sciendo

DOI: 10.2478/bile-2024-0008

*Biometrical Letters*
*Vol. 61 (2024), No. 2, 115-135*

# Comparison of Pearson's and Spearman's correlation coefficients for selected traits of *Pinus sylvestris* L.

**Jan Bocianowski[1], Dorota Wrońska-Pilarek[2], Anna Krysztofiak-Kaniewska[3], Karolina Matusiak[3], Blanka Wiatrowska[2]**

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland
[2]Department of Botany and Forest Habitats, Poznań University of Life Sciences, Wojska Polskiego 71C, 60-625 Poznań, Poland
[3]Department of Forest Engineering, Poznań University of Life Sciences, Wojska Polskiego 71C, 60-625 Poznań, Poland

## SUMMARY

The Spearman rank correlation coefficient is a non-parametric (distribution-free) rank statistic proposed by Charles Spearman as a measure of the strength of the relationship between two variables. It is a measure of a monotonic relationship that is used when the distribution of the data makes Pearson's correlation coefficient undesirable or misleading. The Spearman coefficient is not a measure of the linear relationship between two variables. It assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. Unlike Pearson's product-moment (linear) correlation coefficient, it does not require the assumption that the relationship between variables is linear, nor does it require that the variables be measured on interval scales; it can be applied to variables measured at the ordinal level. The purpose of this study is to compare the values of Pearson's product-moment correlation coefficient and Spearman's rank correlation coefficient and their statistical significance for six morpho-anatomical traits of *Pinus sylvestris* L. (original – for Pearson's coefficient, and ranked – for Spearman's coefficient) estimated from all observations, object means (for trees), and medians. The results show that the linear and rank correlation coefficients are consistent (as to direction and strength). In cases of divergence in the direction of correlation, the correlation coefficients were not statistically significant, which does not imply consistency in decision-making. Estimation of correlation coefficients based on medians is robust to outlier observations and factors that linear correlation is then very similar to rank correlation.

**Key words:** linear correlation; rank correlation; Scots pine; median

## 1. Introduction

The study of trait interdependence (otherwise known as correlation) is a key area of data analysis in the life sciences, medicine, social sciences, economics and many other fields (Fontana et al. 2020; Shimizu et al. 2020; Tundys et al. 2021). Correlation allows us to understand how different traits or variables interact with each other, which is important in the data analysis process (Thielmann et al. 2020). Studying the correlation of traits makes it possible to identify patterns and relationships between different variables (Wright et al. 2007; Orsini et al. 2013). This allows the discovery of hidden relationships between data, which can lead to new research conclusions and hypotheses. Correlation enables the profiling of phenomena through the analysis of their interrelationships (Alaimo et al. 2022). This facilitates a deeper understanding of the associations between various characteristics and helps identify factors that may influence the observed phenomena. The exploration of correlation enables the prediction of behaviors and trends by discerning patterns within historical data (Moews et al. 2019). If two traits have a strong correlation, one can be used to predict the value of the other, which can be very important in stand research. Correlation analysis can help optimize production, marketing or resource management processes (Min et al. 2019). Studying correlations can lead to a better understanding of which factors affect efficiency and how performance can be improved (Lin 2007; Lateef and Keikhosrokiani 2023).

In research, especially in statistics, correlation analysis helps select variables for further analysis. If two traits are highly correlated, it may suggest that both yield similar information, which may lead to the elimination of one of them to simplify the research model. While correlation does not imply causality, it can provide a starting point for further research into the causes of relationships between traits. Identifying a strong correlation can inspire researchers to dig deeper to understand why these traits are related (Dormann et al. 2013; Iqbal et al. 2021; Noor et al. 2020).

In statistics, correlation is a measure of the strength and direction of a relationship between two variables. Correlation results are easy to understand and interpret. The correlation value ranges from –1 to 1, where 1 means perfect positive correlation, –1 means perfect negative correlation, and 0 means no correlation. This simple scale makes it easy to compare relationships between different variables and allows quick inferences about the strength of the relationship.

There are several different methods for calculating correlations, the two most popular being linear correlation (Pearson correlation) and rank correlation (Spearman correlation). Both types of correlation are used in scientific research, but differ in their assumptions and interpretation of results. Linear correlation measures the strength and direction of the relationship between two variables, assuming that the relationship between them is linear. Linear correlation assumes that changes in one variable are proportionally related to changes in the other variable. Linear correlation is sensitive to outliers, meaning that extreme values in the data can significantly affect the correlation result. Rank correlation is a non-linear measure of the relationship between two variables. Unlike linear correlation, it does not assume any restrictions on the shape of the relationship between variables. Instead, rank correlation is based on the order of the values in the two variables, transforming the observations into their respective ranks. Spearman correlation is then calculated based on the ranks of these observations. Rank correlation is less sensitive to outliers than linear correlation because it is based on the order of the ranks rather than the exact values of the data. Correlation analysis is very often performed on forest stand data (Ahmadi et al. 2020; Hyyppä et al. 2000; Puliti et al. 2020; Moradi et al. 2022; Wrońska–Pilarek et al. 2023a).

The history and properties of the Pearson correlation coefficient have been described by Pearson (1920), Weida (1927), Walker (1928), Stigler (1988) and Piovani (2008). Pearson's correlation coefficient was discovered by Bravais in 1846 (Bravais 1846), but Karl Pearson was the first to describe a standard method of calculating it in 1896 (Pearson 1896) and showed that it was the best possible method. Pearson's important assumption is the normality of the analyzed variables, which can only be true for quantitative variables. Pearson's correlation

coefficient is a measure of the strength of a linear relationship between two such variables. In 1904, Spearman (1904) adopted the Pearson correlation coefficient as a measure of the strength of the relationship between two variables that cannot be measured quantitatively.

The purpose of this study was to compare the results obtained from correlation analysis of six morpho-anatomical traits of the needles of ten Scots pine (*Pinus sylvestris* L.) trees based on Pearson's linear correlation coefficients and Spearman's rank correlation. Correlation coefficients were estimated based on (1) all collected data from 30 needles for each tree, (2) mean values for individual trees and (3) medians for individual trees. Scots pine needles served only as an example to achieve this goal. There is a knowledge gap in this area that needs to be addressed: Pearson and Spearman correlations are frequently used in morphometric studies, such as taxonomical and ecological research, but they are rarely applied simultaneously. Consequently, there is limited information on how convergent the results from different statistical approaches applied to the same research sample may be.

## 2. Materials and Methods

### 2.1. Study area

The analyses were carried out near the Belzec-Podlesin aeolian sand mine located near the town of Belzec (50°23′05″N, 23°24′04″E), in Eastern Roztocze (Solon 2018), southeastern Poland, which operated in 2001–2008. The area was reclaimed after the closure of the open-pit mine. In 2019, one permanent experimental plot was established in the immediate vicinity of the sand mine by staff of the Department of Forest Engineering at the Poznań University of Life Sciences. The area of the experimental plot was 6.04 ha. The distance of the center of the experimental plot from the topographical center of the sand mine was 255.9726 m, and the range of distances for individual trees was from 205.28 m to 293.24 m. Other characteristics of experimental plots were as follows:

stand/share – pine/10, age/quality class – 74/I, forest habitat type – fresh coniferous forest, moisture content – fresh, and soil type – rusty podzolic soils.

### 2.2. Sampling

The field research was conducted in May 2019. The experimental plot was established at a distance of 255.9726 m from the topographic center of the sand mine. Then, ten randomly selected Scots pine trees growing in the plot in similar habitat conditions were cut and numbered from 1 to 10. The plant material was stored in a freezer at –18 °C. For each of the ten common pine trees tested, 30 needles (one year old) were measured. A total of 300 randomly selected and correctly formed needles were measured.

### 2.3. Morphological and anatomical analysis

The needles were measured in the period from December 2020 to June 2021. These needles were extracted from branches that had been thawed to ambient temperature. Fresh needle samples were meticulously prepared from specimens stored in refrigeration, methodically extracted, and sectioned utilizing a manual cylindrical microtome (Euromex). Given the inherent variability in the structural composition of pine needles along their longitudinal axis (Jankowski et al. 2021), a consistent approach was employed: the needles were uniformly bisected longitudinally. Subsequent to this preparation, the specimens were meticulously preserved utilizing glycerin. Needle dimensions, specifically thickness and width, were measured under 10x magnification. Additionally, resin canals and stomatal apparatus were quantified on both the concave and convex surfaces of the needles. Following this preliminary analysis, each individually chosen needle underwent further scrutiny. Measurements were carried out employing a Dexter digital caliper and a Levenhuk 850B digital light microscope, featuring LevenhukLite software tailored to enhance the precision of these measurements. The length of each needle was measured with a precision of 0.01 cm. Subsequent analyses included determining epidermal and cuticular thickness, the thickness of individual hypodermal cells, and the diameter of individual resin canals. These

parameters were scrutinized at a 40x magnification level, with measurements recorded to an accuracy of 0.01 µm. A total of six quantitative morphological and anatomical characteristics of needles were analyzed: needle length (NL), needle width (NW), needle thickness (NT), thickness of epidermis and cuticle (TEC), hypodermal cell thickness (HCT), resin duct diameter (RD). The method used to make the measurements is shown in Figure 1. The experiment is described in detail in Wrońska-Pilarek et al. (2023b).

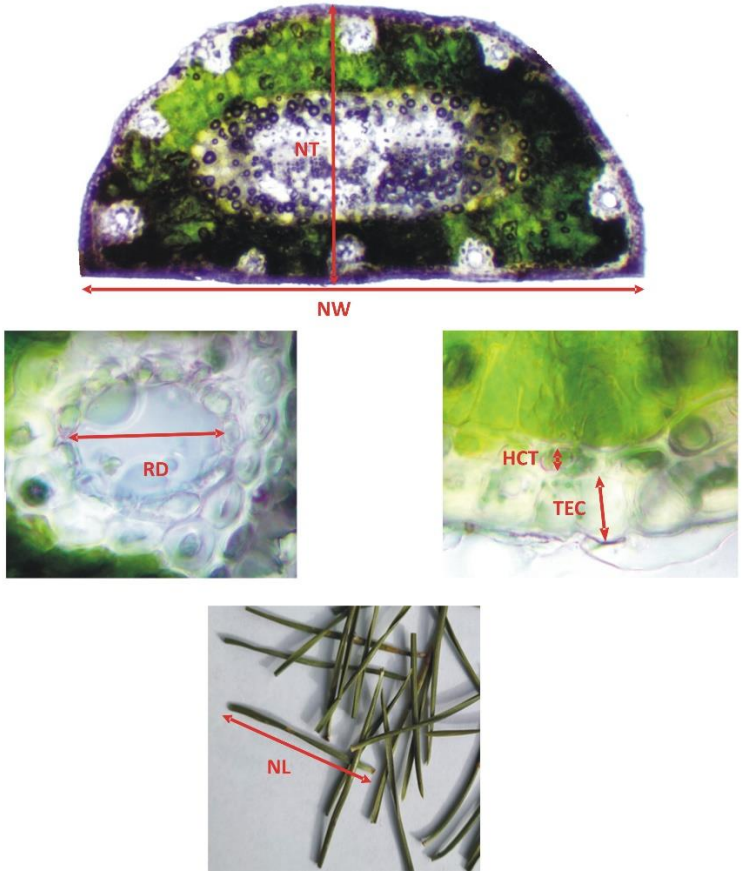



**Figure 1.** The method used to measure analyzed needle traits: NL – needle length, NW – needle width, NT – needle thickness, TEC – thickness of epidermis and cuticle, HCT – hypodermal cell thickness, RD – resin duct diameter

## 2.4. Statistical analysis

### Pearson's correlation

Pearson's correlation coefficient ($\rho_P$) is a measure of the linear correlation between two sets of data $\boldsymbol{X}$ and $\boldsymbol{Y}$ (Pearson 1908):

$$\rho_P = \frac{cov(\boldsymbol{X}, \boldsymbol{Y})}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{Y}}}, \quad (1)$$

where $cov(\boldsymbol{X}, \boldsymbol{Y})$ is the covariance of two variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\sigma_{\boldsymbol{X}}$ is the standard deviation of $\boldsymbol{X}$, and $\sigma_{\boldsymbol{Y}}$ is the standard deviation of $\boldsymbol{Y}$. The formula for $cov(\boldsymbol{X}, \boldsymbol{Y})$ can be expressed in terms of mean and expectation:

$$cov(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{Y} - \mu_{\boldsymbol{Y}})], \quad (2)$$

where E($\boldsymbol{Z}$) is the expected value of a random variable $\boldsymbol{Z}$, $\mu_{\boldsymbol{X}}$ is the mean of $\boldsymbol{X}$, and $\mu_{\boldsymbol{Y}}$ is the mean of $\boldsymbol{Y}$. The formula for Pearson's correlation coefficient can be obtained by substituting estimates of the covariance and variance based on a sample into the formula (1). Given paired data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ consisting of *n* pairs, $\rho_P$ is defined as:

$$\rho_P = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3)$$

where *n* is the sample size, $x_i$ and $y_i$ are the individual sample points indexed with *i*, $\bar{x}$ is the sample mean of $\boldsymbol{X}$, and $\bar{y}$ is the sample mean of $\boldsymbol{Y}$.

### Spearman's correlation

Spearman's rank correlation coefficient ($\rho_S$) is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables) (Spearman 1904). The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the rank variables. For a sample of size *n*, the *n* raw scores $X_i$, $Y_i$ are converted to ranks $R(X_i)$, $R(Y_i)$, and $\rho_S$ is computed as:

$$\rho_S = \frac{cov\big(\mathrm{R}(\boldsymbol{X}), \mathrm{R}(\boldsymbol{Y})\big)}{\sigma_{\mathrm{R}(\boldsymbol{X})}\sigma_{\mathrm{R}(\boldsymbol{Y})}}, \quad (4)$$

where $cov\big(\mathrm{R}(\boldsymbol{X}),\mathrm{R}(\boldsymbol{Y})\big)$ is the covariance of the rank variables, $\sigma_{\mathrm{R}(\boldsymbol{X})}$ is the standard deviation of the rank variable $\boldsymbol{X}$, and $\sigma_{\mathrm{R}(\boldsymbol{Y})}$ is the standard deviation of the rank variable $\boldsymbol{Y}$.

**Data sets**

Pearson's product-moment (linear) correlation coefficients and Spearman's rank correlation coefficients for the six *Pinus sylvestris* L. traits were calculated for three sets: (i) based on all observations, (ii) based on object means (for trees), and (iii) based on medians from empirical distributions.

Testing correlation coefficients

For pairs from an uncorrelated bivariate normal distribution, the sampling distribution of the Studentized Pearson's correlation coefficient follows Student's *t*-distribution with *n*–2 degrees of freedom under the null hypothesis:

$$t = r_P \sqrt{\frac{n-2}{1-r_P^2}}. \tag{5}$$

where $r_P$ is the sample linear Pearson's correlation coefficient. For sample sizes approaching infinity, the $\rho_S$ Spearman rho distribution approaches a normal distribution with expected value equal to the true value $\rho_S$ in the population and variance 1/(*n*–1), hence the statistic used is:

$$z = r_S\sqrt{n-1}, \tag{6}$$

whose distribution approaches the standard normal distribution *N*(0,1) with increasing sample size, where $r_S$ is the sample Spearman's rank correlation coefficient. As the sample size increases, the precision of inference improves (the critical region expands). We should expect to see more statistically significant correlation coefficients when evaluating all original observations than when evaluating data that have been converted (to mean or median values). This is because there are more degrees of freedom in testing. For all observations, the number of degrees of freedom will be 298, whereas when testing on means and medians, it will be only 8.

All analyses were performed using the GenStat v. 23 statistical software package (VSN International Genstat for Windows 2023).

## 3. Results

### 3.1. Correlation based on all observations

The direction of correlation, characterized by the sign of the correlation coefficient calculated from all observed data, was always consistent for linear correlation and rank correlation (Figure 2). The concordance of significance (or lack thereof) of linear correlation with rank correlation was almost perfect. Simultaneous significance of $r_P$ and $r_S$ correlation coefficients was not observed only between RD and TEC. For these two traits, a significant rank correlation was observed ($r_S$=0.133), with no linear correlation ($r_P$=0.066) (Figure 2).

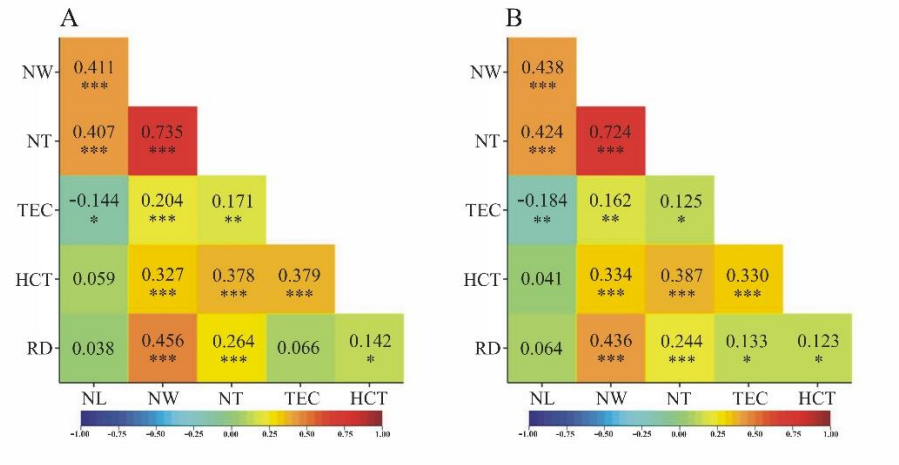



**Figure 2.** Comparison of Pearson's (a) and Spearman's (b) correlation coefficients of selected *Pinus sylvestris* L. traits calculated from all observations. NL – needle length, NW – needle width, NT – needle thickness, TEC – thickness of epidermis and cuticle, HCT – hypodermal cell thickness, RD – resin duct diameter. * – $p$<0.05; ** – $p$<0.01; *** – $p$<0.001; $n$=300

### 3.2. Correlation calculated on object mean values

The signs of Pearson's and Spearman's correlation coefficients estimated from object averages were consistent for all pairs of traits except the RD–NL pair ($r_P$= –0.059, $r_S$=0.042). However, neither coefficient was statistically significant (Figure 3). Analyzing the values of the correlation coefficients, concordance was observed in terms of their statistical significance for all pairs of morpho-anatomical traits except NL–TEC. Although both coefficients were negative, a statistically significant correlation coefficient between the two traits was observed only for ranks ($r_S$= –0.663) (Figure 3b).

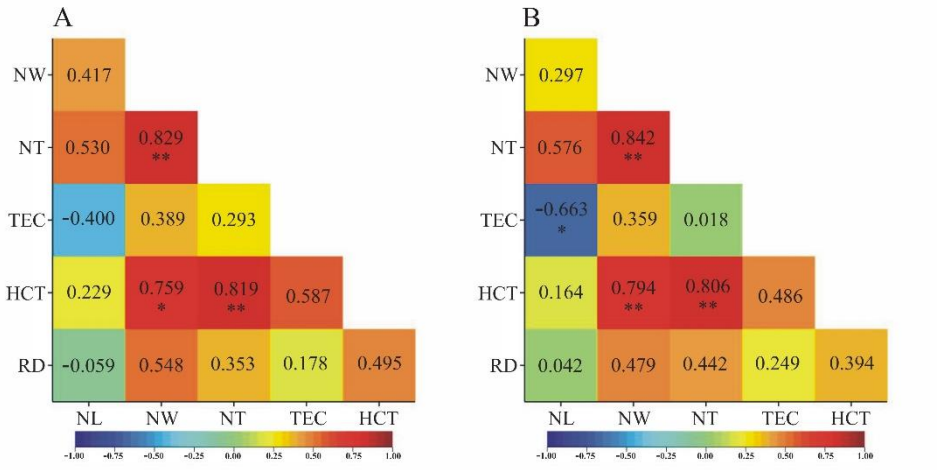



**Figure 3.** Comparison of Pearson's (a) and Spearman's (b) correlation coefficients of selected *Pinus sylvestris* L. traits calculated from tree averages. NL – needle length, NW – needle width, NT – needle thickness, TEC – thickness of epidermis and cuticle, HCT – hypodermal cell thickness, RD – resin duct diameter. * – $p$<0.05; ** – $p$<0.01; $n$=10

### 3.3. Correlation calculated on medians from empirical distributions

Correlation coefficients estimated from medians were characterized by sign agreement for linear and rank correlation except in two situations. For the pairs of traits NL–TEC and NL–RD, linear correlation was negative (–0.432 and

–0.082, respectively), and rank correlation was positive (0.555 and 0.079, respectively). All four coefficients were not statistically significant (Figure 4). For most pairs of traits, the linear and rank correlation coefficients were statistically significant or insignificant at the same time (Figures 4a and 4b). Only between NL and NT was the opposite significance relationship observed: $r_P$ was not statistically significant (0.612; Figure 4a), and $r_S$ had a significance level of $\alpha$=0.05 (0.675; Figure 4b).

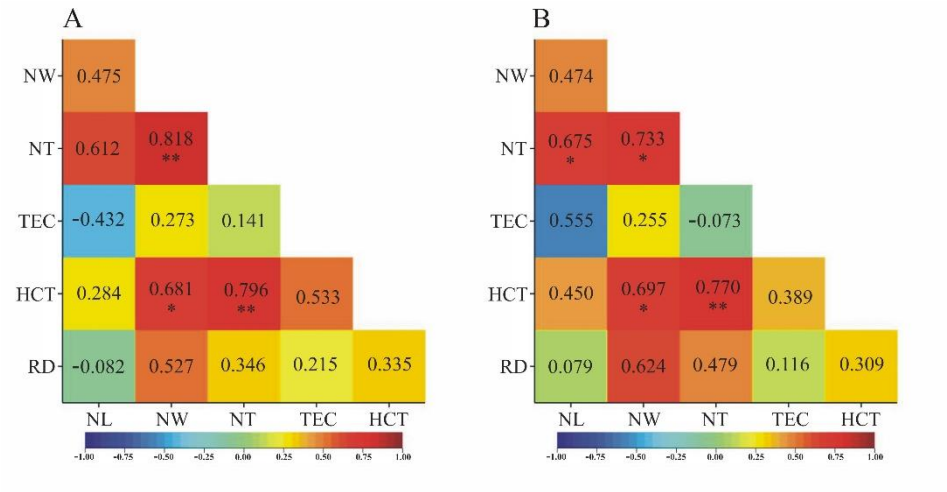



**Figure 4.** Comparison of Pearson's (a) and Spearman's (b) correlation coefficients of selected *Pinus sylvestris* L. traits calculated from medians. NL – needle length, NW – needle width, NT – needle thickness, TEC – thickness of epidermis and cuticle, HCT – hypodermal cell thickness, RD – resin duct diameter. * – $p$<0.05; ** – $p$<0.01; $n$=10

### 3.4. Comparison between estimation methods

The most significant correlation coefficients were observed when correlation was estimated on the basis of all original observations – without recalculations (Figure 2). Much less significant correlation coefficients were found when they were estimated on the basis of object means or medians (Figures 3 and 4). If a pair of traits was characterized by a significant correlation coefficient estimated from tree means or medians, it also had a significant correlation coefficient calculated

on the basis of all observations. Pearson's linear correlation coefficients were equally significant (or simultaneously insignificant) when comparing estimation methods based on means and medians (Figures 3a and 4a). For Spearman's rank correlation coefficients, such a relationship was not clear in two cases. NT correlated significantly with NL when this relationship was assessed based on medians, while it was not significant for the correlation coefficient estimated based on tree means (Figures 3b and 4b). The opposite was true for the TEC–NL trait pair: the correlation coefficient based on mean values was statistically significant, while that based on medians lacked statistical significance (Figures 3b and 4b).

## 4. Discussion

Correlation analysis is an extremely important tool in biology for understanding relationships between various biological variables (Carter and Wiles 2014; Rosato et al. 2018; Rutledge et al. 2022; Saalidong et al. 2022). Its importance includes identifying relationships between traits of organisms, studying the inheritance of genetic traits, analyzing the influence of environmental factors, and studying disease and health. This provides more precise information on the functioning of living organisms and the influence of external factors on their life and health (Tortella et al. 2020; Waszak et al. 2021; Samal et al. 2022).

Nevertheless, simply conducting a correlation analysis is only the beginning. By evaluating the results of such analysis, deeper insights into complex biological processes can be gained, which is crucial for further research, interpretation of results, and decision-making in both science and practice (Nowosad et al. 2023).

Evaluation of correlation analysis helps identify potential errors and pitfalls in the data. Before evaluating correlation, one should make sure that the analyzed data is reliable and properly collected. Otherwise, the results of the analysis may be erroneous or misleading. When assessing the quality of the data, it is important to avoid false conclusions and ensure the reliability of the research.

Another key aspect of evaluating correlation analysis is to consider the biological context. Not all correlations are causal, meaning that even if there is a relationship between two variables, it is not necessarily the case that one variable causes changes in the other. Correlation analysis should always be subjected to an evaluation of the biological context to avoid false interpretations. Otherwise, there is a risk of drawing incorrect conclusions about biological mechanisms (Broadhurst and Kell 2006; Nakagawa and Cuthill 2007).

Last but not least, the importance of assessing correlation analysis is to be able to predict the behavior of future variables based on known correlational relationships. In biology, especially in fields such as ecology or evolutionary biology, the ability to predict changes in populations or ecosystems is crucial for taking effective conservation measures or managing species populations (Clutton-Brock 2010; Dwyer et al. 2020; Hohenlohe et al. 2021).

In data analysis, especially in scientific research, the choice of appropriate statistical tools is crucial to obtaining reliable results. When the relationship between variables is studied, the question is often posed: is it worthwhile to simultaneously evaluate linear correlation and rank correlation? The answer to this question depends on the characteristics of the data and the purpose of the study (Horváth 2010). In many cases, it is worth using both methods because they offer different perspectives on the relationship under analysis. Linear correlation is useful when the relationship between variables is linear. However, if the data contain outliers or do not meet the assumptions of normality, linear correlation can produce erroneous or misleading results. In such cases, rank correlation may be a more reliable alternative because it is less sensitive to outliers and does not require assumptions about the distribution of the data. In the research presented here, both measures were used to assess the correlation between the observed characteristics. Similarly, both measures, Pearson’s and Spearman’s, were used in the following studies, among others: Bonett and Wright (2000), Artusi et al. (2002), Hauke and Kossowski (2011), and Eisinga et al. (2013).

Assessing linear and rank correlation simultaneously provides a more complete picture of the relationship between variables. It may be that both

methods indicate a similar relationship, which will strengthen our confidence in interpreting the results. In the study presented here, there was general agreement in the direction and significance of linear and rank correlation coefficients (Figures 2, 3 and 4). However, in the case of differing results, analysis of the two types of correlation can help identify specific traits of the data that affect the results of the analysis. Cases where the correlations did not agree were noted in the results. In cases where the direction of correlation diverged, the correlation coefficients were not statistically significant, which does not imply consistency in decision-making. Linear and rank correlation coefficients with different signs were also obtained by other researchers: Udovičić et al. (2007), and Song and Park (2020). Simultaneous analysis of linear correlation and rank correlation can also help verify the results and ensure their reliability. If both methods indicate a similar relationship, our confidence in the veracity of the relationship increases. Such situations occurred in the results presented above on the correlation of six pine traits. If the results diverge, a closer examination of the data and potential reasons for the differences is necessary. A discrepancy in the context of significance/non-significance between Pearson's and Spearman's coefficients for the TEC–RD trait pair requires a thorough investigation of the biological relationships between the traits.

The higher number of statistically significant correlation coefficients in the case of evaluation on the basis of all original observations than for the converted data (to mean or median values) is due to the higher number of degrees of freedom in testing. For all observations, the number of degrees of freedom was 298, while for testing based on mean and median values it was 8. This resulted in smaller critical values when testing correlation coefficients based on all original observations.

Linear correlation is essential in regression analysis, which is a widely used technique in the social and natural sciences. Linear correlation makes it possible to assess the extent to which one variable can predict another. The results of linear correlation analysis are the basis for linear regression models, which make it possible to predict the value of one variable from another. This is especially

important in studies of forest stands (Neumann and Starlinger 2001; Lefsky et al. 2005; Ali 2019).

The estimation of rank correlation coefficients has its own unique advantages over the estimation of linear correlation coefficients. Its robustness to outliers, lack of requirements for data distribution, ability to detect nonlinear relationships, adaptability to different measurement scales, and greater flexibility in analyzing nonlinear data make it an attractive choice in many scientific fields. The estimation of rank correlation coefficients is particularly valuable for analyzing data that do not meet assumptions about the linear nature of relationships between variables, making it a versatile tool in scientific research (Schober et al. 2018; Lindinger-Sternart et al. 2021).

Both correlation methods have their place in data analysis, and the choice between them depends on the nature of the variables and the assumptions of the study. If the relationship between the variables appears to be linear and there are no extreme outliers, linear correlation may be an appropriate tool. On the other hand, if a linear relationship cannot be assumed or there are outliers, rank correlation may provide more reliable results. With rank correlation, critical values are less prone to change at different levels of statistical significance, which can be useful in analyses where statistical significance is crucial. Since in the results of the presented study the number of observations for each tree was relatively large (30 observations), the choice of method for estimating correlation coefficients (based on means, medians or all data) does not seem to matter much. Evaluation based on medians “compensates” for outlier observations to a large extent. In contrast, the median-based approach is completely immune to outlier observations, even when there are more of them. Rank correlation is less sensitive to differences in measurement scale between variables. This means that it can be used to analyze measurement data on both interval and ordinal scales, making it more flexible in application. It is important always to understand the characteristics of the data and use the appropriate correlation method to obtain reliable and interpretable results in scientific research.

## 5. Conclusions

Studying the interdependence of characteristics is crucial in data analysis, as it allows understanding of the relationships between different variables. It makes it possible to discover new facts, make predictions, optimize processes and conduct effective scientific research. Correlation helps identify patterns and relationships, which is a fundamental step in the process of data analysis and scientific inference. Evaluation of the results of correlation analysis is as important as the analysis itself. Through careful evaluation, misinterpretations can be avoided, relationships between variables can be understood in the appropriate biological context, and changes in the systems under study can be predicted, all of which contribute to scientific progress and effective activities in the field of biology. Simultaneous evaluation of linear correlation and rank correlation is an effective strategy in analyzing relationships between variables. It makes it possible to get a comprehensive picture of the relationship between data, taking into account the diversity of data and differences in the characteristics of the variables under study. Using both methods, more certain and reliable results are obtained, which provide a solid basis for interpreting the relationships between the analyzed variables. The simultaneous use of linear correlation and rank correlation allows a more complete understanding of the phenomena under study, which contributes to the progress of science and the formation of more precise theories and hypotheses. Estimation of correlation coefficients based on medians is robust to outlier observations and factors that linear correlation is then very similar to rank correlation. The direction of correlation, characterized by the sign of the coefficients, was in the vast majority of cases consistent for linear correlation and rank correlation. The largest number of significant correlation coefficients was observed when the correlation was estimated based on all original observations – without recalculations. This is because as the sample size increases, the precision of inference improves (the critical region expands). The explanation for the discrepancy in the context of significance/insignificance between Pearson and

Spearman coefficients for the TEC–RD trait pair should be sought in the specific properties of pollen and the biological relationships between these two traits.

## REFERENCES

Ahmadi K., Kalantar B., Saeidi V., Harandi E.K.G., Janizadeh S., Ueda N. (2020): Comparison of Machine Learning Methods for Mapping the Stand Characteristics of Temperate Forests Using Multi-Spectral Sentinel-2 Data. Remote Sens 12:3019. https://doi.org/10.3390/rs12183019

Alaimo L.S., Arcagni A., Fattore M., Maggino F., Quondamstefano V. (2022): Measuring Equitable and Sustainable Well-Being in Italian Regions: The Non-aggregative Approach. Soc Indic Res 161: 711–733. https://doi.org/10.1007/s11205-020-02388-7

Ali A. (2019): Forest stand structure and functioning: Current knowledge and future challenges. Ecol Indic 98: 665–677. https://doi.org/10.1016/j.ecolind.2018.11.017

Artusi R., Verderio P., Marubini E. (2002): Bravais-Pearson and Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval. Int J Biol Markers 17: 148–151. DOI:10.1177/172460080201700213

Bonett D.G., Wright T.A. (2000): Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika 65: 23–28. https://doi.org/10.1007/BF02294183

Bravais A. (1846): Analyse mathématique sur les probabilités des erreurs de situation d'un point. *Mémoires présentés par divers savants à l'Académie Royale des Sciences de l'Institut de France* 9: 255–332.

Broadhurst D.I., Kell D.B. (2006): Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2: 171–196. https://doi.org/10.1007/s11306-006-0037-z

Carter B.E., Wiles J.R. (2014): Scientific consensus and social controversy: exploring relationships between students' conceptions of the nature of science, biological evolution, and global climate change. Evo Edu Outreach 7: 6. https://doi.org/10.1186/s12052-014-0006-3

Clutton-Brock T., Sheldon B.C. (2010): Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends Ecol Evol 25: 562–573. https://doi.org/10.1016/j.tree.2010.08.002

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B., Leitão P.J., Münkemüller T., McClean C., Osborne P.E., Reineking B., Schröder B., Skidmore A.K., Zurell D., Lautenbach S. (2013): Collinearity: a review of methods to deal with it and a simulation study evaluating their performance Ecography 36: 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dwyer R.G., Krueck N.C., Udyawer V., Heupel M.R., Chapman D., Pratt H.L., Garla R., Simpfendorfer C.A. (2020): Individual and Population Benefits of Marine Reserves

for Reef Sharks. Curr Biol 30: 480–489. https://doi.org/10.1016/j.cub.2019.12.005

Eisinga R., Grotenhuis M.T., Pelzer B. (2013): The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? Int J Public Health 58: 637–642. https://doi.org/10.1007/s00038-012-0416-3

Fontana M.D., de Araújo Moreira F., Di Giulio G.M., Malheiros T.F. (2020): The water-energy-food nexus research in the Brazilian context: What are we missing? Environ Sci Policy 112: 172–180. https://doi.org/10.1016/j.envsci.2020.06.021

Hauke J., Kossowski T. (2011): Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data. Quaestiones Geographicae 30: 87–93. DOI:10.2478/v10117-011-0021-1

Hohenlohe P.A., Funk W.C., Rajora O.P. (2021): Population genomics for wildlife conservation and management. Mol Ecol 30: 62–82. https://doi.org/10.1111/mec.15720

Horváth I.G., Németh Á., Lenkey Z., Alessandri N., Tufano F., Kis P., Gaszner B., Cziráki A. (2010): Invasive validation of a new oscillometric device (Arteriograph) for measuring augmentation index, central blood pressure and aortic pulse wave velocity. J Hypertens 28: 2068–2075. DOI:10.1097/HJH.0b013e32833c8a1a

Hyyppä J., Hyyppä H., Inkinen M., Engdahl M., Linko S., Zhu Y.H. (2000): Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. For Ecol Manage 128: 109–120. https://doi.org/10.1016/S0378-1127(99)00278-9

Iqbal W., Tang Y.M., Chau K.Y., Irfan M., Mohsin M. (2021): Nexus between air pollution and NCOV-2019 in China: Application of negative binomial regression analysis. Process Saf Environ Prot 150: 557–565. https://doi.org/10.1016/j.psep.2021.04.039

Jankowski A., Wyka T.P., Oleksyn J. (2021): Axial variability of anatomical structure and the scaling relationships in Scots pine (*Pinus sylvestris* L.) needles of contrasting origins. Flora 274: 151747. https://doi.org/10.1016/j.flora.2020.151747

Lateef M., Keikhosrokiani P. (2023): Predicting Critical Success Factors of Business Intelligence Implementation for Improving SMEs' Performances: a Case Study of Lagos State., Nigeria. J Knowl Econ 14: 2081–2106. https://doi.org/10.1007/s13132-022-00961-8

Lefsky M.A., Hudak A.T., Cohen W.B., Acker S.A. (2005): Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. Remote Sens Environ 95: 517–531. https://doi.org/10.1016/j.rse.2005.01.004

Lin W.B. (2007): Factors affecting the correlation between interactive mechanism of strategic alliance and technological knowledge transfer performance. J High Technol Manage Res 17: 139–155. https://doi.org/10.1016/j.hitech.2006.11.003

Lindinger-Sternart S., Kaur V., Widyaningsih Y., Patel A.K. (2021): COVID-19 phobia across the world: Impact of resilience on COVID-19 phobia in different nations. Couns Psychother Res 21: 290–302. https://doi.org/10.1002/capr.12387

Min Q., Lu Y., Liu Z., Su C., Wang B. (2019): Machine Learning based Digital Twin Framework for Production Optimization in Petrochemical Industry. Int J Inf Manage 49: 502–519. https://doi.org/10.1016/j.ijinfomgt.2019.05.020

Moews B., Herrmann J.M., Ibikunle G. (2019): Lagged correlation-based deep learning for directional trend change prediction in financial time series. Expert Syst Appl 120: 197–206. https://doi.org/10.1016/j.eswa.2018.11.027

Moradi F., Darvishsefat A.A., Pourrahmati M.R., Deljouei A., Borz S.A. (2022): Estimating Aboveground Biomass in Dense Hyrcanian Forests by the Use of Sentinel-2 Data. Forests 13: 104. https://doi.org/10.3390/f13010104

Nakagawa S., Cuthill I.C. (2007): Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev 82: 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Neumann M., Starlinger F. (2001): The significance of different indices for stand structure and diversity in forests. For Ecol Manage 145: 91–106. https://doi.org/10.1016/S0378-1127(00)00577-6

Noor M.B.T., Zenia N.Z., Kaiser M.S., Al Mamun S., Mahmud M. (2020): Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. Brain Inf 7: 11. https://doi.org/10.1186/s40708-020-00112-2

Nowosad K., Bocianowski J., Kianersi F., Pour-Aboughadareh A. (2023): Analysis of Linkage on Interaction of Main Aspects (Genotype by Environment Interaction, Stability and Genetic Parameters) of 1000 Kernels in Maize (*Zea mays* L.). Agriculture 13: 2005. https://doi.org/10.3390/agriculture13102005

Orsini L., Vanoverbeke J., Swillen I., Mergeay J., De Meester L. (2013): Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol 22: 5983–5999. https://doi.org/10.1111/mec.12561

Pearson K. (1896): Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. Phil Trans R Soc A 187: 253–318.

Pearson K. (1908): On a mathematical theory of determinantal inheritance, from suggestions and notes of the late W. F. R. Weldon. Biometrika 6: 80–93. https://doi.org/10.1093/biomet/6.1.80

Pearson K. (1920): Notes on the history of correlation. Biometrika 13: 25–45.

Piovani J.I. (2008): The historical construction of correlation as a conceptual and operative instrument for empirical research. Qual Quant 42: 757–777.

Puliti S., Breidenbach J., Astrup R. (2020): Estimation of Forest Growing Stock Volume with UAV Laser Scanning Data: Can It Be Done without Field Data? Remote Sens 12: 1245. https://doi.org/10.3390/rs12081245

Rosato A., Tenori L., Cascante M., De Atauri Carulla P.R., dos Santos V.A.P.M., Saccenti E. (2018): From correlation to causation: analysis of metabolomics data using systems biology approaches. Metabolomics 14: 37. https://doi.org/10.1007/s11306-018-1335-y

Rutledge J., Oh H., Wyss-Coray T. (2022): Measuring biological age using omics data. Nat Rev Genet 23: 715–727. https://doi.org/10.1038/s41576-022-00511-7

Saalidong B.M., Aram S.A., Out S., Lartey P.O. (2022): Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. PLoS ONE 17: e0262117. https://doi.org/10.1371/journal.pone.0262117

Samal K., Mahapatra S., Ali H. (2022): Pharmaceutical wastewater as Emerging Contaminants (EC): Treatment technologies, impact on environment and human health. Energy Nexus 6: 100076. https://doi.org/10.1016/j.nexus.2022.100076

Schober P., Boer C., Schwarte L.A. (2018): Correlation Coefficients: Appropriate Use and Interpretation. Anesthesia & Analgesia 126: 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Shimizu I., Kikukawa M., Tada T., Kimura T., Duvivier R., van der Vleuten C. (2020): Measuring social interdependence in collaborative learning: instrument development and validation. BMC Med Educ 20: 177. https://doi.org/10.1186/s12909-020-02088-3

Solon J., Borzyszkowski J., Bidłasik M., Richling A., Badora K., Balon J., Brzezińska-Wójcik T., Chabudziński Ł., Dobrowolski R., Grzegorczyk I., Jodłowski M., Kistowski M., Kot R., Krąż P., Lechnio J., Macias A., Majchrowska A., Malinowska E., Migoń P., Myga-Piątek U., Nita J., Papińska E., Rodzik J., Strzyż M., Terpiłowski S., Ziaja W. (2018): Physico-geographical mesoregions of Poland: Verification and adjustment of boundaries on the basis of contemporary spatial data. Geogr Pol 91: 143–170. DOI:10.7163/GPol.0115

Song H.Y., Park S. (2020): An Analysis of Correlation between Personality and Visiting Place using Spearman's Rank Correlation Coefficient. KSII Trans Internet Inf Syst 14: 1951–1966. http://doi.org/10.3837/tiis.2020.05.005

Spearman C. (1904): The Proof and Measurement of Association between Two Things. The Amer J Psychol 15: 72–101. https://doi.org/10.2307/1412159

Stigler S.M. (1988): Francis Galton's account of the invention of correlation. Stat Sci 4: 73–86.

Thielmann I., Spadaro G., Balliet D. (2020): Personality and prosocial behavior: A theoretical framework and meta-analysis Psychological Bulletin 146: 30–90. https://doi.org/10.1037/bul0000217

Tortella G.R., Rubilar O., Durán N., Diez M.C., Martínez M., Parada J., Seabra A.B. (2020): Silver nanoparticles: Toxicity in model organisms as an overview of its hazard for human health and the environment. J Hazard Mater 390: 121974. https://doi.org/10.1016/j.jhazmat.2019.121974

Tundys B., Bretyn A., Urbaniak M. (2021): Energy Poverty and Sustainable Economic Development: An Exploration of Correlations and Interdependencies in European Countries. Energies 14: 7640. https://doi.org/10.3390/en14227640

Udovičić M., Baždarić K., Bilić-Zulle L., Petrovečki M. (2007): What we need to know when calculating the coefficient of correlation? Biochemia Medica 17: 10–15.

VSN International Genstat for Windows (2023): *VSN International Genstat for Windows*., 23rd Edition; VSN International: Hemel Hempstead, UK.

Walker H.M. (1928): The relation of Plana and Bravais to theory of correlation. Isis 10: 466–484.

Waszak N., Robertson I., Puchałka R., Przybylak R., Pospieszyńska A., Koprowski M. (2021): Investigating the Climate-Growth Response of Scots Pine (*Pinus sylvestris* L.) in Northern Poland. Atmosphere 12: 1690. https://doi.org/10.3390/atmos12121690

Weida F.M. (1927): On various conceptions of correlation. Ann Math 29: 276–312.

Wright I.J., Ackerly D.D., Bongers F., Harms K.E., Ibarra-Manriquez G., Martinez-Ramos M., Mazer S.J., Muller-Landau H.C., Paz H., Pitman N.C.A., Poorter L., Silman M.R., Vriesendorp C.F., Webb C.O., Westoby M., Wright S.J. (2007): Relationships Among Ecologically Important Dimensions of Plant Trait Variation in Seven Neotropical Forests. Ann Bot 99: 1003–1015. https://doi.org/10.1093/aob/mcl066

Wrońska-Pilarek D., Krysztofiak-Kaniewska A., Matusiak K., Bocianowski J., Wiatrowska B., Okoński B. (2023b): Does distance from a sand mine affect needle features in *Pinus sylvestris* L.? For Ecol Manage 546: 121276. https://doi.org/10.1016/j.foreco.2023.121276

Wrońska-Pilarek D., Maciejewska–Rutkowska I., Lechowicz K., Bocianowski J., Hauke–Kowalska M., Baranowska M., Korzeniewicz R. (2023a): The effect of herbicides on morphological features of pollen grains in *Prunus serotina* Ehrh. in the context of elimination of this invasive species from European forests. Sci Rep 13: 4657. https://doi.org/10.1038/s41598-023-31010-2