FEBRUARY 1962
VOL. XI No. 2

TAXON

Official News Bulletin of the International Association for Plant Taxonomy, Edited and Published for I.A.P.T. by the International Bureau for Plant Taxonomy and Nomenclature, 106 Lange Nieuwstraat, Utrecht, Netherlands

THE COMPARISON OF DENDROGRAMS BY OBJECTIVE METHODS[1]

Robert R. Sokal and F. James Rohlf (Lawrence, Kansas)[2]

INTRODUCTION

The purpose of this paper is to present a technique for comparing dendrograms resulting from numerical taxonomic research with one another and with dendrograms produced by conventional methods. One of the most frequent ways of depicting the results of studies in numerical taxonomy (Sokal, 1960; Sneath and Sokal, 1962) is by so-called dendrograms or diagrams of relationships. These are tree-like schemes which indicate the affinity of taxa to their nearest relatives (on the basis of similarity or *phenetic* resemblance alone, without any necessary phylogenetic implications). These diagrams resemble the customary phylogenetic trees, but are preferred for classificatory purposes; first, because phylogenetic inferences are speculative, while similarities are factual; secondly, because they are quantitative evaluations of these similarities; and thirdly, because they lack some of the other meanings often implied in phylogenetic trees (Sneath and Sokal, 1962). Such dendrograms have been published in bacteriological work (Sneath and Cowan, 1958), in studies of bees (Michener and Sokal 1957; Sokal and Michener 1958), butterflies (Ehrlich, 1961), rice (Morishima and Oka, 1960), members of the nightshade genus *Solanum* (Soria and Heiser, 1961) and others.

With the increasing acceptance of the philosophy of numerical taxonomy an experimental phase in using various types of coefficients is beginning, which will involve the comparison of the results of numerical taxonomic research based on these different coefficients. So far we have lacked a procedure for such comparisons. The cophenetic correlations which will be developed below provide an extremely simple and effective method for comparing dendrograms of various sorts.

Before proceeding to a detailed account of the technique, it will be useful to discuss briefly the four types of comparisons of dendrograms that we wish to make in numerical taxonomy and the reasons for them:

1. A major use of the methods proposed below will be to compare dendrograms arrived at on the basis of numerical taxonomic techniques with dendrograms prepared earlier on the basis of conventional taxonomic methods. In a crude manner such a comparison was undertaken by Michener and Sokal (1957). Such techniques will provide some estimate of the magnitude of the differences between numerical taxonomic classifications and the classifications produced on the basis of the currently used methods. While experience to date has shown that numerical taxonomic

[1]) For footnotes see end of article.

classifications are generally not too different from the conventional ones, it is desirable to quantify such comparisons.

2. Recent efforts in numerical taxonomy have resulted in the development of several methods of clustering of taxa from an original matrix of similarity values among taxa. These different methods include those of Sneath (1957), Sokal and Michener (1958) and Rogers and Tanimoto (1960) and are based on somewhat different approaches to cluster analysis. They will therefore necessarily result in somewhat different dendrograms. Within any method many variations of the technique can be employed. Thus the weighted variable group method of Sokal and Michener (1958) can be modified into a pair group method, both weighted and unweighted, and the method of averaging the relations can be modified from the Spearman sums of variables method to one of several possible alternatives. It is important to have a procedure for comparing the dendrograms resulting from these techniques to determine the amount of similarity among them.

3. As a useful by-product, not intended originally, the method developed below turned out to be a valuable test of the amount of distortion of the mutual relations among taxa introduced by a given clustering method. By comparing a dendrogram arrived at by a given technique with the actual similarity coefficients which exist between each pair of taxa within the study, one can obtain a measure of the amount of distortion which the clustering method leading to the dendrogram has imposed upon the system. This will be discussed in greater detail below by means of an example.

4. The occasion may arise when one may wish to compare two different conventional dendrograms of the same taxonomic group with each other. Previously there has not existed an objective and quantitative criterion for expressing the similarity of two dendrograms of the same taxa prepared by different authors. While not related directly to numerical taxonomy, such a comparison will be possible by means of the technique proposed below.

TECHNIQUE

We may start our considerations with a dendrogram as shown in Figure 1 (taken from Rohlf and Sokal, 1962). This represents the weighted pair group method of clustering, applied to 23 species of the bee subgenera *Chilosima* and *Ashmeadiella s. str.* from Sokal and Michener (1958), amended slightly by Rohlf and Sokal (1962). While in this example the units being clustered are species, other taxonomic units can be clustered by numerical taxonomy. The lowest ranked units possible would be individuals, which could be clustered to form groups of individuals phenetically resembling each other and representing infraspecific taxa, if all the individuals concerned were conspecific. Higher taxonomic entities, such as genera, tribes or families, could be used as units in a taxonomic study and they could be grouped to form still higher taxa. Sokal and Sneath (book, in preparation) have therefore called the lowest taxonomic unit being classified in any given study an *operational taxonomic unit* (OTU). These OTU's conventionally represent the tips of the dendrogram in any given study.

In a dendrogram the abscissa has no particular meaning, except for spacing out the original taxa employed in the study. The ordinate, on the other hand, represents similarity values, which may be on one of several conventionally used scales. Thus they may be association or resemblance coefficients (Sneath 1957; Sneath and Sokal, 1962), correlation coefficients (Sokal and Michener, 1958), as in Figure 1, or distances (Sokal, 1961). The upper end of the scale will represent maximum similarity, unity in the case of correlation and association and zero with distances, while the lower end of the scale will extend as far as is necessary in order to unite all stems in the particular dendrogram. The meaning of such a dendrogram is restricted to showing the level at which two or more stems join with each other to form a common stem. It should be clear that from such a diagram one can only obtain a rough idea of the relation of every OTU with every other one. Any given pair of OTU's may be more closely or more distantly related than is indicated by the level of the junction of the stems bearing them. This is so because the level of junction represents the average resemblance of the OTU's of one cluster with the OTU's of the second cluster.

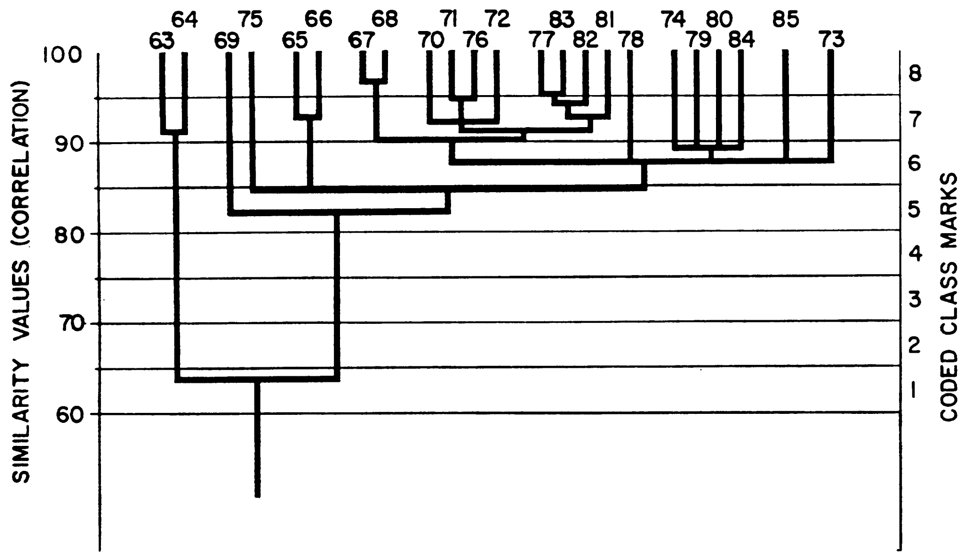

Fig. 1. Dendrogram or diagram of relationships among 23 species of the bee subgenera *Chilosima* and *Ashmeadiella s. str.*, taken from Rohlf and Sokal (1962) and based on data by Sokal and Michener (1958). The relationships were obtained by the weighted pair group method (WPGM). The ordinate is graduated in a Pearson product-moment correlation coefficient scale (coded by multiplication by 100). Numbers across the top of the figure are species code numbers which are identified in Rohlf and Sokal (1962) or Michener and Sokal (1957). Horizontal lines across the dendrogram are phenon lines defining taxa at the minimum level of similarity at which the phenon line cuts the ordinate. Class intervals along the similarity scale delimited by phenon lines have had their class marks coded (on the right hand side of the dendrogram).

In order to avoid misunderstandings it should be clearly stated that the word "related" in the context of this paper means only similar, and does not necessarily infer relationship by descent. According to the principles of numerical taxonomy (Sneath and Sokal, 1962), the process of classification is based only on observed differences and similarities, so-called phenetic evidence, and not on phylogenetic inferences and speculations, however important and interesting these may be from other viewpoints.

In order to avoid the historical and semantic connotations of the conventional system of higher categories (i.e. subgenera, genera, tribes families, etc.), Sneath and Sokal (1962) have introduced the idea of a *phenon* (distinct from the phenon concept of Camp and Gilly, 1943), which defines taxa by drawing horizontal lines across the dendrograms. All taxonomic units carried by single stems crossed by such a phenon line are called phenons with a numerical prefix indicating the vertical level of the phenon line. Thus, as is shown in Figure 1, there are five 85-phenons and two 75-phenons in our present study. In this manner we are able to define objectively the limits of any given taxon and avoid arguments as to what a genus, tribe or any other category is in a given group, since the level of cohesion of the group can be arbitrarily fixed by the investigator interested in defining the group. We are, however, prevented from drawing other than horizontal lines, since these would conflict with the aim of objectivity and repeatability of numerical taxonomy.

In devising methods of comparing dendrograms one of the initial schemes which occurred to us, but which we have so far not followed up, was to try to compare different arrangements of the same OTU's by counting the number of breaks and rearrangements necessary to convert one dendrogram into another. We have since adopted the simpler approach shown below.

We divide the range of similarity values along the ordinate into a suitable number of equal class intervals by drawing phenon lines as class limits across the dendrogram (see Figure 1, where the range of similarity values has been divided into eight classes). The number of classes into which the range of variation should be divided will depend upon the number of OTU's being classified. As a very rough guide, dendrograms involving less than ten OTU's need not be divided into more than four classes, while dendrograms involving as many as 100 OTU's should probably be divided into at least ten classes. A further consideration should be that the class intervals should be fine enough to reveal a reasonable amount of structural detail in the dendrogram to be analyzed. Persons planning to do such computations on a desk calculator should employ the minimum number of classes necessary. On the other hand, increasing the number of classes never does any harm from a statistical point of view. As a matter of fact the computer program which was developed by one of us (F.J.R.) divides the range of similarity values into 50 classes. Schemes could also be developed which would handle the actual similarity value at which two stems join. The probable statistical consequences of using actual juncture levels rather than cophenetic values are probably slight, based on the well known effect of grouping in frequency distributions.

Once the class intervals along the ordinate have been established, each class mark should be coded on a scale starting with unity at the lower end, i.e. the end having the lowest similarity value, and going up in unit steps. Thus with ten classes the highest class should be coded 10. These values will then be proportional to the similarity values, except in the case of distances, where they will be complementary and where coding in reverse, i.e. starting with unity at the highest level, might be advised. The coding is a computational convenience for desk calculator operations. Actual class marks can be used in digital computer programs.

We shall define the *cophenetic value* of two OTU's as the class mark of the class (between phenon lines) in which their stems are connected. For example, in Figure 1 we can see that species 63 and species 64 are connected in class interval 7. Hence their cophenetic value is 7. Similarly the cophenetic value of species 69 with species 71 is 5, since that is the level at which these OTU's are connected. The closer the relationship between OTU's, the higher will be their cophenetic values. It is convenient to record cophenetic values in matrix form, resembling a matrix of similarity values (see Table 1). The actual procedure of recording the cophenetic values for each pair of OTU's in a study, may be appreciably simplified, since the cophenetic values between all OTU's on any one stem with all other OTU's are identical. Thus, for example, in Figure 1 species 63 and 64 will have a cophenetic value of 1 with all of the remaining species.

TABLE 1.

	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
63	x																						
64	7	x																					
65	1	1	x																				
66	1	1	7	x																			
67	1	1	5	5	x																		
68	1	1	5	5	8	x																	
69	1	1	5	5	5	5	x																
70	1	1	5	5	7	7	5	x															
71	1	1	5	5	7	7	5	7	x														
72	1	1	5	5	7	7	5	7	7	x													
73	1	1	5	5	6	6	5	6	6	6	x												
74	1	1	5	5	6	6	5	6	6	6	6	x											
75	1	1	5	5	5	5	5	5	5	5	5	5	x										
76	1	1	5	5	7	7	5	7	7	7	6	6	5	x									
77	1	1	5	5	7	7	5	7	7	7	6	6	5	7	x								
78	1	1	5	5	6	6	5	6	6	6	6	6	5	6	6	x							
79	1	1	5	5	6	6	5	6	6	6	6	6	5	6	6	6	x						
80	1	1	5	5	6	6	5	6	6	6	6	6	5	6	6	6	6	x					
81	1	1	5	5	7	7	5	7	7	7	6	6	5	7	7	6	6	6	x				
82	1	1	5	5	7	7	5	7	7	7	6	6	5	7	7	6	6	6	7	x			
83	1	1	5	5	7	7	5	7	7	7	6	6	5	7	8	6	6	6	7	7	x		
84	1	1	5	5	6	6	5	6	6	6	6	6	5	6	6	6	6	6	6	6	6	x	
85	1	1	5	5	6	6	5	6	6	6	6	6	5	6	6	6	6	6	6	6	6	6	x

Matrix of cophenetic values for figure 1.

We are now able to proceed to the actual comparison of dendrograms. This is done quite simply by calculating an ordinary product-moment correlation coefficient between the corresponding elements of the two matrices of cophenetic values to be compared. We propose to call these coefficients *cophenetic correlations*. For these procedures the half-matrix can be imagined as strung out in single file, column by column. For n OTU's there will be $n\ (n-1)/2$ elements in a similarity coefficient matrix. When the number of classes is rather small a coefficient of association will have to be used between the cophenetic values to be compared. However, in studies of any magnitude, where we are likely to deal with six or more classes per range of similarity values (and thus a two-way frequency distribution involving 36 classes), ordinary product moment correlation is indicated. This can be done by desk calculation, but will preferably be done on a computer. Some transformation of the correlation coefficients may be necessary in order to insure homoscedasticity.

In case the two dendrograms to be compared contain OTU's which are not common to them, one should simply ignore the non-coincident OTU's and proceed as before.

In small studies, i.e. less than 10 OTU's, it is computationally simplest to perform the operations on a desk calculator. We have found that in larger studies it is more efficient to use a digital computer. An I.B.M. — 650 computer program (DENDRON I) developed by one of us (F.J.R.) may be obtained by writing to the authors. The program was designed to accept the output of our Weighted Pair Group Method (WPGM) programs (TAXON I and II), prepares dendrograms from these and computes cophenetic values if desired.

EXAMPLES

The correlation matrix of 23 species of the subgenera *Chilosima* and *Ashmeadiella s. str.* presented by Sokal and Michener (1958) was used as a test case for comparing the various dendrograms which resulted from several methods of clustering used by these authors.

The dendrogram resulting from the weighted pair group method is illustrated in Figure 1. It was compared with the weighted variable group method, the dendrogram of which can be found in figure 1 of Sokal and Michener (1958). Matrices of cophenetic values based on eight classes were prepared. Only eight classes were chosen to make the example simple to inspect and because the first trials of the method were performed on a desk calculator. Fifty classes were used in a repetition of some of the calculations on a computer, resulting in essentially identical cophenetic correlations. The matrix of cophenetic values for the weighted pair group dendrogram of Figure 1 can be seen in Table 1. A similar matrix was prepared for the weighted variable group dendrogram, but is not shown here. Figure 2 shows a two-way frequency distribution of the 253 pairs of species plotted with respect to the two sets of cophenetic values. From this figure we can see the generally high correlation between the two sets of cophenetic values, representing substantial agreement between the two dendrograms. When a Pearson product-moment correlation coefficient was computed for the two-way frequency distribution, a correlation of 0.95 was obtained. Cophenetic values based on eight classes were also computed for the dendrogram constructed by Sokal and Michener (1958) using the unweighted variable group method. A comparison of these cophenetic values with those discussed above is shown expressed as correlation coefficients in Table 2.

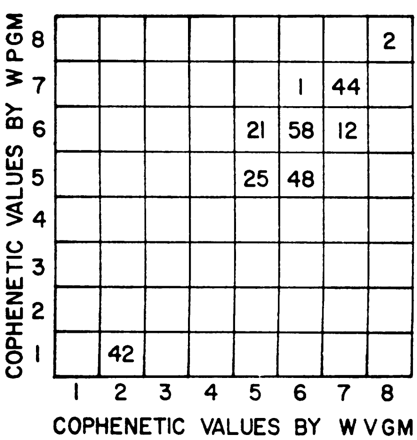

Fig. 2. Two-way frequency distribution of cophenetic values among the 23 species of bees for weighted pair group method (WPGM) and weighted variable group method (WVGM).

Also included in Table 2 is the weighted pair group method with the correlation between joining stems computed by an ordinary arithmetic average of the correlation coefficients, rather than by Spearman's sums of variables method. The last row and column of the correlation matrix in Table 2 represents correlations between the original correlation coefficients (between pairs of species) and the four sets of cophenetic values. These correlations give an indication of the amount of information lost by presenting the correlation matrix in the form of a dendrogram (i.e. the degree of the distortion between the original correlations and the cophenetic values).

TABLE 2.

	WPGM	WVGM	UVGM	WPGA	r
WPGM	x	.95	.84	.86	.80
WVGM	.95	x	.87	.89	.82
UVGM	.84	.87	x	.92	.83
WPGA	.86	.89	.92	x	.86
r	.80	.82	.83	.86	x

Matrix of correlation coefficients among the cophenetic values resulting from four dendrograms and the original correlation coefficients used to prepare the dendrograms. WPGM — weighted pair group method, WVGM — weighted variable group method, UVGM — unweighted variable group method, WPGA — weighted pair group method using arithmetic averages, and r — the original correlations.

Inspection of Table 2 shows that in general the correlation between the original correlation coefficients and the dendrograms resulting from them are high (in the 0.80's range). Of the various clustering methods that of averaging correlation coefficients, rather than Spearman's sums of variables method, seems to show a somewhat closer correlation with the original similarity values and hence the least amount of distortion of the original data. When we consider the various types of dendrograms we note that in general they are more correlated with each other than they are to the original correlations. It would appear that the closest relation between dendrograms is between the weighted pair group method and the weighted variable group method, as has been pointed out before. This validates the choice of the weighted pair group method for computational purposes as being essentially identical with the weighted variable group method, which has been chosen as more suitable for desk calculator operation.

We shall not pursue the implications of the results of this analysis in this paper. It should be obvious that by the use of cophenetic correlations we have a powerful method for evaluating the relative similarities of various dendrograms and the distortions due to different types of cluster analysis. The relative merits of different methods of clustering will be discussed in a later paper.

SUMMARY

The method developed in this paper makes quantitative comparisons among taxonomic dendrograms or diagrams of relationships. Dendrograms obtained by numerical taxonomy can be compared with each other, with dendrograms constructed by conventional taxonomic methods and with the original coefficients of similarity from which a given dendrogram has been derived. The method of comparison is based on the drawing of equidistant phenon lines horizontally across a dendrogram and on the computation of cophenetic values, which are the coded class marks of similarity classes in which any two taxa in a dendrogram are connected. Examples of the method are shown, using a portion of Michener and Sokal's bee data.

LITERATURE CITED

CAMP, W. H. and C. L. GILLY. 1943. The structure and origin of species. Brittonia 4: 323-385.

EHRLICH, P. R. 1961. Has the biological species concept outlived its usefulness? Systematic Zool. 10: 167-176.

MICHENER, C. D. and R. R. SOKAL. 1957. A quantitative approach to a problem in classification. Evolution 11: 130-160.

MORISHIMA, H. and H. OKA. 1960. The pattern of interspecific variation in the genus *Oryza:* its quantitative representation by statistical methods. Evolution 14: 153-165.

ROGERS, D. J., and T. T. TANIMOTO. 1960. A computer program for classifying plants. Science 132: 1115-1118.

ROHLF, F. J. and R. R. SOKAL. 1962. The description of taxonomic relationships by factor analysis. Systematic Zool. (in press).

SNEATH, P. H. A. 1957. The application of computers to taxonomy. J. Gen. Microbiol. 17: 201-226.

——. and S. T. COWAN. 1958. An electro-taxonomic survey of bacteria. J. Gen. Microbiol. 19: 551-565.

——. and R. R. SOKAL. 1962. Numerical taxonomy. (Nature, in press).

SOKAL, R. R. 1960. Die Grundlagen der numerischen Taxonomie. Proc. XI. Intern. Congr. Entomol. (in press).

——. 1961. Distance as a measure of taxonomic similarity. Systematic Zool. 10: 70-79.

——. and C. D. MICHENER. 1958. A statistical method for evaluating systematic relationships. Univ. of Kans. Sci. Bull. 38: 1409-1438.

Soria, J., and C. B. Heiser, Jr. 1961. A statistical study of relationships of certain species of the *Solanum nigrum* complex. Econ. Botany 15: 245-255.

FOOTNOTES

1. Contribution No. 1131 from the Department of Entomology, The University of Kansas, Lawrence, and paper No. 3 in a series entitled "Experiments in Numerical Taxonomy". Paper No. 1 in this series is Sokal (1961) and paper No. 2 is Rohlf and Sokal (1962).

2. This investigation was carried out during the tenure by F. J. Rohlf of a predoctoral fellowship from the Division of General Medical Sciences, United States Public Health Service. The authors are indebted to the University of Kansas General Research Fund for making computer time available at the University of Kansas Computation Center. Mrs. Ann Schlager prepared the drawings. Prof. C. D. Michener of the University of Kansas has read the manuscript and contributed useful suggestions.

GIBT ES EINE REGELUNG FÜR DIE SCHREIBWEISE DER AUTORNAMEN?

Werner Rothmaler (Greifswald)

In den letzten Jahren sind wir dank zielbewußter Arbeit der IAPT und des Intern. Bureau for Plant Taxonomy and Nomenclature gut vorangekommen und haben mancherlei Ordnung in Köpfen und Büchern, in Systemen und Floren schaffen können. Doch scheint mir die Frage des abgekürzten Zitats durch Nennung des Autornamens in der Nomenklatur nicht genügend reglementiert und geordnet.

Es genügt zwar auch meines Erachtens meist der Autorname selbst, anstelle eines vollen Zitats, doch muß der Autorname wirklich erkennbar sein. Wieviel Abkürzungen aber gibt es schon für klassische Autoren und vor allem Autorenpaare, ganz zu schweigen von den Abkürzungen, die sich junge Autoren ausdenken, ohne wirklich ausreichend untersucht zu haben, ob hier jede Verwechslungsmöglichkeit vermieden ist. Wievielerlei Gmelins werden einfach mit Gmel. und welche Schmidts mit Schm. abgekürzt? Neuerdings hat man hier schon durch Vorsetzen der Initialen Ordnung zu schaffen begonnen, doch ist das noch keineswegs ganz gelungen, wie ich bei der Vorbereitung der Bände der Exkursionsflora von Deutschland erkennen mußte. Bei Artnamen ist es dabei noch verhältnismäßig leicht, Klarheit aus dem Index kewensis und ähnlichen Werken zu gewinnen. Sonst aber ist es außerordentlich schwer, wenn als Autor Mey. oder Schm. angegeben wird.

Es erscheint mir doch notwendig, eine einheitliche und dauerhafte Regelung zu schaffen. Aber auch für andere gleichlautende Autorennamen oder ihre Abkürzungen sollte man eine Regelung schaffen. Lam und Lam., Murr und Murr., Moll und Moll., Vent und Vent. unterscheiden sich nur im Punkt am Ende des Namens, der am Satzende entfällt und nicht mehr zu erkennen ist. So sollte man Lamk. schreiben, um Verwechselungen mit H. J. Lam zu verhindern. Man sollte J. Murr im Gegensatz zu Murr(ay) oder W. Vent gegenüber Vent(enat) schreiben. Es würde auch die Arbeit erleichtern, wenn man eine einheitliche Regelung für die Abkürzung von Waldstein und Kitaibel (W.K., W. et K., Waldst. et Kit.) oder von Boissier und Reuter (B.R., B. et R., Boiss. et Reut.) schaffte. Auch die Autoren namens Schwarz müßten wohl ihre Initialen hinzugefügt bekommen, damit A. und O., W. und H. Schwarz unterscheidbar werden.

Ich möchte meinen, daß sich das International Bureau for Plant Taxonomy and Nomenclature mit diesen Fragen befassen müßte. Es sollte eine Liste der Autorennamen und der anzunehmenden Abkürzungen zusammengestellt werden, wobei weit-