Dendrochronologia 52 (2018) 131–145

Contents lists available at ScienceDirect

Dendrochronologia

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

Original article

Evaluating the key assumptions underlying dendro-provenancing: How to spruce it up with a scissor plot

Urs Gut[a,b,*]

[a] Universität Zürich, Institut für Archäologie, Fachbereich Prähistorische Archäologie, Karl Schmid-Strasse 4, 8006 Zürich, Switzerland
[b] ETH Zürich, Department of Environmental System Science, Institute of Terrestrial Ecosystems, Forest Ecology, Universitätstrasse 22, 8092 Zürich, Switzerland

ARTICLE INFO

Keywords:
Dendrochronology
Method
Cross-validation
Nearest neighbor classification
Elevational transect
Proximity measure

ABSTRACT

The pairwise statistical comparison of ring-width series is the basic analysis of dendro-provenancing studies. It is assumed that statistical proximity indicates similar provenance, but this assumption often remains untested. Especially for small areas with high topographic complexity, it is unknown to what extent statistical proximity and geographical provenance are correlated.

In this paper, dendro-provenancing is framed as a search for statistical Nearest Neighbors. The 'k-Nearest Neighbors leave one-out cross-validation' process (k-NN) is proposed as a method for validating dendro-provenancing approaches. Furthermore, it allows researchers to consistently compare and evaluate different proximity measures with respect to their suitability for dendro-provenancing. The validation process is demonstrated on a data set of 401 ring-width series of Norway spruce (*Picea abies (L.) H. Karst.*) encompassing 15 sites along elevational gradients in north-eastern Switzerland. Moreover, a new type of plot, the so-called scissor plot, is introduced to visualize the k-NN validation process.

Results indicate that dendro-provenancing depends heavily on differences in between sites high-frequency signal. Mean classification success for the relevant stages of the k-NN ($C\bar{S}R_{open}$)[1] ranged from 71.8% to 79.2% for the best performing measures. Classification errors occurred mainly between sites at elevations of 1000–1198 m a.s.l. At all other elevations and between different regions of the study area, only moderate differences in classification performance were detected. Thus, the results indicate that dendro-provenancing may be principally feasible even in a small region as studied here.

1. Introduction

Knowing the site of tree growth for timber in archaeological or historical structures and artefacts (e.g., buildings, paintings, ships, etc.) provides crucial information for reconstructing timber trade routes, determining the provenance and authenticity of art historical objects, or estimating forcing factors on tree growth (Wazny, 2002; Eissing and Dittmar, 2011; Jansma et al., 2014; Hellmann et al., 2017). The bulk of this so-called *dendro-provenancing* relies on pair-wise comparisons between ring-width series of unknown provenance and chronologies or single series representing potential sites of origin. Therefore, local reference chronologies form the backbone of dendro-provenancing. Reference chronologies may either be established from living trees or, for provenancing of historical and archaeological timber, they need to be constructed from series of historical or archaeological objects that are thought to represent local timber-sources. The proximity between a reference of known and an object of unknown provenance is usually expressed in statistical terms by calculating either *t*-values or Gleichläufigkeit (percentage of common signs of year-to-year growth changes; cf. Bridge, 2012; Buras and Wilmking, 2015). The spatial distribution of matches is then visualized on maps, which allow for narrowing down the area of provenance to the best matches (Eckstein and Wrobel, 2007; Daly and Nymoen, 2008). Such dendro-provenancing relies on three assumptions:

1. Tree growth varies sufficiently within the study area, thus causing the formation of regionally or locally characteristic ring-width patterns.
2. The (dis-)similarity of tree growth can be quantified by statistical measures of proximity.
3. Highest statistical proximity indicates closest geographical neighbors.

* Correspondence to: Universität Zürich, Institut für Archäologie, Fachbereich Prähistorische Archäologie, Karl Schmid-Strasse 4, 8006 Zürich, Switzerland.
E-mail address: u.gut@posteo.de.

[1] $C\bar{S}R_{open}$ is a figure for the classification success rate of a k-NN and is described in detail in the Methods section.

https://doi.org/10.1016/j.dendro.2018.09.008
Received 25 April 2018; Received in revised form 13 September 2018; Accepted 18 September 2018
Available online 22 September 2018

A review on dendro-provenancing was provided by Bridge (2012). Recently, the 'classical' method described above has been refined. For example, Jansma et al. (2014) combined topographical, geomorphological and soil type information to reconstruct potential sites of origin for *Quercus petraea* and *Quercus robur* in the Roman epoch. Similarly, Eissing and Dittmar (2011) used maps of the potential natural vegetation of Thuringia and Bavaria to determine areas for medieval harvesting of *Picea abies* and *Abies alba*. In addition, they consulted historical sources on timber rafting and considered the hydrological system upstream of a city to locate likely sites of timber supply (Eissing and Dittmar, 2011). However, the original assumptions stated above remain crucial for both studies (Eissing and Dittmar, 2011; Jansma et al., 2014). For a very recent approach of assessing the reliability of best statistical matches, see Drake (2018), who developed a probabilistic framework for dendro-provenancing based on statistical hypothesis testing and Bayesian inference.

While the quantification of similarity in tree growth using proximity measures constitutes a virtually unchallenged assumption, the first and third assumptions are discussed controversially (Bridge, 2000; Haneca et al., 2005; Savva et al., 2006; Garcia-Gonzalez, 2008; Eissing, 2007; Eissing and Dittmar, 2011; Jansma et al., 2014). Concerning the first assumption of regional growth patterns, cluster analysis has sometimes been used to investigate statistical proximity in relation to the spatial distribution of site conditions, albeit often with limited success (Bridge, 2000; Haneca et al., 2005; Savva et al., 2006; Garcia-Gonzalez, 2008). Although Garcia-Gonzalez (2008) found that the clusters actually represent specific ecological conditions, in the studies by Bridge (2000) and Haneca et al. (2005) the spatially diffuse results of the cluster analysis could not be correlated with site factors. Savva et al. (2006) found that the clearest clusters followed elevational belts. Moreover, motivated by studies that showed distinct elevation-specific climate-growth relationships, models predicting the elevational provenance of ring-width series were formulated (Wilson and Hopfmüller, 2001; Frank and Esper, 2005; Eissing and Dittmar, 2011; Dittmar et al., 2012; King et al., 2013; Kolář, Čermák et al., 2017; Lyu et al., 2017). Besides elevation-specific differences, results of pointer year studies suggest distinct regional to site-specific differences in tree growth (Dittmar and Elling, 1999; Rolland et al., 2000; Neuwirth et al., 2004, 2007).

The third assumption was discussed even more controversially. Eissing and Dittmar (2011) pointed out that imported timber was frequently used for the construction of the larger, communal buildings in many old town centers of the medieval and modern epoch. Hence, in such cases reference chronologies for dendro-provenancing cannot be thought of as reflecting local tree growth (Eissing, 2007; Eissing and Dittmar, 2011; Jansma et al., 2014). Best matches to such references must be scrutinized and interpreted critically. Also, a close and strong relation of geographical and statistical proximity is not granted from an ecological point of view (Haneca et al., 2005; Bridge, 2012; Fowler and Bridge, 2015). For example, Bridge (2000) studied oak stands in eastern England and compared oak ring-width series within the studied sites as well as with other British sites, suggesting that series from Hockley Woods (Essex) featured higher statistical proximity when compared to a 270-km distant site chronology from Peckforton (Cheshire), than when compared to much closer sites inside a 100-km radius around Hockley Woods. The contradiction was explained by the fact that both Hockley Woods and Pechforton were located on well-drained, steep slopes, i.e. they were ecologically quite similar in spite of a large geographical distance. Most previous dendro-provenancing studies were conducted in areas that are under an Atlantic climate regime and have only low topographical complexity, such as the Polish, Belgian and Baltic coastal areas and their hinterland (Bridge, 2012). There are only a few such studies in Alpine and pre-Alpine environments (Eissing and Dittmar, 2011). In these latter environments, the complex topography may be advantageous for dendro-provenancing. However, even when there is high diversity of site conditions and micro climate, statistical neighbors may be located at distant sites, where similar environmental factors limit tree growth (Bridge, 2000; Haneca et al., 2005; Boschetti-Maradi and Kontic, 2012).

Inspite of all controversies, dendro-provenancing remains popular and is widely applied. Its core method (i.e., the provenancing via pairwise statistical comparisons), however, has never been evaluated in a small-scale mountain environment. Thus, the objectives of this study are:

1. To develop a statistical procedure for validating the three fundamental dendro-provenancing assumptions.
2. To investigate the suitability of different proximity measures for dendro-provenancing.
3. To assess the most problematic provenancing errors encountered to be able to determine the growth signal that is relevant for dendro-provenancing and thus provide a basis for future research dedicated to a better understanding of this signal in Alpine and pre-Alpine environments.

2. Materials and methods

2.1. Tree-ring data

779 increment cores of Norway spruce (*P. abies* (L.) H. Karst.) were collected between winter 2015 and fall 2016. At 15 sites in the foothills and mountains of the north-eastern Swiss Alps 15–32 healthy spruce with a *diameter at breast height* (DBH) ≥ 30 cm were sampled (Table 1, Figs. 1, A1). The sampling plots were relatively small, ranging between 1110 m^2 and 7840 m^2 and were chosen to represent typical site conditions of the area. For most sites, two cores were extracted per tree. In Sihlwald-Streuboden (*sw*), however, a pilot study was conducted where only one core per tree was acquired.

The measurement of total ring width and cross-dating were performed using standard dendrochronological procedures (Cook and Kairiukstis, 1990; Speer, 2010). Only a few ring-width series had to be excluded from further analysis, after they had failed final cross-dating checks with the program COFECHA (Grissino-Mayer, 2001). Finally, for trees with more than one series of radial measurements, mean ring-width series were calculated. Thus, a data set of 401 ring-width series was created (Table 2, ring-width data in the Online Supplementary Material).

2.2. Dendro-provenancing as a k-Nearest Neighbor Analysis (k-NN)

Dendro-provenancing can be framed as a *Nearest Neighbor (NN) Analysis* (Cover and Hart, 1967; Schmitt, 2006). In One Nearest Neighbor Analysis (1-NN), for example, the object being classified is assigned the same class as its statistical NN. However, in order to evaluate the robustness of a classification, NN Analysis may relate to more than just the 1-NN and inquire the consistency of classifications based on any number k of NN. If $k = 10$, for example, the 10 NN are considered for classification. Consequently, the object (ring-width series) is assigned the same class (site label) as the most abundant class among the 10 NN. This *k-Nearest Neighbor* (k-NN) approach is pursued here. The procedure involves the following steps:

1. A ring-width series is 'anonymized'. All other series of the data set are assigned classes corresponding to their site provenance.
2. A proximity measure is chosen and values are calculated for each pairwise comparison between the anonymized series and all other series of known provenance.
3. Proximity values are sorted in ascending order and ranked accordingly. The number of highest ranking values (k) is chosen and the classes of the respective NN are evaluated. The most abundant class among those k NN determines the class attributed to the anonymized series. Thus, for each possible k (here, $k \in \{1, \ldots 400\}$), the class of the anonymous series is predicted.

Table 1
Site factors. Elevation in m a.s.l. Slope in degrees. Determination of root penetration depth (root pen. depth), nutrient storage capacity (nutr.stor.), and waterlogging (water log.), according to digital soil suitability map of Switzerland (Federal Office for Agriculture: map.geo.admin.ch, 17.8.2017). Duration of growing season in days (grow.s.), annual temperature average in C° (t.av.), and annual precipitation sums in mm (precip.), derived by calculating the mean for the period 1930–2010. Growing season calculated according to the ETCCDI definition (etccdi.pacificclimate.org, 7.9.2018) with *R* Package *climdex.pcic* (Bronaugh, 2018). Monthly and daily mean temperatures and precipitation sums for 1930–2010 were provided by the Land Use Dynamics Research Group at WSL. These data had been derived by a spatial interpolation of data from the MeteoSwiss network using DAYMET (Thornton et al., 1997) to a grid with cell size of 1 h.

Site	Region	Elevation	Slope	Exp.	Root pen.depth	Nutr.stor.	Water log.	Grow.s.	t.av.	Precip.
hw: Glarus-Haltenwald	Linth	627	18	NE	Medium	Medium	Moist	236	8.7	1522
nb: Sool-Nuebaennli	Linth	845	40	S	Very superficial	Very Low	No moist	224	7.6	1603
how: Sool-Hohwald	Linth	1022	41	S	Very superficial	Very low	No moist	213	6.7	1712
gand: Elm-Gandwald	Linth	1180	27	E	Superficial	Low	No moist	212	6.6	1518
ww: Spiringen-Waengiwald	Linth	1707	19	N	Very superficial	Very low	No moist	144	2.4	2019
rw: Elm-Raminerwald	Linth	1723	26	SW	Superficial	Low	No moist	146	2.5	1924
bw: Schmerikon-Bannwald	Obersee	472	15	N	Deep	Medium	Low wet	242	9	1487
ew: Eschenbach-Eggwald	Obersee	618	14	NW	Medium	Low	Moist	234	8.4	1546
sb: Gommiswald-Steibruch	Obersee	856	10	NW	Deep	Medium	Low wet	213	6.8	1853
chw: Eschenbach-Cholwald	Obersee	1106	20	N	Medium	Good	Low wet	192	5.5	1853
sw: Sihlwald-Streuboden	Sihl	646	6	NE	Deep	Good	Low wet	233	8.3	1357
fri: Feusisberg-Friesischwand	Sihl	829	15	N	Medium	Medium	Low wet	219	7.2	1660
kar: Unteriberg-Karenstockwald	Sihl	1000	26	NW	Superficial	Low	Moist	193	5.5	2049
gw: Oberiberg-Gschwaendwald	Sihl	1198	13	SE	Superficial	Medium	Wet	194	5.5	2043
furg: Alpthal-Furggelenstock	Sihl	1506	16	E	Superficial	Medium	Wet	171	4	2243

Fig. 1. Map of sites (46° 48′ 37″ – 47° 29′ 24″ N, 8° 26′ 50″ – 9° 16′ 23″ E). Full names of sites see Table 1. Reproduced by permission of swisstopo (BA17135).

Table 2
Chronology and signal strength statistics as well as site-wise mean classification success rate (*CSR*) for *k-NN* stages $k = 40$ till $k = 1$ conducted with the measure tHO (Table 3) on the complete data set. Mean replication was calculated excluding the years of a chronology with a replication < 2.

Site	*n*	Mean s.length	First yr.	Last yr.	Mean replic.	Rbar	EPS	Mean CSR
bw	27	126.33	1885	2016	25.84	0.32	0.92	87.69
chw	24	93.25	1860	2016	14.60	0.37	0.90	29.06
ew	26	105.85	1875	2016	20.19	0.36	0.92	75.38
fri	31	115.00	1880	2016	26.02	0.47	0.96	93.15
furg	19	210.63	1645	2015	11.23	0.42	0.89	81.18
gand	25	104.84	1890	2015	20.96	0.43	0.94	89.10
gw	15	127.00	1795	2015	13.38	0.38	0.89	48.33
how	23	107.78	1897	2016	20.82	0.49	0.95	62.61
hw	30	130.23	1878	2014	28.72	0.39	0.95	98.67
kar	30	81.90	1881	2016	21.36	0.39	0.93	59.17
nb	30	155.30	1851	2016	28.07	0.49	0.96	97.42
rw	32	160.16	1773	2015	21.09	0.48	0.95	95.94
sb	30	66.77	1938	2016	25.99	0.34	0.93	85.92
sw	30	97.40	1904	2014	27.27	0.26	0.90	71.92
ww	29	184.69	1755	2015	20.67	0.42	0.94	88.19

4. The classification is checked by comparing the true class (i.e., the site of provenance) of the anonymized series with its predicted class.

The k-NN is based on individual ring-width series. Reference chronologies are not used because such chronologies implicitly assume that all individual series share a site-specific growth signal, which is one of the key assumptions (c.f. Introduction) that is investigated here.

2.3. Measures of proximity

In cluster analysis, the term *proximity measure* is used to denote measures of distance, similarity and dissimilarity (see Everitt et al., 2011, p. 43). Similarly, in this paper the term *proximity measure* is used to refer to any value calculated to express statistical proximity, be it measures of distance, similarity or dissimilarity.

Most of the measures applied in this study (Table 3) are well known and frequently used in dendrochronology. Thus, the reader is relegated to the cited references and Supplementary Online Material for more information.

Besides the measures that are based on the agreement of the growth-sign (GL and FGGL, Table 3), two correlation coefficients were investigated: Pearson's correlation coefficient (Pearson, 1895) and Spearman's rank correlation (Hollander and Wolfe, 1973; Best and Roberts, 1975; Table 3). For both correlation coefficients, *t*-values can be calculated (Kendall and Gibbons, 1990). Thus, the dependence on sample size, i.e. years of overlap, of the correlation coefficient is reduced (Wigley et al., 1987). No *p*-values are calculated here because classic dendro-provenancing has almost always been done based on *t*-values (Wazny, 2002; Haneca et al., 2005; Eckstein and Wrobel, 2007; Daly and Nymoen, 2008; Eissing and Dittmar, 2011; Hellmann et al., 2017).

Both correlation coefficients are sensitive to trends and other low frequency fluctuations (Wigley et al., 1987), thus, preprocessing is necessary. However, growth patterns of approximately 5 to 20 years in length, i.e., medium-frequency patterns, are possibly relevant for provenancing, as they may reflect similar local forest management or other growth reactions due to spatially limited disturbances, e.g. storm damages. Thus, several k-NNs were calculated from the raw ring-width series. Additionally, low-frequency tolerant preprocessing methods, such as DET and SPL67pct (Table 3), were applied. All calculations and statistical analyses were done using the statistical software *R* version 3.4.3 (R Core Team, 2017). The *R* code for the main analyses is provided in the Online Supplementary Material.

2.4. Cross-validation using scissor plots

2.4.1. The scissor plot

Because the true class (i.e., the site of provenance) of the anonymized series is known, classifications accomplished by k-NN can be validated rigorously by iteratively anonymizing each ring-width series of the data set once, then predicting its class (i.e., site) for different numbers of *k*. This kind of cross-validation method is known as *leave-one-out cross-validation* (Arlot and Celisse, 2010). From the cross-validated classifications, the *classification error rate* (CER) is derived by dividing the number of wrong classifications by the total number of classification trials.

The probability for a correct classification rises if more NN come from the same site as the anonymized ring-width series. Thus, these *on-*

Table 3
Overview of proximity measures.

Abbreviation	Correlation	*t* statistic	Preprocessing/transformation	Reference
r	Pearson	-none-	-none-	Pearson (1895)
t	Pearson	Yes	-none-	Edgell and Noon (1984)
tHO	Pearson	Yes	Log first differences	Hollstein (1980)
tDIFF	Pearson	Yes	First differences	Stock and Watson (2015)
tBP	Pearson	Yes	5 year running mean	Baillie and Pilcher (1973)
tAR	Pearson	Yes	ar-model	Brockwell and Davis (1996)
tSPL67pct	Pearson	Yes	spline 2/3 series, 0.5 freq. resp.	Cook (1981)
tSPL30yrs	Pearson	Yes	spline 30 years, 0.5 freq. resp.	Cook (1981)
tSPL10yrs	Pearson	Yes	spline 10 years, 0.5 freq. resp.	Cook (1981)
tDET	Pearson	Yes	Deterministic detrending	Fritts (1976)
tARS	Pearson	Yes	Double-detrending & ar-model	Cook (1985)
s	Spearman	-none-	-none-	Best and Roberts (1975)
ts	Spearman	Yes	-none-	Kendall and Gibbons (1990)
tsHO	Spearman	Yes	Log first differences	Hollstein (1980)
tsDIFF	Spearman	Yes	First differences	Stock and Watson (2015)
tsBP	Spearman	Yes	5 year runnig mean	Baillie and Pilcher (1973)
tsAR	Spearman	Yes	ar-model	Brockwell and Davis (1996)
tsSPL67pct	Spearman	Yes	spline 2/3 series, 0.5 freq. resp.	Cook (1981)
tsSPL30yrs	Spearman	Yes	spline 30 years, 0.5 freq. resp.	Cook (1981)
tsSPL10yrs	Spearman	Yes	spline 10 years, 0.5 freq. resp.	Cook (1981)
tsDET	Spearman	Yes	Deterministic detrending	Fritts (1976)
tsARS	Spearman	Yes	Double-detrending & ar-model	Cook (1985)
GL	-none-	-none-	Falling (-) or ascending (+) yearly intervals	Huber (1943)
FGGL	-none-	-none-	Like GL but class 1 intervals (Table A1) excluded	von Jazewitsch (1948)

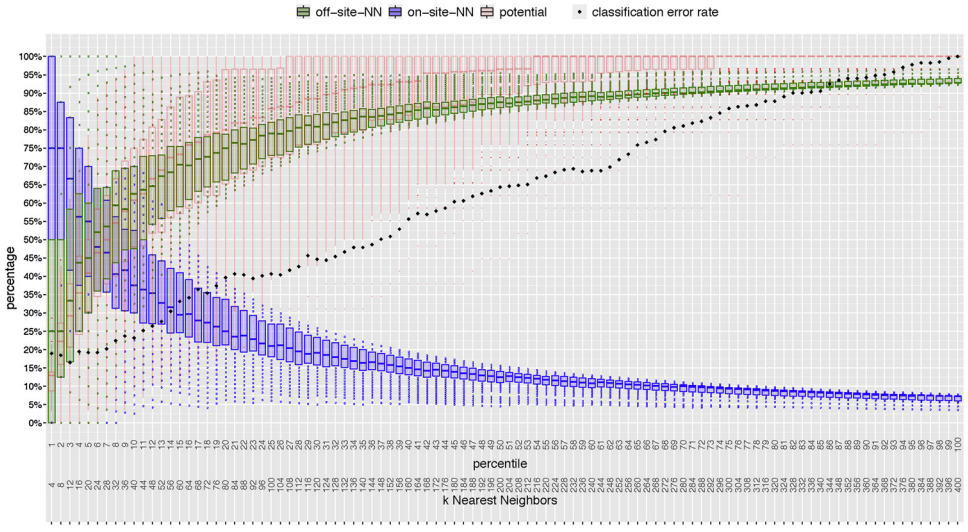

Fig. 2. Scissor plot for measure tHO; complete data set. Scissor plots visualize the leave-one-out cross-validation process. During this process, each ring-width series of the data set is anonymized once and values for on-site-NN/off-site-NN ratio and potential ratio are calculated (ratios plotted as percentages on *y*-axis, see Methods for details). As the complete data set encompasses $n = 401$ ring-widths series, the box-plots represent 401 values. Of the total 400 box-plots that would result from each possible setting for *k*, only 100 box-plots are visualized on the *x*-axis as the figure would get too large otherwise. This subset is called k %-NN (see Methods for details). There is no box-plot for the classification error rate as this rate is a single number, which equals the percentage of wrongly classified series at each setting for *k*.

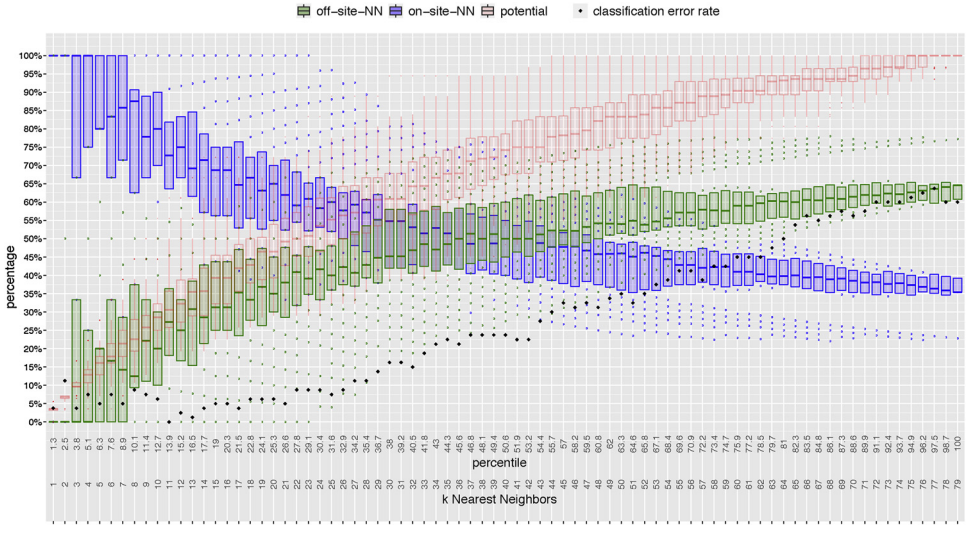

Fig. 3. Scissor plot for measure tHO; high-elevation subset.

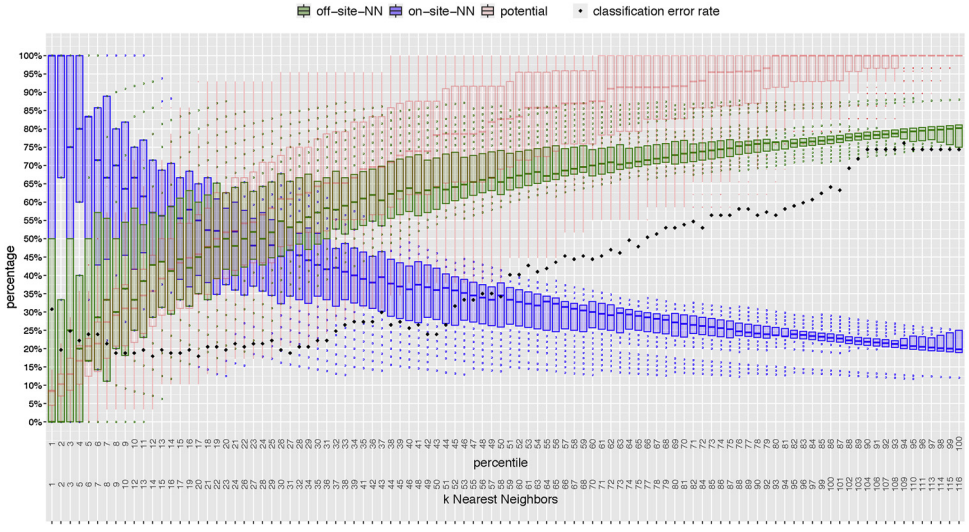

Fig. 4. Scissor plot for measure tHO; medium-elevation subset.

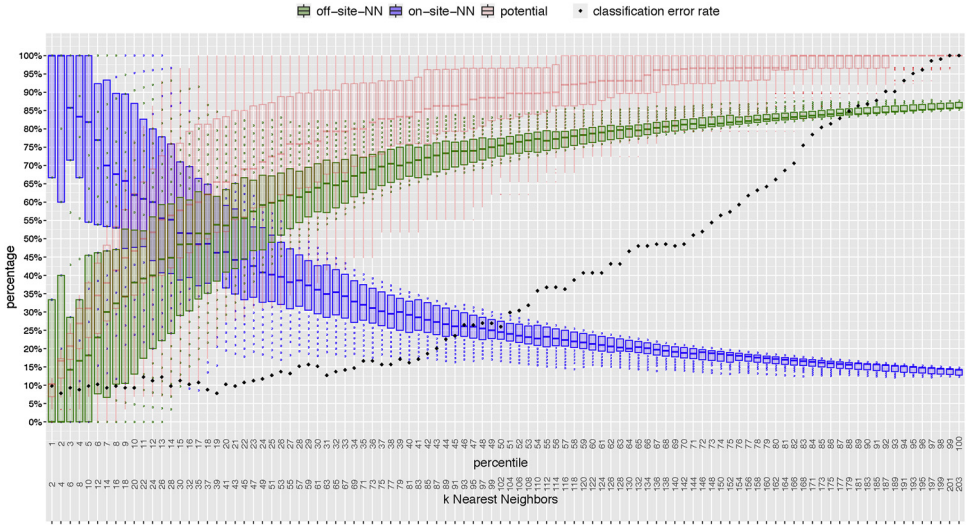

Fig. 5. Scissor plot for measure tHO; low-elevation subset.

site Nearest Neighbors (on-site-NN) are traced for all numbers of k and calculated for all iterations of the leave-one-out cross-validation (i.e., 400 iterations for the complete data set). Box plots are particularly useful to visualize the spread of the percentages of on-site-NN over all the iterations of the leave-one-out cross-validation (Fig. 2). Analogously, the number of NN coming from a different site than the anonymized ring-width series can be visualized. These NN are called *off-site Nearest Neighbors* (off-site-NN). Because the two box plot series of off-site-NN and on-site-NN resemble the crossing blades of a scissor, the result is called a *scissor plot* (Figs. 2–5). The third box plot series drawn for the *potential* is explained further below.

A complete k-NN encompasses $k_{max} = n - 1$ classifications, where n is the sample size. As n gets larger, the classifications become increasingly challenging to visualize. Hence, often only a selection is plotted. To create a subset for k-NNs with $k_{max} \geq 100$, the ranked proximity values are divided into percentiles and classifications are calculated for each percentile of NN. In this paper, such a subset of a k-NN is referred to as a k%-NN.

2.4.2. Rating a proximity measure with a scissor plot

The k%-NN data visualized on scissor plots (for data sets where $k_{max} \geq 100$) represents an arbitrary selection of the data underlying a complete k-NN. However, the information of a scissor plot can be expressed in numbers, and these can be calculated from the complete k-NN data without having to plot the data.

- The rating procedure is based on four indicators (Fig. 6):
  - $\text{opening}_{\text{ratio}}$
  - $\bar{\text{CSR}}$
  - $\text{on} - \text{site} - \text{NN}_{\text{ratio}}$
  - $\text{potential}_{\text{ratio}}$
- These indicators are summarized by the *rating score*. In addition, $\bar{\text{CSR}}_{\text{open}}$ provides a more intuitive rating figure (Fig. 6).

1. opening$_{ratio}$. An *opening* is defined as a classification stage $k \in \{1 \ldots k_{max}\}$ for which the 25% quantile of the on-site-NN box plot series does not overlap with the 75% quantile of the off-site-NN box plot series (for a barely opened scissor plot, cf. $k = 8$, Fig. 2; for a clear opening, cf. $k = 16$, Fig. 5). In these stages, more than 75% of all iterations of the k-NN allow for more on-site-NN than off-site-NN among the set of classifiers. Thus, the probability for a correct classification is high. The *opening$_{ratio}$* is the sum of open stages (N_{open}) expressed as a percentage of the total number of classification stages (k_{max}):

$$\text{opening}_{\text{ratio}} = \frac{N_{open}}{k_{max}} * 100 \tag{1}$$

2. $\bar{\text{CSR}}$. The stability of the classification error rate (CER) or of its complement, the classification success rate (CSR), is crucial for assessing the classification performance of a proximity measure. The overall classification performance is evaluated by calculating the *mean classification success rate* ($\bar{\text{CSR}}$):

$$\bar{\text{CSR}} = \frac{\sum_{i=1}^{k_{max}} (100 - \text{CER}_i)}{k_{max}} \tag{2}$$

3. $\text{on} - \text{site} - \text{NN}_{ratio}$. In general, the higher the $\text{on} - \text{site} - \text{NN}_{ratio}$, the higher the probability of a correct classification of an anonymized ring-width series. The on-site-NN box plot series of a scissor plot shows the spread of the on-site-NN for each stage (Fig. 2). For the rating of the on-site-NN over all stages, the following ratio is calculated, where $\widetilde{\text{on} - \text{site} - \text{NN}_i}$ is the median on-site-NN percent at the k-NN stage $i = k \in \{1 \ldots k_{max}\}$:

$$\text{on} - \text{site} - \text{NN}_{\text{ratio}} = \frac{\sum_{i=1}^{k_{max}} \widetilde{\text{on} - \text{site} - \text{NN}_i}}{k_{max}} \tag{3}$$

4. potential$_{ratio}$. The *potential* characterizes the portion of on-site-NN at a specific stage k of the k-NN (red box plot series in Fig. 2). It takes the value of 100% if all on-site-NN in the data set are found. A high value for *potential* is a sign of a strong common growth signal within the on-site ring-width series. It is also a sign of high signal difference between on-site-NN and off-site-NN. The *potential$_{ratio}$* is defined as follows, where $\widetilde{\text{potential}_i}$ is the median potential at k-NN stage $i = k \in \{1 \ldots k_{max}\}$:

$$\text{potential}_{\text{ratio}} = \frac{\sum_{i=1}^{k_{max}} \widetilde{\text{potential}_i}}{k_{max}} \tag{4}$$

Rating score. These four indicators range from a minimum of 0% to a maximum of 100%. Thus, the overall rating score is calculated as:

$$\text{ratingscore} = \frac{\text{opening}_{\text{ratio}} + \bar{\text{CSR}} + \text{on} - \text{site} - \text{NN}_{\text{ratio}} + \text{potential}_{\text{ratio}}}{4} \tag{5}$$

$\bar{\text{CSR}}_{\text{open}}$. Although the rating score provides the mean of the key indicators of a scissor plot, this mean remains difficult to interpret. The CSR presents a much more intuitive figure. Whenever an opening is present in a scissor plot, the *mean classification success rate from the first opening up to stage $k = 1$* ($\bar{\text{CSR}}_{\text{open}}$) can be calculated. However, for some k-NN, $\bar{\text{CSR}}_{\text{open}}$ cannot be calculated as there is no opening in the scissor

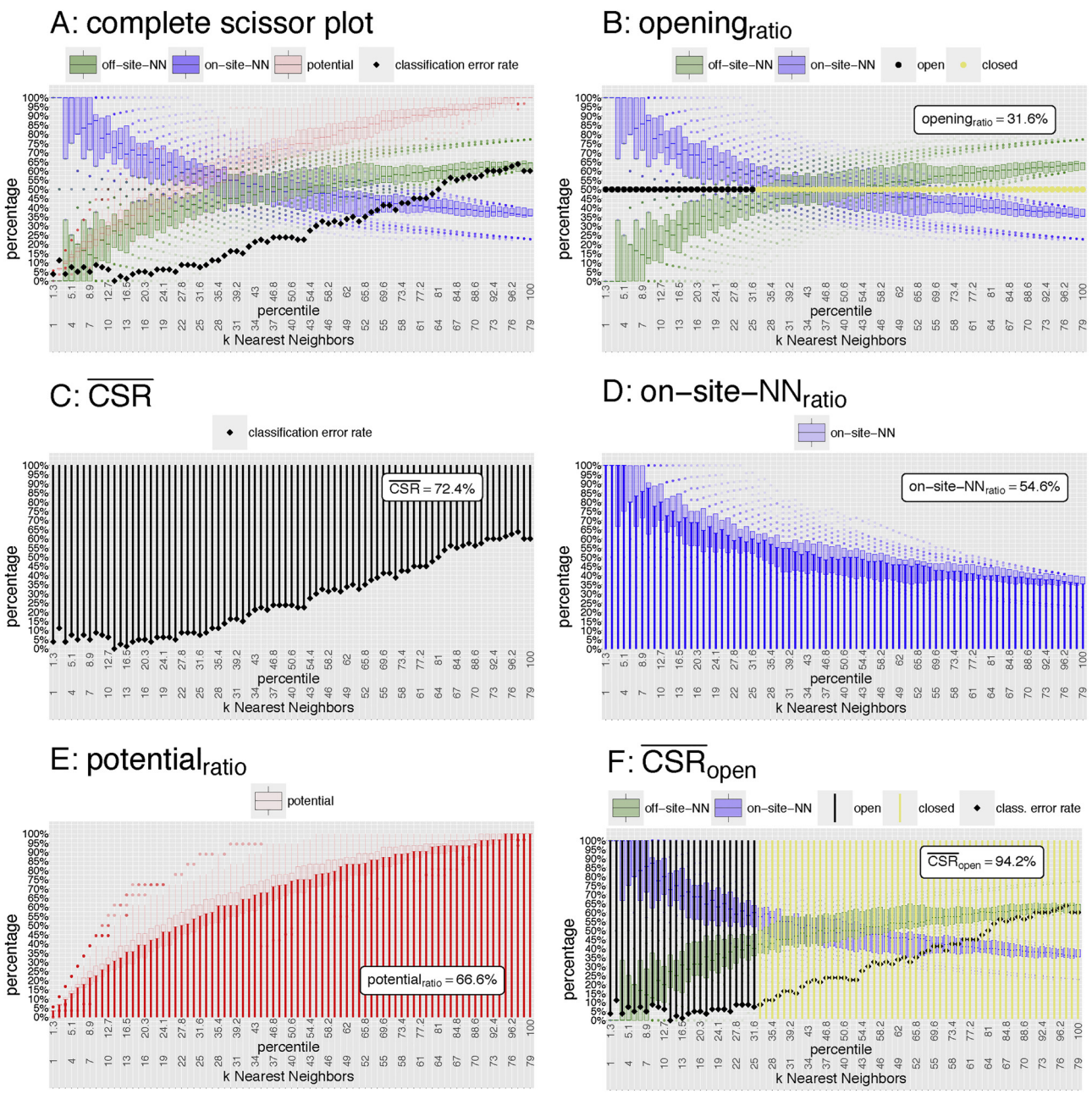

Fig. 6. Indicators used for rating a scissor plot (see Methods for detailed mathematical definitions). Subplot *A* shows a complete scissor plot (i.e., the plot for tHO; high-elevation subset) with all indicators on one plot. Subplots *B-F* illustrate how, for each indicator, the numeric score is connected to the visual representation on the scissor plot.

plot.

2.4.3. *Applying k-NN to the data set*

In a first step, k-NNs were calculated on the complete data set for every proximity measure (Table 3). In a second step, k-NNs were calculated on regional subsets. The three subsets, i.e. Sihl, Linth and Obersee, represent three different watersheds in the study area (for assignment of site to region cf. Table 1, Fig. 1). Because tree growth strongly depends on elevation (Frank and Esper, 2005; Savva et al., 2006; Dittmar et al., 2012; King et al., 2013; Kolář, Čermák et al., 2017; Lyu et al., 2017), in a third step k-NNs were calculated for elevation-specific subsamples of the data set. These subsamples correspond to three elevational bands, representing low (< 1000 m a.s.l.), medium (1000–1500 m a.s.l.) and high elevation sites (> 1500 m a.s.l.). To identify the site pairs that were most difficult for the classification, of each site 15 series were randomly selected and scissor plots were drawn for all possible site pairs. Then, the rating scores were visualized via a so-called *heat plot* (see Section 3.4).

Due to space constraints, scissor plots are shown for tHO only (Figs. 2–5). For the rest of the proximity measures, the results are presented based on the rating scores (Tables 4–6).

3. Results

Firstly, the results for the complete data set are presented. They provide the basis for discussing the suitability of different proximity measures for k-NN classification (see Section 4). Secondly, the results for the regional subsamples, and thirdly, the results for the elevational subsamples are presented. The two subsampling strategies aim at detecting regional and elevational differences in the k-NN classification

Table 4
Rating summary part 1. For information on proximity measures see Table 3. For details on the abbreviated rating figures see Method section.

All sites	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	0.2	0.2	0.8	0.8	0.8	0.8	0.8	0.8
$on-site-NN_{ratio}$	12.9	13.1	12.6	13.1	17.9	18.3	18.7	18.7	18.6	18.7	18.5	18.6
$potential_{ratio}$	72.8	73.1	72.8	73.3	86.7	87.4	88.2	88.1	87.6	88.2	87.5	88.2
$\bar{CSR}$	24.0	24.1	23.6	24.7	36.0	36.6	38.2	38.0	39.5	39.1	40.6	41.4
rating score	27.4	27.6	27.3	27.8	35.2	35.6	36.5	36.4	36.6	36.7	36.9	37.2
$\bar{CSR}_{open}$	-none-	-none-	-none-	-none-	76.3	75.3	79.2	76.9	75.9	76.7	78.5	76.0

All sites	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	0.5	0.5	-none-	-none-	-none-	-none-	-none-	-none-	0.8	0.5	-none-	-none-
$on-site-NN_{ratio}$	17.9	18.1	13.5	13.5	15.4	15.4	16.7	16.6	18.4	18.4	15.4	16.5
$potential_{ratio}$	87.8	87.7	73.7	74.3	80.7	81.2	84.2	84.1	87.6	87.9	83.0	84.9
$\bar{CSR}$	36.3	37.1	28.5	30.1	34.0	34.5	33.8	34.2	36.7	37.5	35.5	35.6
rating score	35.6	35.8	28.9	29.5	32.5	32.8	33.7	33.7	35.9	36.1	33.5	34.2
$\bar{CSR}_{open}$	75.3	72.0	-none-	-none-	-none-	-none-	-none-	-none-	75.7	71.8	-none-	-none-

Sihl subset	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	1.6	9.7	11.3	14.5	16.9	19.4	16.1	18.5	19.4
$on-site-NN_{ratio}$	37.7	38.1	37.8	38.2	43.3	43.1	44.9	44.7	45.4	45.2	44.8	44.9
$potential_{ratio}$	71.5	72.0	72.0	72.6	77.9	77.5	79.2	79.0	79.7	79.4	79.0	79.0
$\bar{CSR}$	47.9	48.5	47.5	48.4	55.1	54.3	57.8	57.5	61.2	59.6	60.9	60.6
rating score	39.3	39.6	39.3	40.2	46.5	46.5	49.1	49.5	51.4	50.1	50.8	51.0
$\bar{CSR}_{open}$	-none-	-none-	-none-	70.1	85.2	84.9	88.4	88.1	90.0	88.7	89.7	91.4

Sihl subset	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	14.5	13.7	0.8	0.8	4.0	4.0	4.0	2.4	10.5	10.5	2.4	-none-
$on-site-NN_{ratio}$	43.3	43.6	37.4	38.1	38.3	39.1	38.9	39.3	43.1	43.1	39.5	41.2
$potential_{ratio}$	77.7	78.2	71.3	72.1	71.4	72.2	72.1	72.7	76.9	77.2	73.5	75.7
$\bar{CSR}$	57.3	58.4	47.1	47.3	53.5	55.3	55.5	55.5	55.5	56.1	51.2	52.8
rating score	48.2	48.5	39.2	39.6	41.8	42.6	42.6	42.5	46.5	46.7	41.6	42.4
$\bar{CSR}_{open}$	87.4	87.3	68.5	75.2	78.9	75.8	79.2	76.0	85.0	85.8	73.7	-none-

Linth subset	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	0.6	-none-	-none-	17.3	19.6	16.7	18.5	16.1	16.1	17.3	17.9
$on-site-NN_{ratio}$	30.8	31.0	30.0	30.3	40.9	41.2	41.3	41.1	40.7	41.1	40.8	41.3
$potential_{ratio}$	74.7	75.0	73.2	73.7	86.1	86.3	86.4	86.6	86.1	86.5	85.8	86.5
$\bar{CSR}$	44.2	44.7	43.8	45.5	57.6	59.7	56.3	57.5	54.0	57.4	57.5	59.0
rating score	37.4	37.8	36.8	37.4	50.4	51.7	50.2	50.9	49.2	50.2	50.4	51.2
$\bar{CSR}_{open}$	-none-	77.5	-none-	-none-	89.2	89.8	90.8	89.5	87.8	88.5	91.6	90.2

Linth subset	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	14.9	15.5	1.2	1.2	5.4	3.0	10.7	10.1	16.1	20.2	1.8	13.7
$on-site-NN_{ratio}$	41.2	41.4	32.3	32.4	37.4	37.7	39.4	39.7	41.3	41.5	36.7	38.6
$potential_{ratio}$	86.1	86.7	75.5	75.7	81.9	81.8	84.6	84.7	86.5	86.8	83.0	84.5
$\bar{CSR}$	55.0	57.0	49.1	50.5	55.0	55.6	55.0	55.5	56.7	58.2	57.6	58.8
rating score	49.3	50.1	39.5	39.9	44.9	44.5	47.4	47.5	50.1	51.7	44.8	48.9
$\bar{CSR}_{open}$	89.8	90.5	72.2	70.2	82.5	81.8	86.8	86.5	89.4	90.2	73.5	87.0

Obersee subset	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	11.3	13.2	22.6	17.9	25.5	21.7	17.9	20.8
$on-site-NN_{ratio}$	37.1	37.1	36.8	37.0	44.5	45.9	46.5	45.4	47.1	46.5	46.1	46.2
$potential_{ratio}$	65.1	65.1	64.9	65.1	74.0	75.2	75.6	75.0	76.0	75.9	75.4	75.4
$\bar{CSR}$	50.4	50.8	49.6	50.1	63.9	67.1	64.9	68.6	69.3	71.3	66.9	69.9
rating score	38.2	38.2	37.8	38.1	48.4	50.4	52.4	51.7	54.4	53.9	51.6	53.1
$\bar{CSR}_{open}$	-none-	-none-	-none-	-none-	81.1	84.2	84.9	87.5	89.7	90.1	87.4	89.1

Obersee subset	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	14.2	6.6	-none-	-none-	1.9	-none-	4.7	2.8	20.8	20.8	-none-	0.9
$on-site-NN_{ratio}$	44.9	44.4	35.2	35.9	41.0	40.9	42.9	42.0	46.2	46.1	40.9	41.7
$potential_{ratio}$	74.1	74.2	62.1	63.0	69.4	69.5	70.9	71.0	75.1	75.6	71.8	71.4
$\bar{CSR}$	67.4	66.7	49.3	50.0	59.8	58.7	63.6	63.6	67.5	68.9	63.1	61.2
rating score	50.1	48.0	36.7	37.2	43.0	42.3	45.5	44.9	52.4	52.8	44.0	43.8
$\bar{CSR}_{open}$	86.1	80.2	-none-	-none-	70.7	-none-	80.1	76.4	88.3	87.1	-none-	66.7

Table 5
Rating summary part 2. For information on proximity measures see Table 3. For details on the abbreviated rating figures see Method section.

High-elevation sites	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	1.3	-none-	-none-	24.1	31.6	31.6	25.3	26.6	27.8	26.6	24.1
$on-site-NN_{ratio}$	43.5	44.4	43.6	43.8	54.2	56.4	54.6	54.6	53.2	54.0	54.7	55.8
$potential_{ratio}$	57.7	57.8	58.1	58.2	66.2	68.0	66.6	66.7	66.1	66.4	66.6	68.2
$C\bar{S}R$	49.0	49.2	49.9	50.2	69.2	72.7	72.4	71.5	69.4	70.6	71.1	71.4
rating score	37.6	38.2	37.9	38.1	53.4	57.2	56.3	54.5	53.8	54.7	54.7	54.9
$C\bar{S}R_{open}$	-none-	76.2	-none-	-none-	91.1	91.2	94.2	90.5	90.7	89.3	91.3	89.5

High-elevation sites	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	30.4	27.8	3.8	-none-	8.9	15.2	20.3	30.4	22.8	30.4	6.3	22.8
$on-site-NN_{ratio}$	55.6	56.4	47.0	46.4	51.1	52.4	54.8	56.8	54.9	56.2	49.6	52.3
$potential_{ratio}$	67.4	68.3	59.9	60.2	64.4	65.8	66.7	68.7	66.8	68.1	63.8	64.6
$C\bar{S}R$	70.8	72.0	53.9	55.6	62.5	65.7	68.1	70.5	68.6	71.9	65.4	70.2
rating score	56.0	56.2	41.1	40.5	46.7	49.8	52.4	56.6	53.3	56.6	46.3	52.5
$C\bar{S}R_{open}$	92.4	91.0	71.2	-none-	80.9	82.7	86.4	87.1	92.4	91.8	77.3	88.9

Medium-elevation sites	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	2.6	0.9	2.6	2.6	6.0	3.4	5.2	4.3
$on-site-NN_{ratio}$	32.5	32.7	32.6	32.8	37.8	38.0	38.8	38.8	37.5	37.9	37.6	38.1
$potential_{ratio}$	68.0	68.6	67.7	68.1	72.9	74.1	73.8	74.8	73.9	74.9	73.8	74.3
$C\bar{S}R$	37.3	36.6	36.7	36.1	56.7	56.6	59.5	60.2	66.0	63.4	63.4	62.2
rating score	34.5	34.5	34.3	34.2	42.5	42.4	43.7	44.1	45.9	44.9	45.0	44.7
$C\bar{S}R_{open}$	-none-	-none-	-none-	-none-	75.0	76.1	77.1	79.0	81.6	79.4	80.3	81.9

Medium-elevation sites	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	2.6	0.9	-none-	-none-	0.9	-none-	2.6	-none-	3.4	4.3	-none-	-none-
$on-site-NN_{ratio}$	38.0	37.1	32.7	33.4	35.0	33.7	36.5	35.5	38.7	38.4	31.4	33.7
$potential_{ratio}$	74.0	73.3	67.7	68.0	70.1	69.4	71.1	70.5	73.2	73.9	67.4	68.9
$C\bar{S}R$	57.1	55.7	42.5	44.5	52.5	52.1	57.3	56.6	58.5	59.4	48.6	59.1
rating score	42.9	41.7	35.7	36.5	39.6	38.8	41.9	40.7	43.5	44.0	36.9	40.4
$C\bar{S}R_{open}$	74.2	76.1	-none-	-none-	68.1	-none-	75.9	-none-	78.3	77.4	-none-	-none-

Low-elevation sites	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	1.5	3.0	6.4	5.4	6.9	7.4	5.4	5.4
$on-site-NN_{ratio}$	25.8	26.1	25.1	25.7	32.0	32.4	32.6	32.6	33.2	33.6	32.7	33.3
$potential_{ratio}$	72.2	72.5	71.0	71.8	79.2	80.6	80.7	80.7	81.3	81.9	80.5	81.7
$C\bar{S}R$	42.8	43.2	41.7	43.1	53.4	53.4	61.6	59.2	64.4	61.3	63.9	63.5
rating score	35.2	35.4	34.5	35.1	41.5	42.3	45.3	44.5	46.5	46.0	45.6	46.0
$C\bar{S}R_{open}$	-none-	-none-	-none-	-none-	79.2	82.5	89.3	87.0	89.4	88.2	87.3	85.7

Low-elevation sites	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	3.4	2.0	-none-	-none-	-none-	-none-	1.0	1.0	5.9	4.9	-none-	-none-
$on-site-NN_{ratio}$	31.3	31.9	23.9	24.4	27.0	27.6	28.9	29.0	32.3	32.4	27.8	29.7
$potential_{ratio}$	79.1	80.3	67.4	68.4	72.6	73.9	75.6	76.1	80.1	80.8	76.4	77.7
$C\bar{S}R$	57.1	56.3	44.3	46.0	50.2	49.3	48.5	48.9	56.7	56.9	53.5	54.7
rating score	42.7	42.6	33.9	34.7	37.5	37.7	38.5	38.8	43.7	43.7	39.4	40.5
$C\bar{S}R_{open}$	83.6	79.8	-none-	-none-	-none-	-none-	72.5	73.2	88.1	86.3	-none-	-none-

performance (discussed in Section 4).

Finally, the results emerging from investigating classification errors are presented. Because performance was low for the medium-elevation subsample (see Section 3.3), this subsample was further split into one consisting of the sites between 1000 and 1106 m, and the other between 1180 and 1198 m. Surprisingly, these subsamples posed no difficulties for the classification. To identify other subsets of sites that are particularly difficult for the classification, k-NNs were calculated for all site pairs. The effect of the thus identified, problematic sites on classification performance was quantified by dropping the respective sites from the complete data set and calculating k-NNs anew. Based on these results potential causes underlying the classification errors are discussed in Section 4.

3.1. Complete data set

All k-NNs calculated on the complete data showed poor performance using correlation-based proximity measures that lacked preprocessing (see upper two blocks of Table 4; i.e., r, s, t and ts). None of these measures showed an opening and the rating score was quite low (cf. Table 4). In fact, the score increased progressively the more rigorously the preprocessing removed the medium- and low-frequency variation. The highest rankings were accomplished by measures that were based on either AR, ARS, BP, SPL10yrs, HO or DIFF preprocessing whereas preprocessing methods that removed only little of the medium- to low-frequency variation such as DET or SPL67pct performed worse (Table 4; Fig. 2). Particularly, they showed no opening and reached low rating scores only.

Differences in performance depending on the correlation coefficient in use were rather small (cf. Table 4). Except for the DIFF methods, all measures using the Pearson correlation coefficient attained a slightly higher $C\bar{S}R_{open}$ than their Spearman counterparts. The rating score, however, was very similar for both correlation coefficients, and for both coefficients openings were limited to stages with small values of *k*

Table 6
Rating summary part 3. For information on proximity measures see Table 3. For details on the abbreviated rating figures see Method section.

Subset 1000–1106 m	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	1.3	2.6	3.9	2.6	26.3	39.5	27.6	36.8	35.5	38.2	40.8	38.2
$on-site-NN_{ratio}$	49.2	49.7	50.4	50.6	54.5	55.8	56.6	57.1	56.0	56.4	56.6	57.0
$potential_{ratio}$	62.5	62.7	62.6	62.7	69.7	71.9	70.5	71.3	70.5	71.9	71.1	71.2
$\bar{CSR}$	58.6	58.5	59.9	59.7	75.6	77.1	76.3	79.8	81.2	81.2	81.0	81.0
rating score	42.9	43.4	44.2	43.9	56.5	61.1	57.8	61.3	60.8	61.9	62.4	61.8
$\bar{CSR}_{open}$	71.0	72.7	72.0	73.0	87.9	88.3	88.5	92.7	94.9	93.9	92.8	92.7

Subset 1000–1106 m	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	31.6	31.6	5.3	7.9	18.4	11.8	22.4	26.3	28.9	34.2	13.2	21.1
$on-site-NN_{ratio}$	55.8	55.0	50.7	51.1	53.1	52.5	53.9	54.4	55.7	56.6	50.4	51.4
$potential_{ratio}$	69.7	70.2	65.8	66.7	66.9	67.1	67.0	67.7	69.5	70.8	66.9	68.3
$\bar{CSR}$	76.6	76.3	65.0	64.1	69.5	69.5	72.1	72.0	77.2	77.5	71.4	77.5
rating score	58.4	58.3	46.7	47.5	52.0	50.2	53.8	55.1	57.8	59.8	50.5	54.6
$\bar{CSR}_{open}$	89.2	91.1	77.7	79.7	81.6	82.1	86.4	84.1	90.6	89.2	81.1	87.0

Subset 1180–1198 m	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	69.2	69.2	69.2	69.2	46.2	33.3	59.0	61.5	69.2	69.2	69.2	69.2
$on-site-NN_{ratio}$	85.6	85.6	85.7	85.9	85.7	84.8	85.6	86.3	79.7	82.4	80.3	82.3
$potential_{ratio}$	70.3	70.3	70.1	70.1	66.9	65.9	67.3	68.1	67.0	67.5	68.0	68.0
$\bar{CSR}$	83.8	83.7	83.8	83.9	74.1	71.0	78.8	79.5	85.5	83.4	84.7	83.7
rating score	77.2	77.2	77.2	77.3	68.2	63.7	72.7	73.9	75.4	75.6	75.6	75.8
$\bar{CSR}_{open}$	93.3	93.2	93.5	93.6	83.8	79.1	89.7	89.2	95.7	92.7	94.6	93.1

Subset 1180–1198 m	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	69.2	69.2	61.5	69.2	69.2	69.2	61.5	61.5	41.0	51.3	56.4	46.2
$on-site-NN_{ratio}$	83.6	83.5	81.3	79.9	79.6	78.4	78.0	75.4	85.6	84.1	79.0	80.0
$potential_{ratio}$	68.2	67.5	68.4	68.8	66.9	66.6	66.2	64.0	66.8	66.0	65.1	65.8
$\bar{CSR}$	84.6	83.9	79.1	80.7	84.0	82.0	80.1	78.0	74.7	73.7	75.1	71.5
rating score	76.4	76.1	72.6	74.6	74.9	74.1	71.5	69.7	67.0	68.8	68.9	65.8
$\bar{CSR}_{open}$	94.4	93.4	89.9	89.3	93.5	90.6	91.1	87.7	88.1	81.5	82.8	77.5

Data without subset 1000–1106 m	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	1.9	2.8	4.3	4.3	5.0	5.0	5.0	4.0
$on-site-NN_{ratio}$	16.5	16.8	16.4	16.7	23.5	24.0	23.9	23.9	24.0	24.1	23.8	24.1
$potential_{ratio}$	73.8	74.1	73.3	74.0	88.5	89.4	89.4	89.9	89.2	90.0	88.8	89.6
$\bar{CSR}$	29.4	29.8	29.5	30.6	41.2	41.4	43.9	43.0	45.6	44.3	46.6	46.6
rating score	29.9	30.2	29.8	30.3	38.8	39.4	40.4	40.3	40.9	40.9	41.0	41.1
$\bar{CSR}_{open}$	-none-	-none-	-none-	-none-	81.1	83.6	89.3	87.7	89.3	88.7	88.5	86.5

Data without subset 1000–1106 m	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	3.4	2.8	-none-	-none-	-none-	-none-	0.9	0.9	3.7	4.0	-none-	-none-
$on-site-NN_{ratio}$	23.4	23.7	17.2	17.4	20.0	20.2	21.6	21.4	23.7	23.8	20.7	21.9
$potential_{ratio}$	89.5	89.9	74.7	75.4	82.5	83.3	86.1	86.2	89.1	89.5	85.5	87.1
$\bar{CSR}$	42.2	42.3	33.8	35.1	39.5	39.4	38.5	38.7	42.1	43.0	40.9	41.1
rating score	39.6	39.7	31.4	32.0	35.5	35.7	36.8	36.8	39.6	40.1	36.8	37.5
$\bar{CSR}_{open}$	87.2	83.4	-none-	-none-	-none-	-none-	76.4	75.6	87.5	85.9	-none-	-none-

Data without chw, gw, how, kar	r	t	s	ts	tBP	tsBP	tHO	tsHO	tDIFF	tsDIFF	tAR	tsAR
$opening_{ratio}$	-none-	-none-	-none-	-none-	3.2	3.9	5.2	5.2	5.8	5.5	5.2	4.5
$on-site-NN_{ratio}$	17.2	17.6	17.2	17.5	24.8	25.2	25.0	25.2	25.1	25.4	25.0	25.4
$potential_{ratio}$	72.8	73.2	72.5	73.1	88.5	89.3	89.2	89.6	88.8	89.6	88.4	89.1
$\bar{CSR}$	31.1	31.3	31.1	32.2	43.1	43.2	45.8	44.5	46.9	45.8	48.0	48.0
rating score	30.3	30.5	30.2	30.7	39.9	40.4	41.3	41.1	41.6	41.6	41.6	41.8
$\bar{CSR}_{open}$	-none-	-none-	-none-	-none-	83.1	85.4	91.0	89.0	90.0	89.1	88.9	86.9

Data without chw, gw, how, kar	tARS	tsARS	tDET	tsDET	tSPL67pct	tsSPL67pct	tSPL30yrs	tsSPL30yrs	tSPL10yrs	tsSPL10yrs	GL	FGGL
$opening_{ratio}$	3.6	2.9	-none-	-none-	-none-	-none-	1.0	1.0	5.2	5.5	-none-	0.3
$on-site-NN_{ratio}$	24.6	24.9	18.0	18.3	21.0	21.4	22.8	22.9	24.9	25.2	21.8	23.2
$potential_{ratio}$	89.0	89.4	73.9	74.6	81.8	82.6	86.1	86.2	89.2	89.4	85.3	87.0
$\bar{CSR}$	43.6	43.7	35.2	36.6	41.1	40.9	39.8	40.1	43.9	44.5	42.6	43.0
rating score	40.2	40.2	31.8	32.4	36.0	36.2	37.4	37.5	40.8	41.2	37.4	38.4
$\bar{CSR}_{open}$	86.8	83.9	-none-	-none-	-none-	-none-	76.9	77.0	89.5	87.9	-none-	75.4

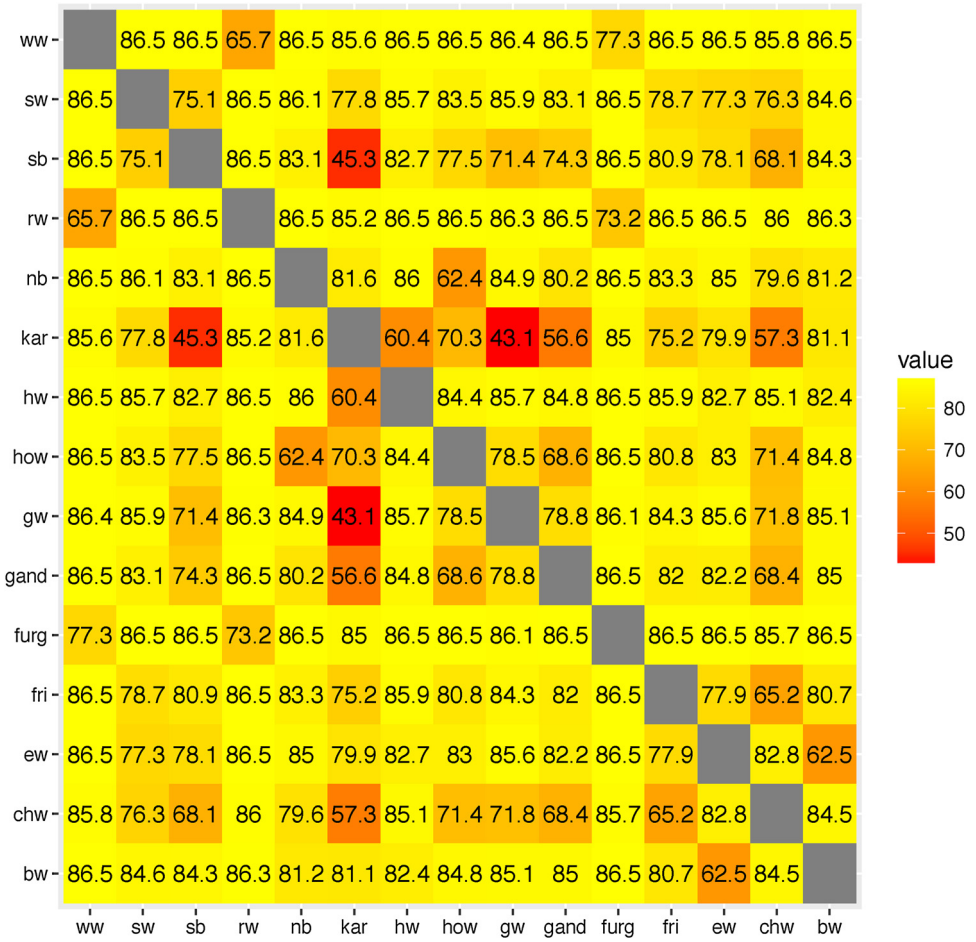

Fig. 7. Heat plot with rating scores accomplished by tHO. Rating scores were calculated from pairwise scissor plots that were drawn for all sites of the data set.

(≤ 5).

The performance of the sign-based proximity measures GL and FGGL lay somewhat above the performance of the measures that relied on DET or SPL67pct preprocessing. However, both GL and FGGL performed much better than the correlation-based methods that lacked preprocessing (i.e., *r*, *s* and *t*). Still, neither GL nor FGGL showed an opening. Additionally, the rating score of these two measures was distinctly lower than that of the correlation-based proximity measures using high-pass filters.

3.2. Regional subsamples

There were few differences in classification performance between the regions (see lower six blocks of Table 4). The subsample Linth had a higher k-NN performance than the other two subsamples, i.e. Sihl and Obersee. The more rigid detrending methods, e.g. SPL67pct, and the sign-based methods, especially FGGL, performed quite well with the Linth subset. For all measures, the overall rating scores were somewhat higher for the regional subsets than for the complete data set. Similar to the complete data set, correlation-based proximity measures that used high-pass filters performed best.

3.3. Elevation-specific subsamples

For the high-elevation subsample, k-NN classification was quite successful. Scissor plots opened for most of the measures (Table 5). Rating scores and $C\bar{S}R_{open}$ reached relatively high values for some measures that in other subsamples performed only poorly, such as tSPL67pct, tsSPL67pct, tDET, FGGL and GL. Again, high-pass filter based measures performed best, with some accomplishing $C\bar{S}R_{open} \geq 90\%$ (e.g. tHO, tAR, sAR, tSPL10yrs).

Within the medium-elevation subsample, k-NN classification issues were more common than within the other two elevation-specific subsamples (Table 5; cf. Figs. 3–5). The performance for the medium-elevation subsample dropped for all measures when compared to the performance reached for the high-elevation subsample. In addition, for most measures the performance was lower with the medium-elevation subsample compared to the low-elevation subsample. This was especially true for the best performing measures tHO, tsHO, tAR, tsAR, tDIFF, tsDIFF, tSPL10yrs and tsSPL10yrs.

On the low-elevation subsample, only correlation-based measures that relied on high-pass filters performed well (Table 5). Consequently, rating scores decreased rapidly the more low-frequency variation was left in an index series (SPL67pct, DET, ts, t, r, s). The sign-based measures both did not show any opening. FGGL again reached a distinctly higher rating score than GL. The rating score of FGGL even lay above that of the low-frequency tolerant, correlation-based measures (SPL67pct, DET, ts, t, r, s).

3.4. Classification errors

Heat plots of all high-performance measures (i.e., tHO, tsHO, tDIFF, tsDIFF, tAR, tsAR and tSPL10yrs) were very similar; thus only one example is shown (tHO, Fig. 7). The generally high rating scores indicated that most pairs of sites were distinguished easily via k-NN. However, there were also some lower rating scores, most of which were associated with the site Unteriberg-Karrenstock (*kar*). A second, less pronounced set of low rating scores was found for Eschenbach-Cholwald

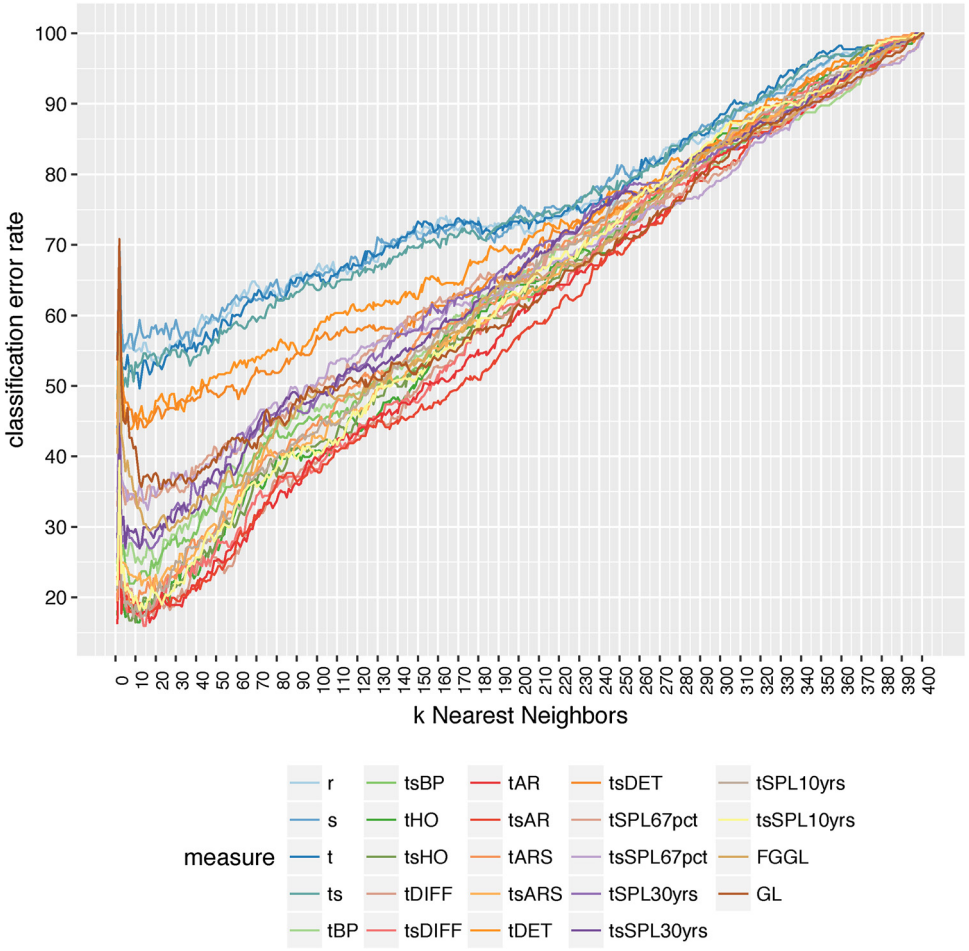

Fig. 8. Trend in classification error rate: Almost linear decrease until k ≈ 40; afterwards, only slow further decrease or even stagnation (except for the jump at k = 2).

(*chw*). Also, a somewhat low rating score resulted for the site pair Sool-Hohwald (*how*) and Sool-Nübännli (*nb*). These problematic cases formed a subset consisting of the sites at 1000–1106 m.

Generally, CER calculated on the complete data set decreased approximately linearly until $k \approx 40$. Afterwards, it decreased very slowly for most measures, or even stagnated (Fig. 8). An abrupt rise in CER occurred at stage $k = 2$ (Figs. 8 and A2) for all proximity measures. But overall, the stages $k = 40$ to $k = 1$ exhibited a relatively stable CER, and thus a relatively stable CSR, for most measures (Fig. 8). However, exactly in these stages, the sites *chw*, *kar* and *how* had a distinctly lower mean CSR than the others (Table 2), which coincided with the findings from the heat plot (Fig. 7). In addition, the site Gschwändwald (*gw*) attained a very low mean CSR although the heat plot indicated classification difficulties for the site pair *gw-kar* only. These four sites belong to the medium-elevation subsample, mostly misclassified to either sites on the same elevation, or sites that are located on the next lower elevational band (i.e., *fri*, *nb* and *sb*; Fig. 9). Moreover, some classification errors were caused by direct geographical neighbors (e.g., *how* series that were frequently misclassified as *nb* series; Fig. 9).

k-NN calculations excluding the sites *chw*, *how* and *kar* exhibited a rise in rating scores by an average of 3.4% (range 2.5% to 4.3%; blocks 5 and 6 in Table 6). In addition, $C\bar{S}R_{open}$ rose by an average of 10.8% (range 4.8% to 14.1%). However, the rating scores and $C\bar{S}R_{open}$ exhibited only a marginal further increase when the small sample of *gw* (n = 15) was excluded as well.

Classification errors occurred especially often in cases where the difference between the median of proximity values calculated for within-site comparisons and the median of proximity values calculated for between-site comparisons was low (e.g., for the site pairs *kar-sb* and *how-nb*, Table 7). In contrast, the classification worked well between pairs of sites for which the difference of median proximity values calculated for between-site and within-site comparisons was high (e.g., site pairs *kar-sw* and *ww-rw*, Table 7).

4. Discussion

The main objective of this paper is to present a consistent method for evaluating the three key assumptions of dendro-provenancing stated in the Introduction. The k-NN classification was designed as a method whose success is highly dependent on the validity of the three key assumptions. Thus, the k-NN classification results need to be discussed before any inference on the validity of the dendro-provenancing assumptions can be drawn.

4.1. Suitability of different proximity measures for dendro-provenancing

Most proximity measures tested here involved preprocessing of the raw ring-widths prior to calculating a proximity value. Thus, the k-NN classification performance indicates which preprocessing method is most successful in enhancing the growth signal that is relevant for dendro-provenancing.

Correlation-based proximity measures that relied on high-pass filters performed best for all regional and elevation-specific subsets (i.e., tHO, tsHO, tDIFF, tsDIFF, tAR, tsAR, tSPL10yrs and tsSPL10yrs). Interestingly, the differencing-based methods, i.e. HO and DIFF, were as efficient in enhancing the high-frequency variation as the very flexible

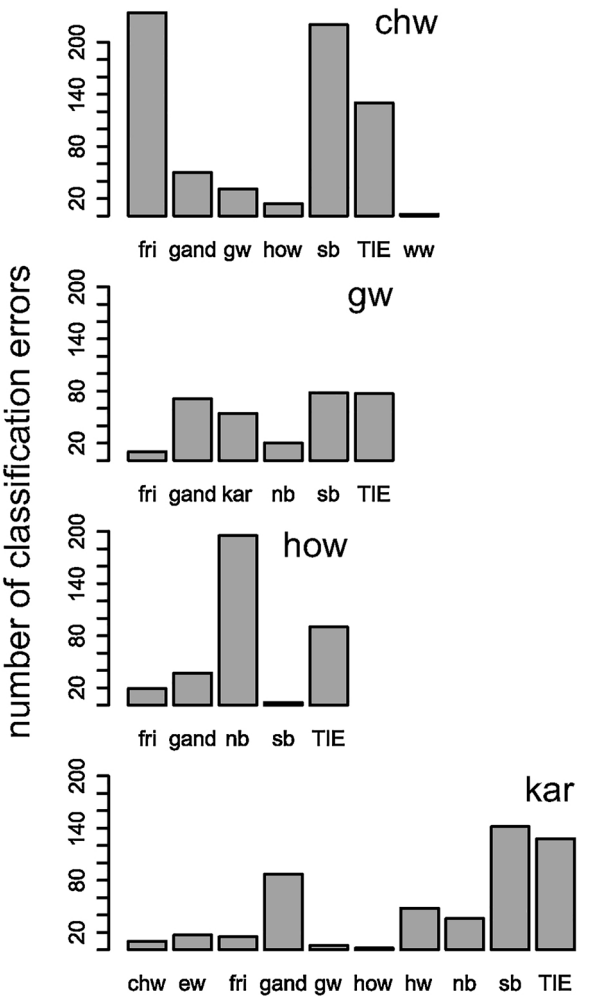

Fig. 9. Distribution of misclassified series per site. Upper right corner: site label of the site investigated (full names of sites see Table 1). The sites to which series have been misclassified are on the *x*-axis. The 'TIE' label denotes classification errors that resulted from a tie, i.e., no site label had a majority. The number of classification errors is on the *y*-axis and represents the sum of errors over the stages $k = 40$ till $k = 1$.

smoothing-splines of SPL10yrs or the auto-regressive modeling of AR. Using either the Pearson correlation coefficient or Spearman's rank correlation coefficient scarcely affected the classification performance. Although the Pearson correlation coefficient is known for its sensitivity to outliers (Edgell and Noon, 1984), this did not flaw the k-NN performance. In fact, the synchronous high frequency outliers (i.e., pointer years) to which the Pearson correlation coefficient is even more sensitive than the Spearman's rank correlation seem to occur often at the site level and thus are potentially provenancing-relevant.

The performance of ARS did not reach that of AR for the bulk of subsets, and the total data set. Hence, double detrending largely had a negative effect on classification performance. In contrast, although direct autoregressive modeling of raw ring-width series probably resulted in biased autocorrelation coefficients (Monserud, 1986; Monserud and Yamaguchi, 1989; Cook and Kairiukstis, 1990; Stock and Watson, 2015), the autoregressive residual series as used by tAR and tsAR were successful here (cf. Supplementary Online Material). Nevertheless, future studies could profit from the individual single detrending of each series and the subsequent use of autoregressive moving-average (ARMA) models, which often are more parsimonious than AR models (Monserud, 1986; Monserud and Yamaguchi, 1989; Speer, 2010).

Measures implementing the BP transformation frequently reached quite high ratings, but not in all settings. Especially on the medium-elevation subsample, they performed worse than HO, DIFF, AR and SPL10yrs. Facing these alternatives, the BP transformation is considered a suboptimal choice for enhancing the provenancing relevant signal. Moreover, running mean transformations have been shown to cause spurious correlation (see Monserud, 1986. This most likely caused the lower k-NN classification performance of BP.

The relatively low performance of the sign-based measures FGGL and GL further supports that high-frequency filters in combination with correlation coefficients should be used for enhancing the signal relevant for dendro-provenancing. The rating for FGGL was close to that of tBP for many subsets. Moreover, FGGL, which takes the amplitude of growth changes into account (cf. Supplementary Online Material), performed much better than GL, which relies purely on sign-agreement. This again underlines the importance of synchronous high-frequency outliers for dendro-provenancing.

Even GL, with its rather low rating scores, performed better than the correlation-based methods that preserved low-frequency variation or completely lacked preprocessing (i.e., DET, ts, t, r, s). Clearly, preserving the low-frequency variation did not pay off in terms of classification success, as demonstrated by the worsening performance of the sequence SPL10yrs, SPL30yrs, SPL67pct, DET, ts, t, r, and s. However, the 1180–1196 m subsample featured completely diverging persistence patterns, but it consisted of just two sites (*gw* and *gand*). In this exceptional case, the high-pass filtering removed the diverging persistence patterns and thus removed the feature that differed most between the ring-width series of both sites. Nevertheless, high-pass filter based methods performed just a little worse on this subsample than low-frequency tolerant methods. Moreover, persistence patterns, like those differentiating the sites of the 1180–1196 m subsample, lost their relevance for classification as soon as the data set was extended. Thus, in the data set studied, similar medium- to low-frequency patterns were not as site-specific as similarities in high-frequency variation. Correlation coefficients can only detect common trends and medium-frequency growth patterns in the time domain when the preprocessing preserves these patterns. At the same time, exactly these patterns (i.e., trends and medium frequency patterns) obscure the high-frequency correlation. Hence, other proximity measures are probably more suitable to reveal persistent growth variations that could be relevant for dendro-provenancing. This may be accomplished for example by measures that account for distances in the frequency domain or calculate distances between auto-regressive model coefficients (Hennig, 2016).

Table 7
Median *t*-values (tHO) for within-site and between-site pairwise comparisons. Due to space constraints, only these exemplary results are presented and are limited to tHO, as the findings were quite similar for all high-performance proximity measures (i.e., tHO, tsHO, tAR, tsAR, tDIFF, tsDIFF, tSPL10yrs and tsSPL10yrs).

site A	site B	within A	within B	between AB	within A - within B	within A- between AB	within B - between AB
kar	sb	2.97	4.31	2.57	−1.34	0.40	1.74
how	nb	5.91	7.82	5.33	−1.92	0.57	2.49
kar	sw	2.97	4.05	1.71	−1.08	1.27	2.34
ww	rw	6.87	6.86	5.79	0.01	1.08	1.07

4.2. Regional and elevation-specific differences in classification performance

The regional subsets consist of sites spread across all elevational belts of a region, whereas climate-growth relationships are known to be most pronounced along elevational gradients (Frank and Esper, 2005; Savva et al., 2006; Eissing and Dittmar, 2011; Dittmar et al., 2012; King et al., 2013; Kolář, Čermák et al., 2017; Lyu et al., 2017). Hence, differences in between-site growth variation within a regional subset are likely attributable to differences in elevation. As the classification performance was similar for all regional subsets, the effect of elevation on the between-site growth variation seems to be equal across the regions.

Within the elevational subsets, which comprise sites of all regions, series were successfully classified to their original sites by correlation-based measures that involve high-pass filtering. Thus, the high-frequency variation between sites of the same elevation differs strongly enough in most cases, except for the medium-elevation subsample.

4.3. Analysis of classification errors

Classification difficulties arose for all measures when the classification was based on two NN. In such instances just one off-site-NN needed to be among the pair of classifier-NN to cause a tie. Since the classification error rate increased dramatically at $k = 2$, many of the ring-width series must have had at least one off-site NN among their two NN.

Besides this $k = 2$ classification problem, notable difficulties were caused by medium-elevation sites that were misclassified to sites at 829–1198 m. Investigating the within-site and between-site median tHO values showed that the lack of signal difference between on-site and off-site series increased the frequency of classification errors. Thus, between-site variability was seemingly less pronounced between upper low-elevation and medium-elevation than at other elevational belts. Within the upper low-elevation and medium-elevation belt, classification errors were not exclusively confined to sites that exhibited similar site-factor combinations and/or were located in direct geographical proximity (such as the site pairs *kar-gw*, *how-nb*, *chw-sb* and *chw-fri*; cf. Table 1 and Figs. 7, 9). Problems also arose between distant sites with differing site-factor combinations (such as *kar-sb*, *kar-gand*; cf. Table 1 and Figs. 7, 9). This contrasts the findings of Bridge (2000) who found good matches between series from distant sites provided that they share key site factors.

Between-site high-frequency variability is thought to be controlled primarily by climatic factors and weather conditions (Fritts, 1976). Studies of the climate-growth relationship of spruce in the Bavarian forest (Dittmar and Elling, 1999; Wilson and Hopfmüller, 2001) may offer some preliminary explanation for the classification problems encountered here, as discussed below.

Wilson and Hopfmüller (2001) found that growth below 680 m was predominantly controlled by moisture availability. In their study, no significant correlations between growth and climate parameters were found for sites between 780 and 970 m. Even at ≥1070 m, site chronologies still exhibited a relatively weak climate signal. Unexpectedly, temperature had no dominant control upon growth at these higher sites.

Dittmar and Elling (1999) also investigated climate-growth relationships, but in their study the sites where no dominant controls on growth could be identified lay somewhat lower than those reported by Wilson and Hopfmüller (2001), i.e. at elevations between 600 and 800 m, and above 800 m temperature gradually became the dominant factor limiting growth. Below 600 m, water supply was found to predominantly control growth, which agreed with the finding by Wilson and Hopfmüller (2001).

Thus, although their findings agreed in some cases, the two studies found somewhat different elevational belts to be most challenging for interpreting climate-growth relationships. Hence, more research is needed to determine the elevational belt where such complex interactions occur within the region studied here. However, Wilson and Hopfmüller (2001) found that sites between 780–970 m shared much of the year-to-year variability with chronologies located at ≥1070 m (of these chronologies, four are located at 1070–1230 m and two are located at 1325–1420 m). This elevational belt with lower between-site variability and no significant correlations between growth and meteorological data matches well with the elevational belt where k-NN classification problems occurred here (829–1198 m).

5. Conclusions

Scissor-plot rating of k-NN offers a consistent method for evaluating the three fundamental dendro-provenancing assumptions prior to determining the provenance of wood. The successful k-NN classification of ring-width series to their original site indicates that growth of spruce diverges sufficiently within the study area to allow for the rise of site-specific high-frequency growth patterns. Thus, the first assumption of dendro-provenancing seems to be adequate for the data set studied here. Moreover, the results imply that similar growth can be quantified by statistical measures of proximity (the second assumption), and that best proximity values are closely correlated with geographical neighbors (except for classifications using $k = 2$), which is the third key assumption of dendro-provenancing.

The between-site variation of the high-frequency signal is paramount for a successful classification. Consequently, correlation-based measures that rely on high-pass filters performed best (tHO, tsHO, tDIFF, tsDIFF, tAR, tsAR, tSPL10yrs and tsSPL10yrs). Preserving the mid- to low-frequency variation did not have any positive effect on classification performance.

Classification problems arose where the difference between within-site and between-site signal was small. In the data set presented here, such small differences were predominantly limited to medium-elevation sites between 1000 and 1198 m. Moreover, the low performance of the classification using $k = 2$ illustrated that relying on a single setting for k may easily lead to unreliable classifications. The climatic and possibly ecological causes underlying the between-site growth variation need to be analyzed further.

Using scissor plots, the stability of classifications can be investigated. In the future, scissor plots could also be evaluated for k-NN classifications based on tree-ring variables other than ring-width data, such as blue intensity, density, or stable isotopes.

Acknowledgements

I thank N. Bleicher for comments, discussions and suggestions; H. Bugmann for comments, discussions and suggestions and providing me with meteorological data; Ph. Della Casa and H. Bugmann for supervising my PhD thesis; and T. Eissing and two anonymous reviewers for additional comments and discussions; also, I thank M. Bolliger, K. Emmenegger, S. Forabosco, F. Gut, L. Gut, T. Jagoulis, N. Lengacher, J. Schmidt, T. Wäckerle, F. Walder and S. Wicki for support in the field work. This study was financially supported by the Swiss National Science Foundation, grant no. P0ZHP1_162299.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.dendro.2018.09.008.

References

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4, 40–79.
Baillie, M.G.L., Pilcher, J.R., 1973. A simple cross-dating programe for tree-ring research.

Tree-Ring Bull. 33, 7–14.
Best, D.J., Roberts, D.E., 1975. Algorithm AS 89: the upper tail probabilities of Spearman's rho. Appl. Stat. 24, 377–379.
Boschetti-Maradi, A., Kontic, R., 2012. Möglichkeiten und Schwierigkeiten dendrochronologischer Untersuchungen in Mittelalterarchäologie und Bauforschung. Mitteilungen der Deutschen Gesellschaft für Archäologie des Mittelalters und der Neuzeit 24, 49–60.
Bridge, M., 2000. Can dendrochronology be used to indicate the source of oak within Britain? Vernac. Archit. 31, 67–72.
Bridge, M., 2012. Locating the origins of wood resources: a review of dendroprovenancing. J. Archaeol. Sci. 39, 2828–2834.
Brockwell, P.J., Davis, R.A., 1996. Introduction to Time Series and Forecasting. Springer Texts in Statistics. Springer, Berlin and Heidelberg.
Bronaugh, D., 2018. climdex.pcic: PCIC Implementation of Climdex Routines. R package version 1.1-9.
Buras, A., Wilmking, M., 2015. Correcting the calculation of Gleichläufigkeit. Dendrochronologia 34, 29–30.
Cook, E.R., 1981. The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bull. 41, 45–53.
Cook, E.R., 1985. A time series approach to tree ring standardization. University of Arizona, Tucson (Dissertation).
Cook, E.R., Kairiukstis, L.A. (Eds.), 1990. Methods of Dendrochronology. Kluwer, Dordrecht.
Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.
Daly, A., Nymoen, P., 2008. The Bøle ship, Skien, Norway–Research history, dendrochronology and provenance. Int. J. Nautic. Archaeol. 37, 153–170.
Dittmar, C., Eissing, T., Rothe, A., 2012. Elevation-specific tree-ring chronologies of Norway spruce and Silver fir in Southern Germany. Dendrochronologia 30, 73–83.
Dittmar, C., Elling, W., 1999. Jahrringbreite von Fichte und Buche in Abhängigkeit von Witterung und Höhenlage. Forstwissenschaftliches Centralblatt 118, 251–270.
Drake, B.L., 2018. Source & sourceability: towards a probabilistic framework for dendroprovenance based on hypothesis testing and Bayesian inference. Dendrochronologia 47, 38–47.
Eckstein, D., Wrobel, S., 2007. Dendrochronological proof of origin of historic timber – retrospect and perspectives. In: Haneca, K., Verheyden, A., Beeckman, H., Gärtner, H., Helle, G., Schleser, G. (Eds.), TRACE – Tree Rings in Archaeology, Climatology and Ecology, Vol. 5: Proceedings of the Dendrosymposium 2006, April 20th–22nd 2006, Tervuren, Belgium, pp. 8–20.
Edgell, S.E., Noon, S.M., 1984. Effect of violation of normality on the t test of the correlation coefficient. Psychol. Bull. 95, 576–583.
Eissing, T., 2007. Zum Problem der Interpretation dendrochronologischer Datierung bei Flossholz am Beispiel Bambergs. In: für Hausforschung, A. (Ed.), Spuren der Nutzung in historischen Bauten. Jonas, Marburg. Jahrbuch für Hausforschung, 54, pp. 23–36.
Eissing, T., Dittmar, C., 2011. Timber transport and dendroprovenancing in Thuringia and Bavaria. In: Fraiture, P. (Ed.), Treerings, Art and Archaeology, pp. 137–149.
Everitt, B.S., Sabine, L., Morven, L., Stahl, D., 2011. Cluster Analysis. Wiley Series in Probability and Statistics 848, 5th edition. Wiley, Chichester.
Fowler, A., Bridge, M., 2015. Mining the British Isles oak tree-ring data set. Part A: Rationale, data, software, and proof of concept. Dendrochronologia 35, 24–33.
Frank, D., Esper, J., 2005. Characterization and climate response patterns of a high-elevation, multi-species tree-ring network in the European Alps. Dendrochronologia 22, 107–121.
Fritts, H.C., 1976. Treerings and Climate. Academic Press, London, New York, San Francisco.
Garcia-Gonzalez, I., 2008. Comparison of different distance measures for cluster analysis of tree-ring series. Tree-Ring Res. 64, 27–37.
Grissino-Mayer, H.D., 2001. Evaluating crossdating accuracy: a manual and tutorial for the computer program COFECHA. Tree-Ring Res. 57, 205–221.
Haneca, K., Wasny, T., van Acker, J., Beeckman, H., 2005. Provenancing Baltic timber from art historical objects: success and limitations. J. Archaeol. Sci. 32, 261–271.
Hellmann, L., Tegel, W., Geyer, J., Kirdyanov, A.V., Nikolaev, A.N., Eggertsson, O., Altman, J., Reinig, F., Morganti, S., Wacker, L., Büntgen, U., 2017. Dendro-provenancing of Arctic driftwood. Q. Sci. Rev. 162, 1–11.
Hennig, C.M., Meilă, M., Murtagh, F., Rocci, R., 2016. Handbook of cluster analysis. Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Taylor & Francis Group, Boca Raton.
Hollander, M., Wolfe, D.A., 1973. Nonparametric Statistical Methods. Wiley Series in Probability and Statistics. John Wiley and Sons Ltd., New York.
Hollstein, E., 1980. Mitteleuropäische Eichenchronologie. Trierer dendrochronologische Forschungen zur Archäologie und Kunstgeschichte 11, Phillipp von Zabern, Mainz am Rhein.
Huber, B., 1943. Über die Sicherheit jahrringchronologischer Datierung. Holz als Roh- und Werkstoff 6 263 ff.
Jansma, E., Haneca, K., Kosian, M., 2014. A dendrochronological reassessment of three Roman vessels from Utrecht (the Netherlands): Evidence of inland navigation between the lower-Scheldt region in Gallia Belgica and the limes of Germania inferior. J. Archaeol. Sci. 50, 1–13.
von Jazewitsch, W., 1948. Über die Möglichkeiten einer jahrringchronologischen Individualdiagnose von Bäumen mit Beiträgen zur Methodik der Jahrringforschung. Ludwig-Maximilians-Universität, München (Dissertation).
Kendall, M.G., Gibbons, J.D., 1990. Rank Correlation Methods, 5th ed. Edward Arnold, London.
King, G.M., Gugerli, F., Fonti, P., Frank, D.C., 2013. Tree growth response along an elevational gradient: climate or genetics? Oecologia 173, 1587–1600.
Kolář, T., Čermák, P., Trnka, M., Žid, T., Rybníček, M., 2017. Temporal changes in the climate sensitivity of Norway spruce and European beech along an elevation gradient in Central Europe. Agric. For. Meteorol. 239, 24–33.
Lyu, L., Suvanto, S., Nöjd, P., Henttonen, H.M., Mäkinen, H., Zhang, Q.B., 2017. Tree growth and its climate signal along latitudinal and altitudinal gradients: comparison of tree rings between Finland and the Tibetan Plateau. Biogeosciences 14, 3083–3095.
Monserud, R.A., 1986. Time-series analysis of tree-ring chronologies. For. Sci. 32, 349–372.
Monserud, R.A., Yamaguchi, D.K., 1989. Comments on “Cross-dating methods in dendrochronology” by Wigley et al. J. Archaeol. Sci. 16, 221–224.
Neuwirth, B., Esper, J., Schweingruber, F.H., Winiger, M., 2004. Site ecological differences to the climatic forcing of spruce pointer years from the Lötschental, Switzerland. Dendrochronologia 21, 69–78.
Neuwirth, B., Schweingruber, F.H., Winiger, M., 2007. Spatial patterns of central European pointer years from 1901 to 1971. Dendrochronologia 24, 79–89.
Pearson, K., 1895. Note on regression and inheritance in the case of two parents. Proc. R. S. Lond. U.S.A. 58, 240–242.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rolland, C., Desplanque, C., Michalet, R., Schweingruber, F.H., 2000. Extreme tree rings in spruce (*Picea abies* [L.] Karst.) and fir (*Abies alba* Mill.) stands in relation to climate, site, and space in the Southern French and Italian Alps. Arctic Antarctic Alpine Res. 32, 1.
Savva, Y., Oleksyn, J., Reich, P.B., Tjoelker, M.G., Vaganov, E.A., Modrzynski, J., 2006. Interannual growth response of Norway spruce to climate along an altitudinal gradient in the Tatra Mountains, Poland. Trees 20, 735–746.
Schmitt, I., 2006. Ähnlichkeitssuche in Multimedia-Datenbanken: Retrival, Suchalgorithmen und Anfragebehandlung. Oldenbourg, München.
Speer, J.H., 2010. Fundamentals of Tree-ring Research. University of Arizona Press, Tucson.
Stock, J.H., Watson, M.W., 2015. Introduction to Econometrics. Always Learning, 3rd global edition. Pearson, Boston.
Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol. 190, 214–251.
Wazny, T., 2002. Baltic timber in Western Europe – an exciting dendrochronological question. Dendrochronologia 20, 313–320.
Wigley, T., Jones, P.D., Briffa, K.R., 1987. Cross-dating methods in dendrochronology. J. Archaeol. Sci. 14, 51–64.
Wilson, R., Hopfmüller, M., 2001. Dendrochronological investigations of Norway spruce along an elevational transect in the Bavarian Forest, Germany. Dendrochronologia 19, 67–79.