.32</td>
<td>1.63</td>
<td>24.39</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>27.61</td>
<td>13.17</td>
<td>0.66</td>
<td>19.19</td>
<td>3.52</td>
</tr>
</tbody>
</table>

Table 2. SWR assessment scale (proposed, based on the literature)

<table>
<thead>
<tr>
<th>Class</th>
<th>Descriptive label</th>
<th>WDPT (s)</th>
<th>MED (%)</th>
<th>CA (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hydrophilic (wettable)</td>
<td>&lt; 5</td>
<td>0 and 3</td>
<td>&lt; 40</td>
</tr>
<tr>
<td>2</td>
<td>Slightly hydrophobic</td>
<td>5-60</td>
<td>5</td>
<td>40-90</td>
</tr>
<tr>
<td>3</td>
<td>Moderately hydrophobic</td>
<td>60-600</td>
<td>8.5</td>
<td>90-110</td>
</tr>
<tr>
<td>4</td>
<td>Strongly and very strongly hydrophobic</td>
<td>600-3600</td>
<td>13 and 24</td>
<td>110-130</td>
</tr>
<tr>
<td>5</td>
<td>Extremely hydrophobic</td>
<td>≥ 3600</td>
<td>36</td>
<td>≥ 130</td>
</tr>
</tbody>
</table>

2.2. Soil water repellency measurements

2.2.1. Contact angle measurement using the sessile drop method

The CA of a water drop placed on the soil surface (sessile drop method) was measured using the CAM 100 optical goniometer (KSV Instruments, Finland). The test samples were prepared according to the procedure described by Bachmann et al. (2000a). A smooth microscope glass slide was covered with a double-sided adhesive tape, while the soil was sprinkled on a 2-3 cm$^2$ area. Particles were pressed to the tape using a 100 g weight for several seconds, after which the slide was shaken carefully to remove non-adhering soil particles. On the sample surface, a drop of water (volume: 10 µl) was deposited manually with a syringe needle, whose outer diameter was equal to 0.71 mm. Video recording software was run at the moment of placing the drop on the sample. Measurements were performed consecutively.
at 1 s intervals for 15 s. After completion of the image video recording, for each consecutive second, the CA between the surface of the soil and water drop was determined by the software (KSV Instruments, 2004). For each soil, the procedure was repeated three times, each time on a new prepared sample. The drops were always placed on the centre of the glass. The CA were measured 10 min after sample preparation. The following data were used for further analysis (average of three drops): \( CA_{\text{ini}} \) - initial CA \( (t = 1 \text{ s}) \), \( CA_{\text{fin}} \) - final CA \( (t = 15 \text{ s}) \), and \( CA_{\text{av}} \) - average CA (average of 15 s). Usually, when the CA is less than 90°, the soil is considered hydrophilic (Letey et al., 2000; Roy and McGill, 2002).

### 2.2.2. Contact angle measurement using the Wilhelmy plate method

The contact angle of the soil, when immersed in water (Wilhelmy plate method), was measured using a DCAT 11 (Dynamic Contact Angle Meter and Tensiometer, DataPhysics Instruments GmbH). Soil particles were attached to both sides of the glass slide using double-sided adhesive tape (Bachmann et al., 2000a). According to Bachmann et al.’s (2006) suggestion, each sample was immersed in water to a depth of 8 mm at a speed of 0.1 mm·s\(^{-1}\), then emerged at the same speed. For an evaluation of the CA, the wetting force was determined through a linear regression of the recorded weight curve as a function of time. CAs were measured continuously during immersion, but also during the ascension of the sample. However, for equalization, only measurements resulting from water immersion at a depth of 6 to 8 mm (plate immersion phase) were taken into consideration. During the emergence phase, only measurements resulting from immersion at a depth of 0 to 2 mm were considered. With the help of an algorithm implemented in the software, two CA values were obtained (upon immersion and ascension of the plate). During the immersion of the plate, the advancing CA \( (CA_A) \) was defined, while the receding CA \( (CA_R) \) was defined during the ascent. The analyses were only carried out in one measurement cycle, consisting of a single immersion and a single emergence of the plate. Based on these results, the average CA was measured by Formula (1), as proposed by Andrieu et al. (1994) and Marmur (1994):

\[
CA_M = \arccos \left( \frac{\cos CA_A + \cos CA_R}{2} \right)
\]

where: \( CA_M \) is the average CA (º), \( CA_A \) is the advancing CA (º), \( CA_R \) is the receding CA (º).

### 2.2.3. Water drop penetration time test

The WDPT test defines the time needed for a single water drop to penetrate into a soil sample (Doerr, 1998; Hallin et al., 2013). Three to five drops of distilled water with a similar volume were placed on the surface of each sample using a medical dropper and time needed for each drop to completely penetrate the soil was measured with a stopwatch. For further analysis, the average value \( (WDPT_{av}) \) and median \( (WDPT_{me}) \) were calculated using the drop penetration time measurement results for individual soil samples.

### 2.2.4. Molarity of an ethanol droplet test
The MED test determines the severity of water repellency and uses standardized solutions of ethanol at different concentration levels: the higher the ethanol concentration in water, the lower the liquid surface tension. Ethanol solutions with seven different ethanol/water concentrations (0, 3, 5, 8.5, 13, 24 and 36%) were used for the test, with the lowest ethanol strength at which at least three out of five droplets, which were applied to the soil surface, penetrated within 5 s (Urbanek et al., 2015), defining the repellency class.

2.3. Basic soil properties

Basic soil properties (Table 1) and material descriptions are presented in the material and method section. SOM content was estimated from loss on ignition at 550°C. OC content was determined spectrophotometrically after oxidation by a potassium dichromate solution (ISO 14235, 1998) and total N content was measured by Kjeldahl’s method, after which the C:N ratio was calculated. Soil pH (H₂O) was determined by potentiometry (van Reeuwijk, 2002).

2.4. Statistical methods

In order to compare SWR assessment methods, the following statistical techniques were used: rater agreement analysis, marginal homogeneity testing and Kendall’s coefficient of concordance. In addition, the results of the comparisons were presented graphically using Bangdiwala’s (1985) Observer Agreement Chart. All calculations were performed using SAS 9.4 software (SAS Institute, 2013, Cary, NC).

Rater agreement analysis was performed to measure the degree of compatibility between different hydrophobicity assessment methods. This statistical technique is commonly used for measuring the compliance between experts rating the same objects. If it is assumed that measurements (ratings) are conducted by two different evaluators, \( \pi_{ij} \) denotes the probability (frequency) that the first evaluator classified an object in category \( i \) and the second classified an object in category \( j \). The total probability of agreement is \( \sum_i \pi_{ii} \), while perfect agreement occurs when the sum equals 1. A popular measure of compliance between the evaluators, expressed by a single number, is the kappa coefficient (Cohen, 1960), which compares the obtained agreement with the agreement expected for the independent classifications. However, kappa treats ratings, as well as significant and insignificant differences (when measured on an ordinal scale), as nominal. Given that this would have been undesirable in our case, due to the natural ordering of hydrophobicity categories, weighted kappa (Spitzer et al., 1967; Cohen, 1968) was applied. In order to perform computations involving weighted kappa coefficients, a standardized SWR assessment scale, as shown in Table 2, was used. The scale for the WDPT was based on a standardized five point hydrophobicity assessment scale (Dekker and Jungerius, 1999), while we adapted the MED test scale (seven points) proposed by Doerr (1998). The difficulty was in selecting appropriate CA ranges. Finally, a modified scale proposed by Aryal and Neuner (2010) for leave wettability was applied.
Using the scale of hydrophobicity categories from Table 2, the application of the rater agreement analysis proceeded along the following lines. The observers were the results of the assessment of the hydrophobicity sample test obtained from two different methods and, as mentioned above, evaluated according to five categories \((k = 1-5)\). The number of tested soil samples used in this calculation was \(N = 106\). The weighted kappa coefficient used weights \((w_{ij})\), satisfying \(0 \leq w_{ij} \leq 1\), \(w_{ii} = 1\), and \(w_{ij} = w_{ji}\), in order to describe the closeness of agreement between two measurement methods. The weighted compliance was defined as \(\Sigma_i \Sigma_j w_{ij} \pi_{ij}\). The weights were determined using Formula (2), as proposed by Cicchetti and Allison (1971), or Formula (3), as suggested by Fleiss and Cohen (1973):

\[
w_{ij} = 1 - \frac{|i-j|}{c-1} \quad \text{(2)}
\]

\[
w_{ij} = 1 - \frac{(i-j)^2}{(c-1)^2} \quad \text{(3)}
\]

where: \(w_{ij}\) is the weight assigned to the classification pair \(i, j\); \(i\) represents the results obtained by the first method; \(j\) represents the results obtained by the second method; and \(c\) is the maximum number of possible results.

The weighted kappa coefficient \(\kappa_w\) compared the weighted compliance with its expected value under independence \(\Sigma_i \Sigma_j w_{ij} \pi_{i+} \pi_{+j}\) (symbols defined below), as calculated by the following formula:

\[
\kappa_w = \frac{P_o(w) - P_e(w)}{1 - P_e(w)} \quad \text{(4)}
\]

where:

\[
P_o(w) = \Sigma_i \Sigma_j w_{ij} \pi_{ij}
\]

\[
P_e(w) = \Sigma_i \Sigma_j w_{ij} \pi_{i+} \pi_{+j}
\]

therefore:

\[
\kappa_w = \frac{\Sigma_i \Sigma_j w_{ij} \pi_{ij} - \Sigma_i \Sigma_j w_{ij} \pi_{i+} \pi_{+j}}{1 - \Sigma_i \Sigma_j w_{ij} \pi_{i+} \pi_{+j}}
\]

where \(w_{ij}\) represents the weights assigned to classification pair \(i, j\) (Model (2) or (3)); \(\pi_{ij}\) is the probability of classifying to the \(i\)-th category in the first method (Method A) or the \(j\)-th category in the second method (Method B) (the probability matrix is included in Table 6); \(\pi_{i+}\) is the sum of the elements in a row of the \(i\)-th method, \(\pi_{i+} = \Sigma_j \pi_{ij}\); and \(\pi_{+j}\) is the sum of items in a column of the \(j\)-th method, \(\pi_{+j} = \Sigma_i \pi_{ij}\).

The weighted kappa coefficient equals 0 when no agreement between observations exists, or equals 1 (maximum value) when a perfect match between observations with different methods is obtained. Generally, the stronger the (weighted) agreement, the bigger the value of the weighted kappa. Figure 1 shows three commonly used interpretations of the occurrence of the weighted kappa coefficient, all of which were used in this study to present the strength of agreement between the methods for assessing SWR.
The next statistical tool used in this analysis was a chi-squared marginal homogeneity test. With two raters and an ordered scale, one sometimes observes significant differences in the marginal totals (lack of marginal homogeneity). The reason for this may be that one rater has a tendency to classify the objects into lower (or higher) categories than the other rater. The aim of the marginal homogeneity test is to verify the null hypothesis that the marginal distributions of the classifiers are identical (this is precisely the case with marginal homogeneity), which means no rater consistently rates lower or higher than the other one. The marginal homogeneity test is the significance test for a coefficient describing the interaction in a linear model, based on the contingency table containing the results of the classification performed by the two raters. As each soil sample was rated twice (by two given SWR assessment methods), the linear model used was a repeated measures model.

The last statistical tool used in the study was Kendall’s coefficient of concordance (Kendall’s W), which, in general, is a measure of the agreement between several (possibly more than two) methods used to rank several objects. The coefficient represents the ratio of the variability of the total ranks for the ordered objects to the maximum possible variability of the total rank. Kendall’s coefficient of concordance ranges from 0 (no overall agreement between the methods, so the ranks may be regarded as essentially random) to 1 (complete agreement) (Legendre, 2010).

Figure 1. Some frequently used interpretations of weighted kappa values (strength of agreement)
3. Results

3.1. Laboratory results

The results of the CA measurements, using the sessile drop method, the Wilhelmy plate method, the WDPT test and the MED test, are summarized in Table 3, which shows the number of samples obtained in five wettability classes.

For the CA measurements obtained by the sessile drop method, the initial CA values $CA_{in}$ were higher than $90^\circ$ for the vast majority of samples (up to 93 samples), whereas the end angle $CA_{fin} > 90^\circ$ for 72 samples. According to the division proposed in Table 2, when analysing the initial angle $CA_{in}$ of 106 soil samples, only three samples were hydrophilic, while as many as 38 samples were extremely hydrophobic. For the end angle $CA_{fin}$, 11 samples were classified as wettable and 23 as extremely hydrophobic.

For the CA measurements obtained by the Wilhelmy plate method, it was found that the CA values at the moment of plate immersion $CA_{A}$ were higher than in the case of CAs obtained using the sessile drop method. Eight samples were classified as wettable and 77 extremely hydrophobic. In turn, the CA values, which were averaged by Formula (1), were lower than the CA values obtained using the sessile drop method. These CA values were 28% lower than the initial CA values and 22% lower than the average angle value $CA_{av}$.

Assuming the classification of WDPT, as proposed by Dekker and Jungerius (1990), for evaluating hydrophobicity, it can be concluded that, among the 106 tested samples, 50 were classified as extremely hydrophobic, 15 as strongly or very strongly hydrophobic, 13 as moderately hydrophobic, and 12 as slightly hydrophobic. The other 16 samples were classified as hydrophilic.

The MED values ranged from 0 to 36%. According to the classification proposed in Table 2, 22 samples were wettable and 38 samples were extremely hydrophobic.

Table 3. Frequency distribution of SWR classes for different measurement methods: sessile drop, Wilhelmy plate, WDPT test and MED test

<table>
<thead>
<tr>
<th>Class</th>
<th>1 Hydrophilic (wettable)</th>
<th>2 Slightly hydrophobic</th>
<th>3 Moderately hydrophobic</th>
<th>4 Strongly hydrophobic</th>
<th>5 Extremely hydrophobic</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDPT_{av} (s)</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td>17</td>
<td>48</td>
</tr>
<tr>
<td>WDPT_{fin} (s)</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>MED (%)</td>
<td>22</td>
<td>6</td>
<td>7</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>$CA_{in}$</td>
<td>3</td>
<td>10</td>
<td>19</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>$CA_{in}$</td>
<td>11</td>
<td>23</td>
<td>18</td>
<td>31</td>
<td>23</td>
</tr>
<tr>
<td>$CA_{av}$</td>
<td>4</td>
<td>23</td>
<td>16</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>$CA_{A}$</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>11</td>
<td>77</td>
</tr>
<tr>
<td>$CA_{M}$</td>
<td>8</td>
<td>51</td>
<td>26</td>
<td>21</td>
<td>0</td>
</tr>
</tbody>
</table>
3.2. Statistical analysis of the methods’ compliance

For rater agreement analysis, eight classifiers were used: CA_in, CA_fin, CA_av, CA_v, WDPT_med, WDPT_av, MED. Seven classifiers presented rates between 1 and 5. The highest score (5) was not obtained with the CA_m classifier, due to CAs measurements being lower than ≤ 130° (rates ranging from 1 to 4). The CA_v classifier was excluded from the analyses due to the fact that the obtained results only varied in the range from 0 to 85.8° (1-2 class). The analyses were conducted for selected pairs of classifiers, while classifiers originating from the same method were not analysed.

The results from marginal homogeneity testing, which compared 20 pairs of variables (classifiers), are presented in Table A.1. In all of the examined pairs, a lack of marginal homogeneity can be shown (low P-values < 0.05), which indicates that the influence of the examined pairs of classifiers was significant.

Table 4 summarizes Kendall’s W coefficient values calculated with two versions (with different numbers of analysed classifiers). Kendall’s W coefficient value for eight tested classifiers was 0.79. As the CA_v classifier (receding CA from the Wilhelmy plate method) significantly decreased compliance (Kendall’s W coefficient was equal to 0.7), it was decided not to include it in the compliance analysis.

The calculated weighted kappa coefficient values, involving the weights proposed by Fleiss and Cohen (1973), ranged from 0.34 to 0.89 (Table 5). The best compliances were achieved by comparing: (i) the median WDPT and the MED test; (ii) the CA_m obtained by the sessile drop method and the median WDPT; and (iii) the CA_in and the median WDPT. The weakest compliance (the lowest kappa coefficient values) was achieved by comparing the CA_in as measured using the sessile drop method, and the CA_m as measured by the Wilhelmy plate method. Alternatively, when applying the weights as proposed by Cicchetti and Allison (1971), the obtained kappa coefficients values were lower and ranged from 0.17 to 0.76 (Table B.1). The best compliances were obtained for the same pairs when applying the weights as proposed by Fleiss and Cohen (1973). In this case, the weakest compliance was achieved by comparing: (i) the CA_in obtained by the sessile drop method and an CA_m obtained by the Wilhelmy plate method; and (ii) the CA_v and the MED test.

Figure 2 presents the graphs of agreement, as proposed by Bangdiwala (1985), for selected pairs of tested classifiers. The area of full compliance (marked blue on the chart) indicates that both of the compared methods gave the same hydrophobicity sample assessment, while the grey area indicates partial compliance (the compared methods gave adjacent evaluations). The higher the methods’ compliance, the closer the blue colour is to the 1:1 line, while in the case of non-compliance, the area deviates from the line.

The highest compliances were obtained by comparing: (i) the CA_av obtained by the sessile drop method and the median WDPT (Figure 2a); and (ii) the median WDPT and the MED test (Figure 2b). Moderate compliances were achieved by comparing: (i) the CA_v as measured by the Wilhelmy plate method, and the CA_in (Figure 2c); and (ii) the median WDPT and the CA_in (Figure 2d). The weakest
compliances were achieved by comparing: (i) the $CA_M$ as measured by the Wilhelmy plate method and the $CA_{in}$ (Figure 2e); and (ii) the $MED$ test and the averaged CA measured by the Wilhelmy plate method (Figure 2f).

Table 4. Kendall’s coefficient of concordance values

<table>
<thead>
<tr>
<th>Variant</th>
<th>Analysed classifiers</th>
<th>Kendall’s coefficient of concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$CA_{in}$, $CA_{fin}$, $CA_{av}$, $CA_A$, $CA_M$, $WDPT_{av}$, $WDPT_{me}$, $MED$</td>
<td>0.70</td>
</tr>
<tr>
<td>2</td>
<td>$CA_{in}$, $CA_{fin}$, $CA_{av}$, $CA_A$, $CA_M$, $WDPT_{av}$, $WDPT_{me}$, $MED$</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Figure 2. Bangdiwala’s (1985) agreement chart obtained by comparing: (a) the average CA measured by the sessile drop method and the median WDPT; (b) the median WDPT and the MED test; (c) the initial CA and the advancing CA, measured by the Wilhelmy plate method; (d) the initial CA with the median WDPT; (e) the initial CA and the averaged CA, measured by the Wilhelmy plate method; (f) the averaged CA, measured by the Wilhelmy plate method, and the molarity of ethanol droplet test. The blue colour represents exact agreement and the grey colour represents partial agreement.
4. Discussion

4.1. Contact angle measurement

There are many CA measurement methods. The choice of the right method depends on many factors, such as the geometry of the system, size and sample shape. Not all available methods were suitable for the organic soils (Histosols) tested in our study. As samples were air-dried and crushed, any method based on the capillary rise phenomenon could not be used. According to Bachmann et al. (2003), in the case of soils, none of the developed methods can be used to determine the whole range of soils’ wettability in a precise and sensitive manner. In the end, two methods were selected: the sessile drop and the Wilhelmy plate methods. The results obtained were not directly comparable, although they theoretically measured the same parameter, i.e., the CA. With the sessile drop method, the initial CA (CA obtained in the first second of the measurement) and the final CA (in the last second of the measurement) were obtained. Using the Wilhelmy plate method, the advancing and receding CAs were achieved. Many CA definitions, depending upon various factors, including measurement methods, can be found (Kumar and Prabhu, 2007; Marmur, 2009). Bachmann et al. (2000a) emphasized that, in the
case of the CA measurements of porous media, such as soil, there are larger difficulties in evaluating the angle than for flat and smooth surfaces. Hence, the results should be considered as relative values.

4.1.1. Sessile drop method

During CA measurements using the goniometric method, a decrease in the angle value as a function of time was observed for all tested samples. Similar results were obtained by Leelamanie and Karube (2009), who studied CA dependency over time for a sand formation, to which stearic acid was added in order to increase its hydrophobicity. According to the authors, adsorption of water molecules on low-energy surfaces of hydrophobic organic matter may be responsible for these changes. Changes in CA over time is commonly observed when examining various materials, such as wood (Rodríguez-Valverde et al., 2002), pharmaceuticals (Muster and Prestidge, 2002) and plant leaves (Xu et al., 2010). CA value changes are caused by a change in drop shape due to a drop spreading over the surface material, absorption and evaporation. Given the lack of a clear methodology concerning the length of CA measurement, when using the sessile drop method, we assumed that a measurement time equal to 15 s was suitable, while the effect of evaporation was negligible. Research carried out by Whelan et al. 2014 on hydrophobized glass beads showed that, during 15 min measurements, CA values were not significantly affected by evaporation (less than 2% variation).

The tested soils were characterized by a wide range of wettability, from completely wettable to extremely hydrophobic samples. The initial measured CA values ranged from 0 to 149° (mean: 117°), which are higher than those reported in the literature: maximum angle values for mineral formations - 122° (Valat et al., 1991; Bachmann et al., 2000a), 114° (Holden, 1998), 110° (Bachmann et al., 2000b) and 109° (Ellies et al., 2005); initial CA maximum values for minerals - 132° for galena, 125° for malachite and 124° for sphalerite (Lourenço et al., 2015). CA values were lower than those reported in the literature for plant leaf surfaces, which average 160° as a result of the presence of hydrophobic waxes on their surfaces (Neinhuis and Barthlott, 1997). The highest values for the initial CA (CA\text{in}) were obtained in a group of peat formations (149° for alder wood peat), muck soil formations (146° for peat proper muck) and gyttja (145° for detritus gyttja), as a result of their very high soil OC content (> 40%). Unlike other studies (Amer et al., 2017), no relation was found between CA and pH. Final CA values were lower than the initial angle and ranged from 0 to 147° (approximately 98°). This was probably the result of a change in drop shape, which caused drops to spread over the surface material, and absorption. Drop dynamics are also related to ambient temperature, as higher temperatures cause a greater CA value, which decreases with time (Amer et al., 2017). The highest values CA\text{fin} were observed for alder wood peat (147°), peat proper muck (146°) and detritus gyttja (145°). For those samples, characterized by a high OC content, the CA values hardly decreased with time (the CA values were stable).

4.1.2. Wilhelmy plate method
The measured average CA values $CA_M$ were lower than the sessile drop method values, and ranging from 0 to 118° (mean: 84°). The highest values were obtained for the detritus gyttja sample (118°) and willow peat (116°). These samples were characterized by a high OC content (in both cases, OC > 40%). Advancing CA ($CA_A$) values averaged 140° and ranged from 0 to 180°, which is similar to results reported by Wang et al., (2010). According to these authors, in the case of hydrophobic soils, the OC content has a greater influence than texture or soil pH. In turn, Woche et al. (2005) emphasized that the quality of soil organic matter is more important than the total amount of soil OC content. Their advancing CA values were slightly lower than ours and ranged from 0 to 125°. These authors also observed that, for most soil profiles, CA varied irregularly with depth.

The Wilhelmy plate method for measuring CA is seldom used for porous media investigations. This is due to the fact that the tested material must be the same on both sides of the plate. According to Bachmann et al. (2006), the method can be applied in a reproducible and consistent way for a wide range of textures. We overcame that issue by grinding organic soils for homogeneity. On the other hand, the problem was still meaningful in the case of organo-mineral soil samples.

4.2. Measurement of drop penetration time and molarity alcohol test

In practice, due to the laboriousness of the above methods, hydrophobicity evaluation is typically carried out by the WDPT test (Letey, 1969; Doerr, 1998; Jaramillo et al., 2000; Lachacz et al., 2009; Oostindie et al., 2017) and the MED test (Letey et al., 2000; Buczko et al., 2002; Abrantes et al., 2017). These tests are commonly used due to their ease of use, particularly under field conditions.

4.2.1. Water drop penetration time test

The WDPT test, given its simplicity and lack of need of expensive instruments, is commonly used to assess the hydrophobicity of soils under both laboratory (Ma’shum and Farmer, 1985; Bisdom et al., 1993) and field (Adams et al., 1970; Doerr et al., 2009) conditions. According to Diehl and Schaumann (2007), factors such as air movement, relative humidity, temperature and drop sizes have a strong influence on WDPT. Therefore, analyses were conducted under similar conditions for all soil samples. Measured WDPT values were in the range from 1 to 29,921 seconds (> 8 h). The longest WDPT was obtained for the detritus gyttja sample (29,921 s) and the muck peat sample (28,126 s). The measured time results were higher than those reported in the literature (e.g., Doerr et al., 1996; Lachacz et al., 2009). This could be due to higher soil OC content and sample preparation, especially the grinding into fine powder, because material roughness has a considerable effect on water repellency, i.e., the smaller the particle diameter, the greater the repellency (McHale et al., 2005).

4.2.2 Molarity of an ethanol droplet test
Although the length of the drop penetration time was greatly reduced by the MED test, as with the WDPT method, there were some drawbacks regarding the interpretation of the results. Ethanol concentration results ranged from 0 to 36%, with the highest alcohol percentage mean value obtained for peats (28.77%) and the lowest for humus sand (5.5%). Those results are related to the soil OC content. The higher the SOC content, the more hydrophobic the soil, which means that a higher ethanol concentration (lower surface tension) is needed for the drop to penetrate the soil within 5 s. The tests were performed on air-dried samples in order to eliminate the influence of sample moisture (King, 1981).

4.3. Conformity of assessment methods

Given the importance of the soil hydrophobicity problem, the methodology determines its proper extent, while the evaluation of the factors affecting this phenomenon is critical. Moreover, understanding the hydrophobicity phenomenon is not limited to soil science (McHale et al., 2007), but concerns other substances and materials, including polymers (Berger et al., 1997), silicon wafers (Lavi and Marmur, 2004), wood (Gindl et al., 2004; Gérardin et al., 2007; Sedighi Moghaddam et al., 2013), carbon (Jańczuk et al., 1996; Park et al., 2000), materials of plant origin (Wagner et al., 2003; Gaskin et al., 2005; Aryal and Neuner, 2010; Holder, 2012; Rosado and Holder, 2013; Mao et al., 2014; Sikorska et al., 2017), textiles (Liu et al., 2007; Hoefnagels et al., 2007; Zimmermann et al., 2008) and building materials (Tanaka et al., 2002; Klein et al., 2012).

The literature lacks clear criteria for method selection regarding hydrophobicity assessment, while many authors have compared the methods for assessing soil hydrophobicity (Table 3). Equations of linear and non-linear regressions proposed in literature are not universal for all types of soil, nor are they widely used. In addition, these equations mainly concern the hydrophobicity of mineral soils, while the literature pays very little attention to the hydrophobicity of organic soils. For instance, Buczko et al. (2002) obtained reasonable linear correlations ($r^2=0.64$) between the MED and WDPT tests, but they considered only one type of soil (podzolic luvisols with prevailing sandy grain). Meanwhile, when Doerr et al. (2009) investigated SWR for mineral forest soils, no significant correlation was observed between WDPT and CA, as measured by the sessile drop method. It was expected that CAs could possibly be based on test results. As our soils samples were characterized by a great diversity (five different groups of soils), regression equations from the literature could not be applied.

In order to estimate the degree of compatibility between experts evaluating the same objects, this paper proposed using a statistical technique known as agreement between observers. This technique is widely used in medicine (Maxwell, 1977; Van Swieten et al., 1988; Armstrong et al., 1996), but has never been previously used in SWR studies, the best of our knowledge.

When compared, the methods with the proposed hydrophobicity categories (Table 2) presented high weighted kappa coefficient values, which suggests that good compliance allows for the CA to be calculated on the basis of a simple test (WDPT or MED). Good compatibility between these methods
(sessile drop and MED tests) has also been confirmed by other studies (Wijewardana et al., 2016), in which MED methods agreed well at an initial CA of 90 to 100° in the case of hydrophobic soils.

Conclusions
The results have illustrated the possibility of using the statistical method known as agreement between observers to compare hydrophobicity assessment methods for soil formations. A good compatibility obtained between the different test methods and CAs values allowed for the CA value to be estimated on the basis of the performed tests. The best compatibility was achieved between the average CA, measured with the use of the sessile drop, and the tests WDPT and MED tests.

It was estimated that, for hydrophilic samples in which WDPT was less than 5 s, the average CA values were smaller than 40°. However, for hydrophobic samples in which WDPT ranged from 5 to 60 s, the angle values were in the range from 40 to 90°. In medium hydrophobic samples, where WDPT ranged from 60 to 600 s, CA values ranged from 90 to 110°. In strongly hydrophobic samples, in which measured WDPT was 600-3600 s, the CA ranged from 110 to 130°, while, in extremely hydrophobic samples, for which the measured WDPT was longer than 3600 s, CA values were higher than 130°.

Satisfactory compatibility for an average CA was also obtained between the MED test and the sessile drop method. One can, therefore, conclude that hydrophilic samples on which a drop (0 to 3% ethanol) is soaking for less than 5 s, will have an average CA lower than 40°. In weakly hydrophobic samples, whose concentration of an aqueous solution of ethanol was 5%, the CA values ranged from 40 to 90°. In hydrophobic medium samples (ethanol concentration equal to 8.5%), the CA values range from 90 to 110°. Strongly hydrophobic samples, which reached an ethanol concentration equal to 13% and 24%, were characterized by CAs from 110 to 130°, while the extremely hydrophobic samples (ethanol concentration equal to 36%) had CAs higher than 130°. In view of our results, WDPT is a suitable method for determining the average CA of organic and organo-mineral soils.
References


Wijewardana, N.S., Müller, K., Moldrup, P., Clothier, B., Komatsu, T., Hiradate, S., de Jonge, L.W., Kawamoto, K., 2016. Soil-water repellency characteristic curves for soil profiles with organic carbon gradients. Geoderma 264, 150-159.


Table A.1. Marginal homogeneity test results

<table>
<thead>
<tr>
<th>Compared classifiers</th>
<th>Degrees of freedom</th>
<th>Test chi-squared marginal homogeneity</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA_{in} - CA_{fn}</td>
<td>4</td>
<td>72.89</td>
<td>5.57·10^{-15}</td>
</tr>
<tr>
<td>CA_{in} - MED</td>
<td>4</td>
<td>25.45</td>
<td>4.08·10^{-5}</td>
</tr>
<tr>
<td>CA_{in} - WDPT</td>
<td>4</td>
<td>42.49</td>
<td>1.32·10^{-8}</td>
</tr>
<tr>
<td>CA_{in} - WDPT_{av}</td>
<td>4</td>
<td>38.09</td>
<td>1.08·10^{-7}</td>
</tr>
<tr>
<td>CA_{in} - CA_{A}</td>
<td>4</td>
<td>80.97</td>
<td>1.08·10^{-16}</td>
</tr>
<tr>
<td>CA_{in} - CA_{av}</td>
<td>4</td>
<td>46.21</td>
<td>2.23·10^{-9}</td>
</tr>
<tr>
<td>CA_{fn} - CA_{av}</td>
<td>4</td>
<td>28.17</td>
<td>1.15·10^{-5}</td>
</tr>
<tr>
<td>CA_{fn} - MED</td>
<td>4</td>
<td>39.59</td>
<td>5.27·10^{-8}</td>
</tr>
<tr>
<td>CA_{fn} - WDPT_{me}</td>
<td>4</td>
<td>54.52</td>
<td>4.09·10^{-11}</td>
</tr>
<tr>
<td>CA_{fn} - WDPT_{av}</td>
<td>4</td>
<td>50.12</td>
<td>3.42·10^{-10}</td>
</tr>
<tr>
<td>CA_{fn} - CA_{A}</td>
<td>4</td>
<td>172.11</td>
<td>3.68·10^{-36}</td>
</tr>
<tr>
<td>CA_{av} - CA_{A}</td>
<td>4</td>
<td>158.51</td>
<td>3.05·10^{-33}</td>
</tr>
<tr>
<td>CA_{av} - MED</td>
<td>4</td>
<td>33.42</td>
<td>9.80·10^{-7}</td>
</tr>
<tr>
<td>CA_{av} - WDPT_{me}</td>
<td>4</td>
<td>48.93</td>
<td>6.04·10^{-10}</td>
</tr>
<tr>
<td>CA_{av} - WDPT_{av}</td>
<td>4</td>
<td>44.81</td>
<td>4.36·10^{-9}</td>
</tr>
<tr>
<td>CA_{A} - MED</td>
<td>4</td>
<td>52.32</td>
<td>1.18·10^{-10}</td>
</tr>
<tr>
<td>CA_{A} - WDPT_{me}</td>
<td>4</td>
<td>47.31</td>
<td>1.32·10^{-9}</td>
</tr>
<tr>
<td>CA_{A} - WDPT_{av}</td>
<td>4</td>
<td>51.40</td>
<td>1.84·10^{-10}</td>
</tr>
<tr>
<td>WDPT_{av} - MED</td>
<td>4</td>
<td>21.29</td>
<td>2.78·10^{-4}</td>
</tr>
<tr>
<td>WDPT_{me} - MED</td>
<td>4</td>
<td>23.76</td>
<td>8.91·10^{-5}</td>
</tr>
</tbody>
</table>

* Small P-values < 0.05 lead to rejection of the null hypothesis in favour of the alternative hypothesis

Table B.1. The weighted kappa coefficient defined by Formula (2), together with the interpretation of compliance by various authors

| Compared pairs of classifiers | Weighted kappa coefficient (|) | Confidence limits (Lower | Upper | Strength of agreement (Landis and Koch, 1977) | Strength of agreement (Cohen, 1968) | Strength of agreement (Peiris et al., 2004) |
|-------------------------------|--------------------------------|----------------------------|-------|-----------------------------------------------|--------------------------------------|------------------------------------------|
| CA_{in} - MED                 | 0.51                           | 0.41 | 0.60 | Moderate | Fair | Poor | Poor | Moderate or good |
| CA_{in} - WDPT_{me}           | 0.59                           | 0.50 | 0.67 | Moderate | Fair | Poor | Poor | Moderate or good |
| CA_{in} - CA_{av}             | 0.17                           | 0.10 | 0.24 | Poor | Poor | Poor | Poor | Poor |
| CA_{in} - MED                 | 0.53                           | 0.43 | 0.62 | Moderate | Fair | Poor | Poor | Poor |
| CA_{in} - WDPT_{me}           | 0.68                           | 0.60 | 0.75 | Substantial | Good | Moderate or good | Moderate or good |
| CA_{in} - WDPT_{av}           | 0.22                           | 0.14 | 0.41 | Fair | Poor | Poor | Poor | Poor |
| CA_{av} - WDPT_{me}           | 0.44                           | 0.32 | 0.56 | Moderate | Fair | Poor | Poor | Poor |
| CA_{av} - MED                 | 0.24                           | 0.16 | 0.31 | Fair | Poor | Poor | Poor | Poor |
| CA_{av} - WDPT_{me}           | 0.26                           | 0.19 | 0.33 | Fair | Poor | Poor | Poor | Poor |
| WDPT_{av} - MED               | 0.76                           | 0.68 | 0.83 | Substantial | Excellent | Excellent | Excellent | Excellent |