# Predicting Phosphorus NMR Shifts Using Neural Networks. 2. Factors Influencing the Accuracy of Predictions

Geoffrey M. J. West

School of Computing and School of Sciences, Staffordshire University, Staffordshire, England

Received November 16, 1993®

Further studies on a novel shift prediction method based on neural networks are described. While illustrated using phosphorus-31 NMR the method only requires that the element's shift is moderated by its $\alpha-\delta$ substituents. Focal region topologies are encoded in a novel substructure code. The network's input vectors are then formed by replacing the code symbols with some numeric property. The network learns to output the shift associated with a given input vector. The results of the studies within show that the prediction method evolved by the network has similar dependencies to those of the $^{13}C$ additivity rules.

## INTRODUCTION

Carbon-13 nuclear magnetic resonance (NMR) chemical shifts are commonly predicted using either atom centered substructure codes[1−3] or additivity rules.[4] Code based predictions are integral to the strategy of many computer assisted structure elucidation (CASE) systems.[5−7] Rule based predictions are usually applied manually, although automated application is an increasing trend.[8−11] While both methods differ substantially in how shifts are derived both rely on $\alpha-\delta$ substituent effects[12,13] and have inherent strengths and weaknesses. Additivity rules offer the convenience of an equation based approach but suffer from a highly fragmented application domain. Here each distinct realm requires a unique, and usually manually derived, set of parameters. As new compounds are discovered within the realms, these sets become obsolete. Substructure codes offer the conveniences of a data driven approach, with most procedures being easily automated. Here one major task is data maintenance as two main factors influencing accuracy are the quality and quantity of the data. Unfortunately most large databases have these two factors inversely related, and chemical databases are no exception.[14] Relative to the additivity rule programs, a data driven approach can also suffer from slower prediction times. This can be debilitating in CASE systems when the estimated spectra are used to prune large candidate lists.

A previous paper introduced a new method of predicting NMR shifts using an artificial neural network (ANN),[15] which is trained to output a shift when given a topological representation. In its prediction mode the network has the speed and convenience of an equation driven approach. In its learning mode it has the convenience of a data driven approach, in automatically extracting a "rule" set from the data. These "rules" are stored succinctly by the network in its weight vector. While illustrated using 31-phosphorus ($^{31}P$) NMR, the method requires no $^{31}P$ specific parameters. It does, however, presuppose that the NMR of the element under consideration shows $\alpha-\delta$ substituent effects, and this, together with evidence from other sources,[16−19] suggests the potential of applying the method to $^{13}C$ NMR. While $\alpha-\delta$ substituent effects are common to $^{31}P$ and $^{13}C$ NMR, however, there are also some differences. Currently, apart from heuristically based methods, no other general method for predicting $^{31}P$ shifts exists. This, and the ubiquitous nature of $^{13}C$ NMR, makes it highly probable that most comparisons will be made to $^{13}C$ prediction methods. These should take into account the differences between the two elements, and, to facilitate this, some details on $^{31}P$ NMR are given before introducing the method and its terminology.

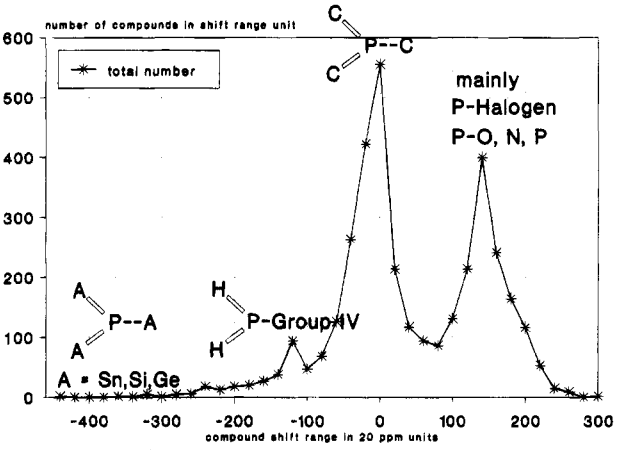


**Figure 1.** Shift distribution 3c3v class. Compound shifts plotted in 20 ppm ranges.

**Variability and Accuracy of Reported $^{31}P$ NMR Shifts.** Phosphorus can commonly exist in several valence states, and specialists normally focus on a particular coordination and valence class. Older $^{31}P$ shift values, where the literature does not state the reference compound, must be treated cautiously since, before standardizing on 85% $H_3PO_4$, various references were used. A more serious problem concerns the polarity of $^{31}P$ NMR shifts, as in 1976 the accepted polarities were reversed.[20] This is exacerbated by the $H_3PO_4$ $^{31}P$ shift being located in the middle of the shift range of most phosphorus classes. The shifts of three valent phosphorus compounds span well over 1200 ppm.[21] This work uses the three coordinate three valent (3c3v) class whose shifts span 750 ppm, and whose shift distribution is shown in Figure 1. When comparisons are made to existing $^{13}C$ predicting methods, simply scaling the results using the ratio of the $^{31}P$:$^{13}C$ shift ranges is unlikely to be reliable. The following summary is intended to show the variability of shifts in the 3c3v class, with this abridged information obtained from phosphorus specialists and the $^{31}P$ database.[15]

® Abstract published in *Advance ACS Abstracts,* December 15, 1994.

0095-2338/95/1635-0021$09.00/0 

The reported $^{31}P$ shifts of the same compound in the same solvent commonly vary by over 2 ppm. In different solvents this variability is often over 5 ppm. Once obvious polarity errors are eliminated, the $^{31}P$ database has 343 3c3v compounds with more than one recorded shift value. These additional shifts are not (knowingly) attributable to stereoisomerism. These 343 compounds show an average difference of 8.9 ppm between their shifts, with over 60% having differences greater than 2 ppm. Like most topological databases, the $^{31}P$ database contains only some of the data on stereoisomers. The 3c3v class has 104 distinct geometric stereoisomer groups with an average intragroup difference of 21.68 ppm. Over 70% of groups have an intragroup difference greater than 5 ppm.

**Prediction Method, Summary, and Terminology.** One assumption made is that a mapping can be used to model the topology-shift relationship. It is well established that neural networks can learn to represent such mappings,[22,23] although many factors, often problem specific, can affect their learning ability. Compound topology in the locale of the resonating $^{31}P$ focus is represented by a novel substructure code, and the network is given a derivative of this code and trained to output the focal $^{31}P$ shift. This work restricts the locale to, at most, the focal $\delta$ substituents and excludes any bond order data. The type of network used has a predefined, fixed architecture and requires any input to have a fixed number of components. Topological sizes are standardized by translating the focal locale onto a graph ***template***. Translation starts from the focus and radiates outwards toward the $\delta$ substituents. For this work the three templates of Figure 2 were used. For template$_{33\alpha\beta\gamma}$ the focus is allowed a maximum of 3 $\alpha$, 9 $\beta$, and 27 $\gamma$ substituents. for template$_{33\alpha\beta\gamma}$ and template$_{33\alpha\beta\gamma\delta}$ these numbers are 3 $\alpha$, 9 $\beta$ and 3 $\alpha$, 9 $\beta$, 27 $\gamma$, 81 $\delta$, respectively. In this work all substituents are restricted to a maximum coordination of four. For 3c3v class compounds and Figure 2 templates, translation causes the three template nodes allocated to the $\alpha$ substituents to become ***occupied***. Each node occupied by an $\alpha$ substituent has three nodes directly linked to it. These hold any directly attached $\beta$ substituents. When the number of directly attached $\beta$ substituents is less than three, the corresponding unoccupied template nodes are filled with the "−" symbol. Figure 3 shows the translation of one substituent "unit", where a unit is defined as the group of substituents that are (a) more distant from the focus than and (b) directly attached to Z, the current atom. The substituents in a unit are ordered on their extended connectivity (EC) score during translation. These scores are obtained by applying the Morgan algorithm to the compound's hydrogen explicit matrix. Translation starts with the focus as the current atom and, for template$_{33\alpha\beta\gamma}$, would be complete when the lowest scoring $\beta$ substituent attached to the lowest scoring $\alpha$ substituent had its unit of $\gamma$ substituents, if any, ordered and translated. Cyclic compounds are treated as if they were acyclic with one difference. When a cycle defining bond is detected, the incident atoms are flagged. When a flagged atom occurs in a substituent unit, its EC score is still accounted for in the ordering of substituents, but the node assigned to it is filled with the "−" symbol. When completed, the template representation is itself translated into a ***molecular abstract graph space*** (MAGS) substructure code. There are three different ***formats*** of this code, linear, branch, and random, with the formats differing in the relationship between the

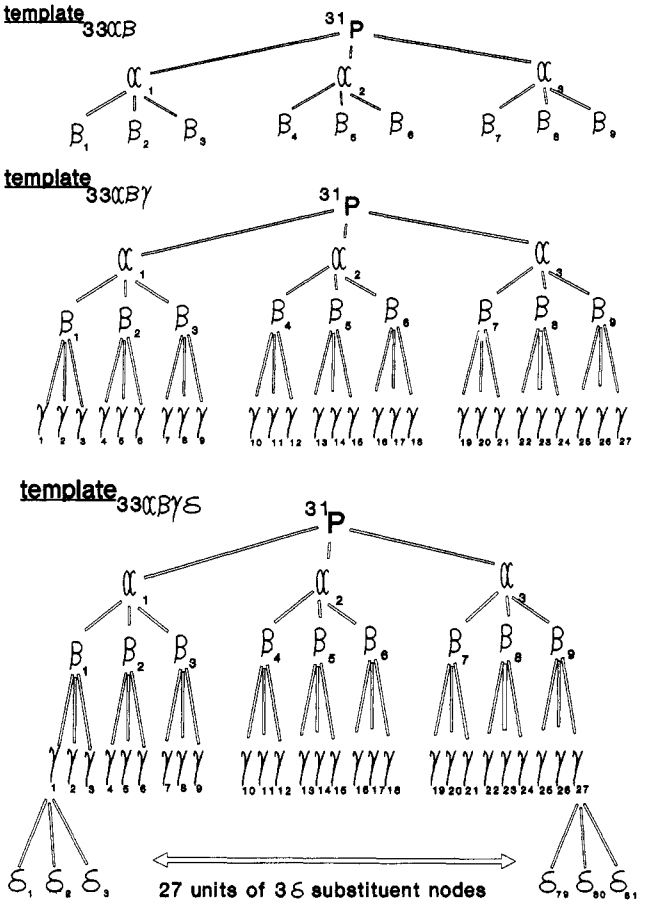


**Figure 2.**

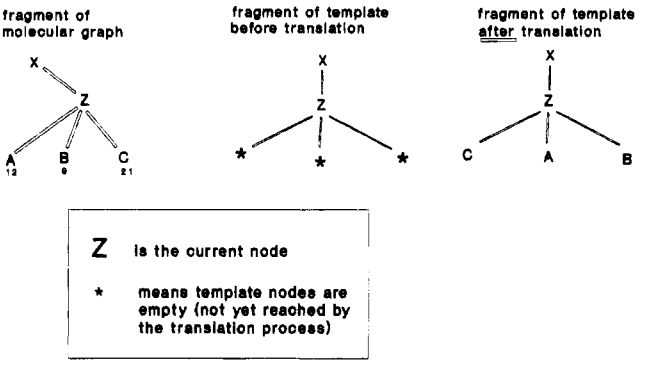


**Figure 3.**

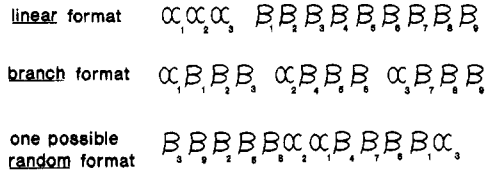


**Figure 4.** Node−position relationship, MAGS code formats derived from template$_{\alpha\beta}$.

code ***positions*** and the template nodes. These relationships are illustrated in Figure 4 for template$_{33\alpha\beta}$ derived codes. Figure 5 shows the linear format of MAGS code derived when the compound number 2525 is translated onto template$_{33\alpha\beta\gamma}$.

Four distinct operations remain before the construct given to the network is obtained. First, from the set of representations, any one to many mappings are eliminated. The source of these are subsets where the subset members have an identical MAGS code and different shift values. Such

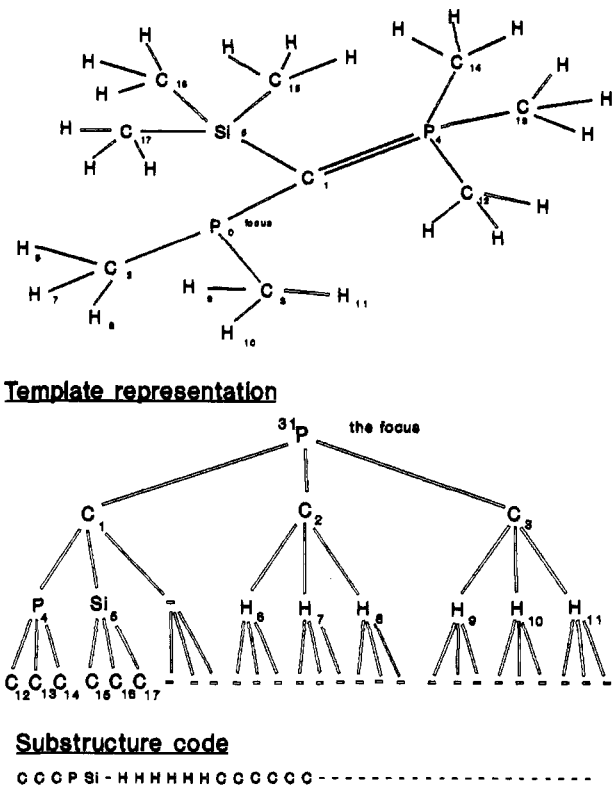


**Figure 5.**

subsets can cause network oscillation during training.[15] The required ***unique set*** of representations is obtained by (a) averaging the shifts within a collection, (b) removing the duplicates, and (c) assigning the subsets average shift to the surviving member. The second operation concerns the prefixing of a five element ring vector to the MAGS code. This is a conjectured means of encoding the involvement of the focus in any ring system(s). In this work only one vector is prefixed, and, where multicyclic systems occur, the smallest ring size is given. The ring vectors are listed in Table 1. Thirdly, the MAGS code symbols are replaced with, in this work, a single numeric value. A file containing a set of symbol-number replacements is termed a ***symbol replacement file*** (SRF). Table 2 lists the SRF files used in this work, all of which have a replacement value of zero for the "−" symbol. The fourth procedure is scaling the shift value associated with each MAGS code. In the 3c3v class, shifts range from [293.7 to −455.5] ppm, while networks use a transfer function with a [0−1] output range. To avoid the extremes of the transfer function's range, shifts are transformed to a range of [0.1−0.9]. In practice the four operations above occur concomitantly. When a unique set is created both the MAGS code *and* the ring vector are taken into account, allowing cyclic and acyclic compounds with identical MAGS codes to survive. The outcome is a unique set with compound vectors of a particular ***type***. Each compound vector consists of (a) the ring vector, (b) the numeric MAGS code, and (c) the scaled shift value. Parts (a) and (b) constitute what is input to the network or the ***input vector*** and part (c) the desired output or ***output vector***. Three factors, all concerned with the MAGS code, determine the type of a compound vector. These are (a) the template used, (b) the format of MAGS code, and (c) the SRF file used. A type is identified by the superscripted template label preceding, and the subscripted SRF label following, the format of the MAGS code. Thus the $^{33\alpha\beta\gamma\delta}\text{linear}_{\text{map\_eneg}}$ type has a linear code format, derived by translation onto $\text{template}_{33\alpha\beta\gamma\delta}$ with the element symbols replaced by their electronegativity values mapped to a [0−0.9] range. The $^{33\alpha\beta\gamma\delta}\text{linear}_{\text{map\_eneg}}$ type has an input vector of 120 components subdivided as follows: five components for the ring vector; three for any $\alpha$ substituents; nine for any $\beta$ substituents; 27 for any $\gamma$ substituents; and 81 for any $\delta$ substituents. Hereafter the terms "unique set" and "type" are used interchangeably to refer to the unique set of compound vectors of a particular type. For this work each unique set was divided into three subsets, one large one for training the network and two smaller ones for monitoring the trained network's performance. For any template, the sizes of these subset ***partitions*** are kept constant, and are $\text{template}_{33\alpha\beta}$ 860, 200, 200, $\text{template}_{33\alpha\beta\gamma}$ 1748, 250, 250, and $\text{template}_{33\alpha\beta\gamma\delta}$ 2109, 300, 300. Partitioning a unique set was achieved by the random selection process described in ref 15.

**The Purpose of the Experiments.** Throughout this work the term "experiment" refers to a *computer based* experiment which, like its laboratory counterpart, is distinguished by its parameters or conditions. This work restricts these parameters to (1) the different network architectures used; (2) the different types used; and (3) the different modified types or ***variants*** used in some monitoring subsets, the form of which is described later. As data errors may bias the results, experiments are repeated using different partitions of the unique set. Such an experiment collection, identical apart from the partitions, is termed an ***experiment category***. The category average results are used in making any comparisons. The categories in this work belong to one of three groups, each group addressing a factor(s) affecting network performance. These factors are (a) whether performance is a function of the environment size contained in the MAGS code; (b) whether performance depends on the values replacing the element symbols; and (c) whether substituent distance from focus affects the contribution to performance. Viewed from a chemical perspective this may be seeking to prove the axiomatic. Any predicting ability acquired by a network is, however, not restricted to solutions with chemically significant and interpretable means.[24]

Previous work showed little learning occurred in networks trained with control types which either (a) lack a ***structured*** format of MAGS code or (b) contained no substituent data.[15] Here the term "structured" refers to the linear or branch code formats which have (EC) ordered positions. A unique set of types with a structured code format has the ***positional equivalence property*** (PEP). This refers to the equivalence of the code positions across the set with respect to the compound topologies. This property's importance was demonstrated before, using $\text{template}_{33\alpha\beta\gamma}$ derived types. The group A categories extend this work to $\text{template}_{33\alpha\beta}$ and $\text{template}_{33\alpha\beta\gamma\delta}$ derived types, to answer the following questions: (a) is importance of the PEP consistent across templates and (b) is performance enhanced by increasing the template and, hence, the environment size. The group B categories sought to determine whether performance is a function of the values that replace symbols. The sole difference between group B categories is in the SRF file used to create the categories type. These SRF files can be subdivided by their replacement values being a scaling of (a) the atomic number, (b) the electronegativity value, or (c) a single global replacement value. Scaled atomic

numbers uniquely identify each element to the network, while, at the other extreme, a global replacement value makes the elements indistinguishable. Electronegativity values have greater chemical significance as there is a well-known correlation between them and shift values. The different scalings are used because the rate at which a network evolves a solution can be affected by the magnitude of its inputs. This is particularly important for atomic number replacements, as the most commonly occurring elements have low atomic numbers. In the [0−0.9] scale the value for the second most common element (in the 3c3v MAGS codes) hydrogen is very low. Different scalings of the three relationships are used to detect such a "magnitude" effect. Aside from the effects of the extremes of input size, or those from data errors, categories differing only in the scale of the replacement values should *eventually* reach identical performance maxima, because (a) the internal ratios between values in the category types are identical and (b) the network weights can have any real numbered value. This means that if a network was trained using two identical training sets which differed only in the magnitude of their inputs the network weights would simply scale accordingly. Categories in group C differ only in the format of the types in the monitoring sets. One category contains the unmodified monitoring sets of the partitions, and four categories contain ***variant*** sets derived by modifying the sets in the first category. The changes involve setting positions in MAGS code ***region(s)*** to zero. A type derived from template$_{33\alpha\beta\gamma}$ has three regions, one each for the $\alpha$, $\beta$, and $\gamma$ substituent data. The use of altered monitoring sets is required because the unique sets derived from the Figure 2 templates all have different characteristics. The differences are introduced when any duplicates are removed and can be measured by the ***element percentage profile*** (E%P) of a unique set. The E%P is the percentage that each element (and "−") occupies each code position. Comparing these profiles shows that differences, while in general small, do exist. The group A results cannot therefore conclusively determine the effects of increasing the environment size. In group C, however, a network is trained and then monitored on an unaltered set and its four variants, reducing any differences to those between the altered regions.

## EXPERIMENTAL SECTION

**1. Common Conditions.** All networks are fully connected, have a bias on each noninput neuron, and use the $(1/[1 + e^{-NET}])$ transfer function throughout. Learning rate and momentum values are static at 0.02 and 0.5, respectively. [See ref 15 for an explanation of why these unusually low values were used.] Once trained, a network is monitored using the test and evaluation subsets of a partition and the performance expressed as the percentage of vectors where the difference between predicted and actual shifts is (a) inside ±20 ppm, (b) inside ±40 ppm, (c) outside ±80 ppm, and (d) outside ±100 ppm. Results are reported as ***average by Epoch*** values for a category. These are calculated according to the formula shown below where

$$CP^E = \frac{1}{k}\sum_{i=1}^{i=k} XP_i^E$$

where $CP^E$ is the category performance at Epoch E, $XP_i^E$ is the performance using the $i$th partition in the category, and $k$ is the number of partitions or individual experiments per category.

**Table 1.** Ring Vectors Encoding the Sizes of Focal Ring Systems

| ring size data | ring vector | | | | |
|---|---|---|---|---|---|
| 3 | 1 | 1 | 1 | 1 | 1 |
| 4 | 0 | 1 | 1 | 1 | 1 |
| 5 | 0 | 0 | 1 | 1 | 1 |
| 6 | 0 | 0 | 0 | 1 | 1 |
| 7 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 |

**2. Unique Conditions. Group A.** The 36 [including template$_{33\alpha\beta\gamma}$ derived types from previous work[15]] group A categories are distinguished by the architecture and type used. To allow comparisons, analogous results from template$_{33\alpha\beta\gamma}$ are included. Four types were derived from each template: the $^{X}$linear$_{map_eneg}$ and $^{X}$branch$_{map_eneg}$ structured types and the $^{X}$random$_{map_eneg}$, and $^{X}$linear$_{zero}$ controls. Here X is a Figure 2 template. Each group A category contains five partitions. Networks were trained for 4000 Epochs, with the network's configuration saved every epoch.

**Group B.** The type used distinguishes the 13 group B categories. Two types have a $^{33\alpha\beta\gamma}$random format, and 11 a $^{33\alpha\beta\gamma}$linear$_Z$ format, where $Y$ is SRF set or 9 or 12, and $Z$ a set from number 2−12 in Table 2. Each group B category contains 10 partitions. Networks were trained for 4000 Epochs, with the network's configuration save every Epoch.

**Group C.** All five categories use the $^{33\alpha\beta\gamma}$linear$_{map_eneg}$ type and the 44 3 1 architecture and are distinguished by the variant sets used to monitor the trained network. Each different kind of variant set is distinguished by a subscript following its monitoring set name. The subscript identifies what substituent data the type in the set *contains*. The test$_{\beta\gamma}$ and the evaluation$_{\alpha\gamma}$ variants were formed by removing the data on $\alpha$ and $\beta$ substituents, respectively. For continuity, sets with unmodified types also have a subscript. Table 3 lists the monitoring set names and the substituent data they contain. In group C, networks were trained with a partition's training$_{\alpha\beta\gamma}$ set for 40 000 Epochs [Training was extended when it became clear (group B) that significant performance variations could occur *after* 4000 Epochs.], saving the configuration every 10 Epochs. Five such partitions were used in group C.

**3. Meaningful Performance Differences.** Comparing performance values requires identifying the size of significant differences, particularly when the training data contain many errors. Two characteristics of training a neural network are (a) long training times and (b) dependence on a number of interrelated parameters. To derive the size of significant performance differences statistically would require substantially more experiments per category than at present and narrow the scope of the research. Instead, therefore, the size of significant differences is measured heuristically, on the different performances of ***category pairs*** with structured types. Within a pair, the only difference in experimental conditions is that their types have different structured formats. As network architectures are fully connected, these should have identical performances. Nine structured category pairs can be formed from the group A categories, with a maximum difference of 4% in the <20 ppm band. This maximum occurs in the pair of $^{33\alpha\beta}$linear$_{map_eneg}$ and $^{33\alpha\beta}$branch$_{map_eneg}$

**Table 2.** Description and Subscript Identifiers of the Different Symbol Replacement Files (SRF) Used To Replace the Element Symbols in the MAGS Substructure Codes

| property contained within the SRF file | SRF file no. | SRF file subscript identifier | description of the result of using the SRF file to replace the MAGS substructure code symbols |
|---|---|---|---|
| global zero replacement | 0 | zero | all element symbols all replaced by a value of 0 |
| mapped electronegativity | 1 | map_energ | element symbol replaced by its electronegativity values scaled to a [0−0.9] range |
| inverse mapped electronegativity | 2 | inv_eneg | element symbol replaced by its electronegativity values scaled to a [1−0.1] range |
| full electronegativity | 3 | ful_eneg | element symbol replaced by its actual electronegativity values (a [0−4.2] range) |
| mapped atomic number | 4 | map_atno | element symbol replaced by its atomic number scaled to a [0−0.9] range |
| inverse mapped atomic number | 5 | inv_atno | element symbol replaced by its atomic number scaled to a [1−0.1] range |
| full atomic number | 6 | ful_atno | element symbol replaced by its actual atomic number value (a [1−103] range) |
| global point 1 replacement | 7 | all_0.1 | all element symbols replaced by a value of 0.1 |
| global point 3 replacement | 8 | all_0.3 | all element symbols replaced by a value of 0.3 |
| global point 5 replacement | 9 | all_0.5 | all element symbols replaced by a value of 0.5 |
| global point 7 replacement | 10 | all_0.7 | all element symbols replaced by a value of 0.7 |
| global point 9 replacement | 11 | all_0.9 | all element symbols replaced by a value of 0.9 |
| global full 5 replacement | 12 | all_5.0 | all element symbols replaced by a value of 5.0 |

**Table 3.** Names of the Different Kinds of Variant Monitoring Sets Used in Group C and the Substituent Region Data They Contain

| monitoring set name | | substituent data that the monitoring set *contains* | | | |
|---|---|---|---|---|---|
| general name given to this type of set | individual names | ring vector | α substituent region | β substituent region | γ substituent region |
| monitoring$_{\beta\gamma}$ sets | test$_{\beta\gamma}$ & evaluation$_{\beta\gamma}$ sets | yes | no | yes | yes |
| monitoring$_{\alpha\gamma}$ sets | test$_{\alpha\gamma}$ & evaluation$_{\alpha\gamma}$ sets | yes | yes | no | yes |
| monitoring$_{\alpha\beta}$ sets | test$_{\alpha\beta}$ & evaluation$_{\alpha\beta}$ sets | yes | yes | yes | no |
| monitoring$_{\alpha}$ sets | test$_{\alpha}$ & evaluation$_{\alpha}$ sets | yes | yes | no | no |
| monitoring$_{\alpha\beta\gamma}$ sets | test$_{\alpha\beta\gamma}$ & evaluation$_{\alpha\beta\gamma}$ sets | yes | yes | yes | yes |

types where a 17 3 1 architecture was used (Table 4, rows 6 and 7).

## RESULTS

The results for groups A, B, and C are given in Tables 4, 5, and 6, respectively, with an "*a*" in column 1 of Tables 5 and 6 signifying previous results.[15] The original purpose of group C was to show the eventual relative performance of the variants as is given in Table 6. Early in the training cycle, however, these performances fluctuate considerably giving valuable information on what occurs during training. Figures 6, 7, and 8 have therefore been included. The figures show performances in the <20 ppm band for the variant and unmodified monitoring sets at an Epoch interval: every 100 epochs in Figure 6 and every 10 Epochs in Figures 7 and 8.

## DISCUSSION AND CONCLUSIONS

The groups are discussed individually first and then collectively.

**Group A.** The results suggest that the scope of the positional equivalence property is generally important. The reasoning follows. The $^{X}$linear$_{zero}$ type control types, where X is a Figure 2 template, contain no substituent data. The $^{X}$random$_{map_eneg}$ control type(s) contains the same amount of substituent data as its structured counterparts. Networks trained with structured types perform significantly better than when trained with control types. Performance differences between pairs of control types ({row$_{4a-1}$,row$_{4a}$} in Table 4) are close to the level of significance. A further consistency across the templates is the relative performance of architecture with (multilayer) and without (Perceptron) hidden layers. In every case, the Perceptron performed significantly better than the controls but worse than the corresponding multilayer architecture. As Perceptrons can only represent linear relationships,[25] this suggests the topology-shift relationship has linear *and* nonlinear components. No significant differences in performance occur between multilayer architectures with different numbers and sizes of layers. With respect to the relative contribution of substituents, as mentioned above, the group A results can only give confirmatory evidence, suggesting the following dependencies in the prediction method evolved by the network.

**1.** From the performance order template$_{33\alpha\beta\gamma\delta}$ > template$_{33\alpha\beta\gamma}$ > template$_{33\alpha\beta}$ the suggestion is that better performance is obtained by increasing the environment size.

**2.** Less of an increase is obtained by including $\delta$ substituent data than is obtained by including $\gamma$ substituent data. This suggests that distance from the focus is a factor affecting the contribution of substituents to the predictions.

**3.** Differences in the performance of Perceptron and multilayer architectures also follow the order template$_{33\alpha\beta\gamma\delta}$ > template$_{33\alpha\beta\gamma}$ > template$_{33\alpha\beta}$. This suggests that a progressive increase in environment causes a progressive increase in the topology-shift relationship's nonlinear component. One explanation for this is that a fully connected architecture may somehow be accounting for the effects of through-space interactions. As the $\gamma$ and then the $\delta$ substituents are included, the number of such interactions detectable by the network also increases.

**Group B.** These results indicate that the class of phosphorus compounds studied in this work has an optimum replacement value for each element. Thus for this (3c3v) class, an optimum SRF *set* exists, inferred from the following. Networks trained with *any* structured type learned (subgroups 1−3 in Table 5) performing better than the controls (subgroup 4). Using scaled electronegativity replacement values (subgroup 1) is much better than using scaled atomic number or global values. Any differences *within* subgroups 1−3 are attributed to the effects of the replacement magnitudes. Evidence for this follows.

**1.** The lowest performing type containing a scaling of atomic numbers (subgroup 2) is the $^{33\alpha\beta\gamma}$linear$_{map_atno}$ type. Due to the element distributions, and in particular considering

**Table 4.** Group A Results Networks Trained with Four Types of Compound Vector Derived from Template$_{33\alpha\beta}$ and Four Types of Compound Vectors Derived from Template$_{33\alpha\beta\gamma\delta}$

| architecture of neural network used | type of compd vector used to train networks | performance values at the Epoch where the <20 ppm band reaches a max.[b] %age of compd vectors inside the tolerance band <20 ppm | <40 ppm | %age of compd vectors outside the tolerance band >80 ppm | >100 ppm | Epoch no. where the <20 ppm band reaches a max. value | max. <20 ppm value obsd in any individual monitoring set |
|---|---|---|---|---|---|---|---|
| 17 3 2 1 | $^{33\alpha\beta}$linear$_{map-eneg}$ | 33 (10) | 60 (24) | 13 (52) | 8 (42) | 2420 | 42.5 |
| 17 3 2 1 | $^{33\alpha\beta}$branch$_{map-eneg}$ | 36 (11) | 64 (25) | 12 (52) | 6 (42) | 3820 | 45.0 |
| 17 3 2 1 | $^{33\alpha\beta}$random$_{map-eneg}$ | 17 (11) | 30 (24) | 50 (47) | 36 (40) | 3740 | 21.5 |
| 17 3 2 1 | $^{33\alpha\beta}$linear$_{zero}$ | 12 (10) | 26 (25) | 47 (49) | 36 (40) | 2290 | 17.0 |
| 17 3 1 | $^{33\alpha\beta}$linear$_{map-eneg}$ | 33 (10) | 58 (23) | 13 (52) | 8 (41) | 3310 | 44.0 |
| 17 3 1 | $^{33\alpha\beta}$branch$_{map-eneg}$ | 37 (11) | 61 (25) | 13 (51) | 7 (41) | 840 | 42.0 |
| 17 3 1 | $^{33\alpha\beta}$random$_{map-eneg}$ | 18 (11) | 35 (24) | 39 (50) | 28 (40) | 3260 | 26.5 |
| 17 3 1 | $^{33\alpha\beta}$linear$_{zero}$ | 12 (11) | 26 (25) | 46 (49) | 37 (40) | 410 | 17.0 |
| 17 1 | $^{33\alpha\beta}$linear$_{map-eneg}$ | 29 (13) | 53 (28) | 19 (44) | 12 (31) | 70 | 36.5 |
| 17 1 | $^{33\alpha\beta}$branch$_{map-eneg}$ | 31 (13) | 56 (29) | 17 (43) | 10 (32) | 150 | 37.5 |
| 17 1 | $^{33\alpha\beta}$random$_{map-eneg}$ | 15 (11) | 29 (22) | 46 (49) | 36 (40) | 3220 | 17.0 |
| 17 1 | $^{33\alpha\beta}$linear$_{zero}$ | 12 (13) | 26 (27) | 46 (50) | 36 (41) | 50 | 15.0 |
| 44 10 1[a] | $^{33\alpha\beta\gamma}$linear$_{map-eneg}$ | 48 (10) | 73 (21) | 8 (49) | 4 (35) | 3910 | 56.0 |
| 44 10 1[a] | $^{33\alpha\beta\gamma}$branch$_{map-eneg}$ | 50 (10) | 74 (21) | 8 (52) | 4 (38) | 3980 | 56.8 |
| 44 10 1[a] | $^{33\alpha\beta\gamma}$random$_{map-eneg}$ | 16 (9) | 30 (20) | 46 (52) | 36 (35) | 2440 | 24.0 |
| 44 10 1[a] | $^{33\alpha\beta\gamma}$linear$_{zero}$ | 11 (9) | 24 (18) | 43 (50) | 33 (31) | 3990 | 14.0 |
| 44 3 1[a] | $^{33\alpha\beta\gamma}$linear$_{map-eneg}$ | 48 (10) | 73 (21) | 8 (50) | 4 (35) | 3840 | 54.8 |
| 44 3 1[a] | $^{33\alpha\beta\gamma}$branch$_{map-eneg}$ | 47 (10) | 72 (21) | 8 (52) | 5 (37) | 3910 | 54.4 |
| 44 3 1[a] | $^{33\alpha\beta\gamma}$random$_{map-eneg}$ | 14 (9) | 28 (19) | 47 (52) | 34 (37) | 3550 | 20.4 |
| 44 3 1[a] | $^{33\alpha\beta\gamma}$linear$_{zero}$ | 11 (9) | 24 (19) | 43 (51) | 31 (34) | 3960 | 14.4 |
| 44 1[a] | $^{33\alpha\beta\gamma}$linear$_{map-eneg}$ | 38 (21) | 62 (41) | 13 (29) | 8 (20) | 3600 | 45.2 |
| 44 1[a] | $^{33\alpha\beta\gamma}$branch$_{map-eneg}$ | 36 (20) | 61 (39) | 13 (32) | 8 (23) | 3030 | 42.8 |
| 44 1[a] | $^{33\alpha\beta\gamma}$random$_{map-eneg}$ | 13 (14) | 27 (27) | 45 (50) | 33 (38) | 3350 | 16.4 |
| 44 1[a] | $^{33\alpha\beta\gamma}$linear$_{zero}$ | 12 (10) | 24 (22) | 43 (46) | 34 (31) | 3620 | 14.4 |
| 125 5 1 | $^{33\alpha\beta\gamma\delta}$linear$_{map-eneg}$ | 53 (8) | 76 (18) | 6 (52) | 3 (34) | 3610 | 56.7 |
| 125 5 1 | $^{33\alpha\beta\gamma\delta}$branch$_{map-eneg}$ | 52 (7) | 77 (16) | 6 (51) | 3 (31) | 3930 | 58.0 |
| 125 5 1 | $^{33\alpha\beta\gamma\delta}$random$_{map-eneg}$ | 13 (7) | 24 (16) | 53 (51) | 43 (33) | 2240 | 20.7 |
| 125 5 1 | $^{33\alpha\beta\gamma}{}_{\delta}$linear$_{zero}$ | 10 (7) | 24 (17) | 42 (50) | 32 (34) | 1150 | 13.0 |
| 125 3 1 | $^{33\alpha\beta\gamma\delta}$linear$_{map-eneg}$ | 52 (8) | 77 (18) | 6 (51) | 3 (34) | 3860 | 58.0 |
| 125 3 1 | $^{33\alpha\beta\gamma\delta}$branch$_{map-eneg}$ | 51 (7) | 76 (16) | 7 (49) | 3 (31) | 3850 | 57.3 |
| 125 3 1 | $^{33\alpha\beta\gamma\delta}$random$_{map-eneg}$ | 13 (8) | 24 (18) | 50 (51) | 37 (34) | 3150 | 20.7 |
| 125 3 1 | $^{33\alpha\beta\gamma\delta}$linear$_{zero}$ | 10 (8) | 24 (16) | 42 (50) | 32 (34) | 2720 | 14.3 |
| 125 1 | $^{33\alpha\beta\gamma\delta}$linear$_{map-eneg}$ | 39 (25) | 66 (47) | 11 (27) | 7 (18) | 70 | 43.7 |
| 125 1 | $^{33\alpha\beta\gamma\delta}$branch$_{map-eneg}$ | 39 (25) | 64 (46) | 12 (26) | 7 (17) | 90 | 44.3 |
| 125 1 | $^{33\alpha\beta\gamma\delta}$random$_{map-eneg}$ | 13 (7) | 23 (17) | 47 (49) | 34 (36) | 60 | 21.0 |
| 125 1 | $^{33\alpha\beta\gamma\delta}$linear$_{zero}$ | 10 (8) | 23 (22) | 42 (45) | 31 (34) | 3900 | 13.0 |

[a] Indicates results from categories previously presented in ref 15. [b] Performance values at one Epoch are shown in parentheses.

the replacement value for hydrogen [The second most common element in MAGS codes obtained from the 3c3v class.], this type has very small input magnitudes. Its "inverse", the $^{33\alpha\beta\gamma}$linear$_{inv-atno}$ type, still has small input magnitudes though these are closer to one than zero. The $^{33\alpha\beta\gamma}$linear$_{inv-atno}$ and $^{33\alpha\beta\gamma}$linear$_{ful-atno}$ types, while having scales differing by two orders of magnitude, have very similar performances.

**2.** Types with global replacement values (subgroup 3, Table 5) show a high degree of correlation between the replacement magnitudes and the type's performance.

**3.** If performance is viewed over the training period the rate of increase in a category's performance and the magnitude of its inputs are highly correlated. One further conclusions made from these results is that with the training parameters currently in use considerable performance variations probably occur *after* 4000 Epochs.

**Group C.** The results demonstrate that, in the network's method of shift prediction, substituent regions are ordered $\alpha > \beta > \gamma$ in terms of their importance. This is shown by the relative performances of the variant and unmodified sets in Table 6 and in Figure 6 and taking into account the element percentage profile of the template$_{33\alpha\beta\gamma}$ derived unique set. Considering the profile is necessary because the same effect would be observed *if* the positions in the $\beta$ and $\gamma$ regions were very sparsely occupied *and* substituents from all regions had equivalent importance. This is not the case, with the ratio of the total number of positions occupied in the $\alpha$, $\beta$, and $\gamma$ regions being approximately 1:2:3. Previous work showed a high correlation between the nature of the $\alpha$ substituents of foci and distinct subranges of their $^{31}$P shifts. These subranges were often 100 ppm distant. Group C categories show that without any $\alpha$ substituent data the network cannot predict shifts to within $\pm$100 ppm. Here a dependency of the network evolved method can be directly related to a parameter in the problem domain. In many problems where good solutions have been evolved by networks, the means are unclear and cannot be related to any problem parameters.[24] This is unfortunate, because when both network and human evolved solutions use equivalent parameters, ways of improving the network's solution can be immediately suggested. During training, network states are saved at intervals and used later to assess performance. If learning is based on general characteristics, then training and monitoring sets should have roughly equivalent performances. There may occur a point where no more learning

**Table 5.** Results from Group B Where Different SRF Files Were Used To Replace the Substructure Code Symbols[b]

| type of compd vector in the category | subgroup classified by the relationship *between* the values in the SRF file used in the category | performance values at the Epoch where the <20 ppm band attains a max.[c] %age of compd vectors inside the tolerance band value <20 ppm | <40 ppm | %age of compd vectors outside the tolerance band value >80 ppm | >100 ppm | Epoch no. where the <20 ppm band reaches a max. value | max. <20 ppm value obsd in any individual monitoring set |
|---|---|---|---|---|---|---|---|
| $^{33}\alpha\beta\gamma$linear$_{map-eneg}$[a] | 1 | 48 (10) | 73 (21) | 8 (50) | 4 (35) | 3840 | 54.8 |
| $^{33}\alpha\beta\gamma$linear$_{inv-eneg}$ | 1 | 49 (10) | 74 (20) | 8 (50) | 5 (36) | 3890 | 58.4 |
| $^{33}\alpha\beta\gamma$linear$_{ful-eneg}$ | 1 | 49 (10) | 75 (21) | 7 (48) | 3 (31) | 3920 | 58.8 |
| $^{33}\alpha\beta\gamma$linear$_{map-atno}$ | 2 | 24 (9) | 46 (21) | 22 (51) | 14 (37) | 3990 | 32.8 |
| $^{33}\alpha\beta\gamma$linear$_{inv-atno}$ | 2 | 37 (9) | 60 (19) | 17 (51) | 11 (35) | 3970 | 43.2 |
| $^{33}\alpha\beta\gamma$linear$_{ful-atno}$ | 2 | 36 (12) | 60 (26) | 15 (44) | 10 (28) | 3430 | 45.2 |
| $^{33}\alpha\beta\gamma$linear$_{all-0.1}$ | 3 | 32 (9) | 52 (20) | 24 (52) | 17 (37) | 3840 | 42.0 |
| $^{33}\alpha\beta\gamma$linear$_{all-0.3}$ | 3 | 34 (8) | 55 (20) | 22 (51) | 16 (36) | 3440 | 39.6 |
| $^{33}\alpha\beta\gamma$linear$_{all-0.5}$ | 3 | 34 (9) | 55 (20) | 22 (51) | 16 (36) | 3970 | 40.0 |
| $^{33}\alpha\beta\gamma$linear$_{all-0.7}$ | 3 | 34 (9) | 55 (21) | 23 (50) | 17 (35) | 3440 | 44.0 |
| $^{33}\alpha\beta\gamma$linear$_{all-0.9}$ | 3 | 35 (9) | 55 (20) | 23 (50) | 17 (36) | 3860 | 40.0 |
| $^{33}\alpha\beta\gamma$linear$_{all-5.0}$ | 3 | 38 (15) | 57 (29) | 22 (40) | 16 (27) | 2040 | 46.0 |
| $^{33}\alpha\beta\gamma$linear$_{zero}$[a] | 4 | 11 (9) | 24 (19) | 43 (51) | 31 (34) | 3960 | 14.4 |
| $^{33}\alpha\beta\gamma$random$_{all-0.5}$ | 4 | 16 (9) | 29 (20) | 43 (52) | 32 (37) | 1630 | 27.6 |
| $^{33}\alpha\beta\gamma$random$_{all-5.0}$ | 4 | 15 (9) | 29 (20) | 46 (50) | 37 (36) | 2880 | 20.8 |
| $^{33}\alpha\beta\gamma$random$_{map-eneg}$[a] | 4 | 18 (10) | 35 (24) | 39 (50) | 28 (40) | 3260 | 26.5 |

[a] Indicates where the results have been presented previously in ref 15. [b] All compound vectors are derived from template$_{33\alpha\beta\gamma}$. In all categories a 44 × 33 × 1 network architecture was used. [c] Performance values at one Epoch are shown in parentheses.

**Table 6.** Results from Group C Where Trained Networks Are Monitored Using Additional Monitoring Set Variants[a]

Section 1. Average Performance Values at the Epoch Where the <20 ppm Band Attains a Maximum Value

| monitoring set type | performance values at the Epoch where the <20 ppm band attains a max. %age of compd vectors inside the tolerance band value <20 ppm | <40 ppm | %age of compd vectors outside the tolerance band value >80 ppm | >100 ppm | Epoch no. where the <20 ppm band reaches a max. value | max. <20 ppm value obsd in any individual monitoring set |
|---|---|---|---|---|---|---|
| monitoring$_{\beta\gamma}$ sets | 18 (11) | 31 (25) | 45 (48) | 33 (32) | 20 | 22.8 |
| monitoring$_{\alpha\gamma}$ sets | 31 (9) | 57 (19) | 16 (47) | 9 (31) | 530 | 34.8 |
| monitoring$_{\alpha\beta}$ sets | 34 (10) | 59 (24) | 14 (47) | 9 (33) | 150 | 41.2 |
| monitoring$_{\alpha}$ sets | 25 (9) | 45 (19) | 20 (49) | 12 (33) | 250 | 30.8 |
| monitoring$_{\alpha\beta\gamma}$ sets | 49 (10) | 74 (24) | 7 (45) | 4 (30) | 5940 | 57.6 |

Section 2. Average Performance Values at 40 000 Epochs

| monitoring set type | performance values at 40 000 Epochs[c] %age of compd vectors inside the tolerance band value <20 ppm | <40 ppm | %age of compd vectors outside the tolerance band value >80 ppm | >100 ppm |
|---|---|---|---|---|
| monitoring$_{\beta\gamma}$ sets | 0.0 | 0.0 | 99.9 | 99.8 |
| monitoring$_{\alpha\gamma}$ sets | 12 | 24.5 | 55.5 | 45.3 |
| monitoring$_{\alpha\beta}$ sets | 34.4 | 61.0 | 12.2 | 7.2 |
| monitoring$_{\alpha}$ sets | 9.6 | 19.7 | 58.8 | 47.2 |
| monitoring$_{\alpha\beta\gamma}$ sets | 51.0 | 74.9 | 6.9 | 4.1 |

[a] All compound vectors are of the type $^{33}\alpha\beta\gamma$linear$_{map-eneg}$. [b] Values observed at 10 Epochs are shown in parentheses. [c] Values are shown to first decimal place in this section.

based on general characteristics is possible. If this point is reached, the network's error driven response will now consider characteristics specific to the training set. Normally, such overtraining of a network is detected by an attenuated performance in the monitoring set(s), while the performance of the training set continues to improve. Figure 6 shows no decline in the performance of the unmodified (monitoring$_{\alpha\beta\gamma}$) sets, suggesting that any learning is based on general characteristics. There are, however, many training intervals where the *variant* set performances decline. An explanation follows.

Where there is no overtraining, the unmodified sets will never show a performance decline because the network responds to the *overall* training error. This response considers the aggregate of the contributions from each region to predicting shifts. As the unmodified set(s) effectively mirrors the training set, consequently performance will either increase or be static. Any decreases in performance of the variant sets will then be due to intervals where (a) the region(s) contained in the variant has an aggregate negative contribution and (b) the omitted region has a positive contribution. When a region contributes negatively, the most probable cause is that, in this training interval, another region *dominates*. Here some adjustments of the dominant region weights give large increases in performance, i.e., the region is the dominant contributor to the error gradient. In such intervals adjustments may be made at the expense of the nondominant regions, whose weight values *on their own* may now have a negative contribution. Once a region's domination diminishes, other region(s) can take precedence. These

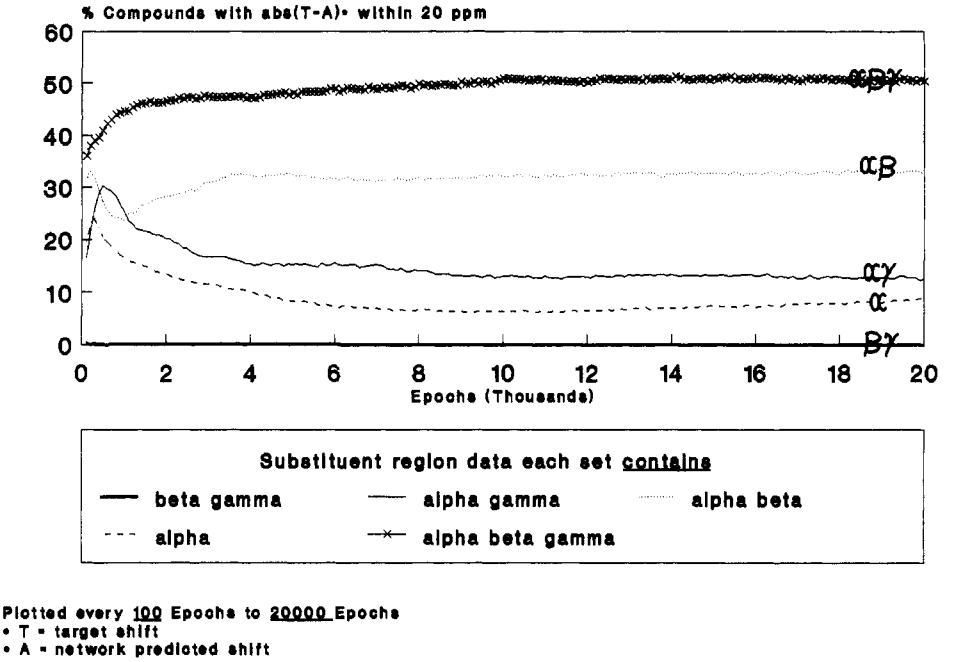


**Figure 6.** Group C 20 000 Epoch performance. Category averages <20 ppm tolerance band for the five types of monitoring set.

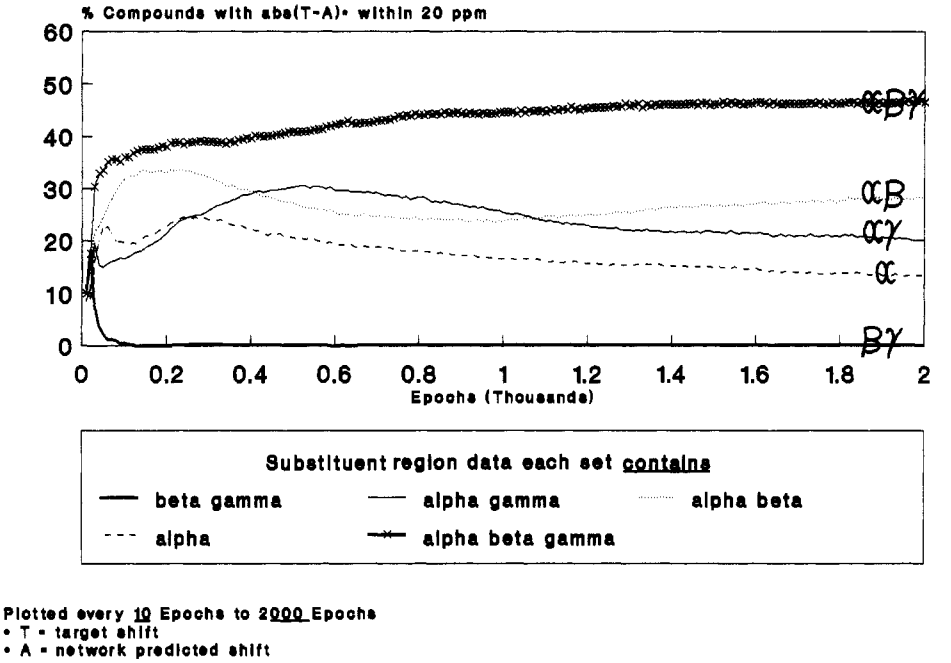


**Figure 7.** Group C 2000 Epoch performance. Category averages <20 ppm tolerance band for the five types of monitoring set.

oscillations, with respect to the partial contributions of regions, are most likely to occur early in training and should eventually dampen down. Despite the reduced detail due to extending the training time, such oscillations are still detectable, as is apparent in Figures 6–8, and can give valuable information. The pattern of these fluctuations suggests the following.

**1.** At or about 20 Epochs the network starts to evolve its prediction method, shown by the rapidly declining performance of the monitoring$_{\beta\gamma}$ sets. Every other set contains α substituent data and shows an increase in performance in this interval.

**2.** The order of the *first* performance decreases in the variants (Figures 7 and 8) is monitoring$_{\beta\gamma}$, monitoring$_{\alpha\gamma}$, monitoring$_{\alpha}$, and monitoring$_{\alpha\beta}$. As outlined above, such a decrease is most probably caused by intervals where the *excluded* region(s) are the dominant contribution to the error gradient. If this order is expressed in terms of the omitted regions it is α, $\beta$, $\beta\gamma$, $\gamma$. The order of importance of a region's contribution derived from the final performances, and that derived from its (initial) contribution to the error gradient during training, is therefore consistent.

**General Conclusions.** Apart from indicating a general importance for the positional equivalence property, the results implicate three other factors as significant in the prediction method evolved by the network. These are (a) a correlation between the proximity of substituents to the focus and their importance in predicting shifts, (b) a correlation between the size of the environment the network is given and the predictive accuracy, and (c) a correlation between the values that replace the element symbols and the predictive accuracy. These three factors have direct parallels in $^{13}C$ additivity rule

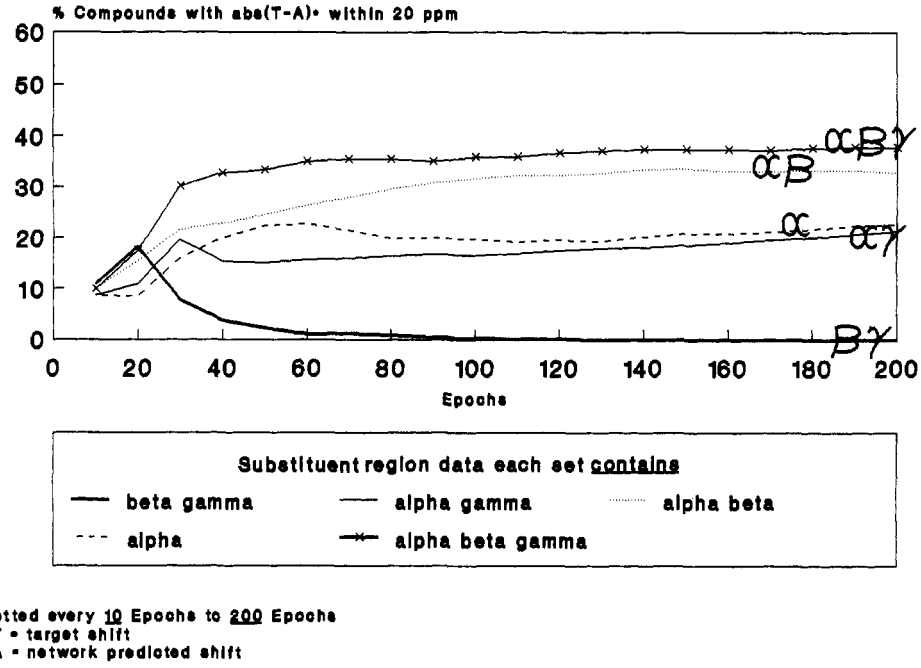


**Figure 8.** Group C 200 Epoch performance. Category averages <20 ppm tolerance band for the five types of monitoring set.

methods which show (a) a dependence on $\alpha-\delta$ substituent effects,[13] (b) a dependence of predictive accuracy on environment size,[4] and (c) the existence of an optimum set-(s) of numeric replacement values for topological features.[26] The parallels between the two methods suggests two immediate ways of increasing the network's predictive accuracy. In the first, divide the application domain into many topologically distinct realms. So far, data from the entire 3c3v class have been used. In the $^{13}C$ additivity rules this is equivalent to having a single set of parameters for *any* $sp^3$ focus. In the second, use the error driven nature of network learning to derive the optimum set of values *automatically* for each topological subclass. Of the replacements (SRF files) used in this work, electronegativity scalings give the highest accuracy. It is most improbable that this is a global maximum for even the entire 3c3v class. Apart from any improvement that may be caused by the refinements above, others will *assure* better performance, as the method in its current form suffers from several imperfections. A list of these follows.

(a) The level of data errors: Aside from errors in the polarity and magnitude of shifts for the reasons given earlier, there is a high probability that many topologies are also incorrect, as these were manually decoded from a novel line formula.[21] These nonunique representations had an original audience of $^{31}P$ experts and require considerable dexterity and chemical knowledge to be unambiguously decoded. In such formulas, topology is represented by a sequential list of functional groups, starting from some arbitrarily chosen point. Connections are also given arbitrarily, and functional groups from outside the allowed set are often included. These attributes made any means of automatic translation to a more algorithmically pliable form difficult. Manual translation is bound to have introduced errors.

(b) Optimizing the network related parameters: Neural network learning is influenced by many interrelated parameters such as the architecture, training algorithm, and transfer functions used. For the structure-shift mapping problem, it is highly unlikely that any of the network parameters above are optimal. There are many more sophisticated training algorithms than the one used in this work,[27,28] with backpropagation chosen for its simplicity. It is anticipated that methods such as genetic algorithms[29] would be employed in the future search for network related optima.

(c) Using a better vertex ordering algorithm: These initial investigations have used a global vertex ordering given by the Morgan algorithm.[30] In some topological types the order given by this algorithm is known to be erroneous,[31] and thus the order in a *fixed percentage* of structures will be incorrect. Many more differentiating global ordering algorithms exist, such as those based on adjacency matrix eigenvalues,[31] the molar graph center,[32] and vertex potentials[33] as well as more refined algorithms based on extended connectivity.[34,35] While a future investigation of these is planned, it is highly probable that using a local ordering algorithm will be more optimal. This work indicates the network's prediction method has substituent-shift dependencies similar to those in $^{13}C$ NMR, where the relationships are better modeled by using a local ordering as in the DARC[3] and HOSE systems.[1]

Three characteristics of neural network based solutions are (i) a notorious difficulty in scaling problems up from the prototype; (ii) good fault tolerance with a gradual decline in performance as errors increase; and (iii) lengthy training times. The first characteristic is one reason which dictates that the initial research should use all the data rather than a highly refined (error free) subset. The other two characteristics, together with the research's focus on viability of the method, are the reasons why the Morgan and backpropagation algorithms are retained. Although performance will be attenuated, the previous experiments using these algorithms provide a convenient baseline for performance comparisons, especially given that any errors they introduce are constant.

## REFERENCES AND NOTES

(1) Bremser, W. HOSE-A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.
(2) Munk, M. E.; Lind, R. J.; Clay, M. E. Computer Mediated Reduction of Spectral Properties to Molecular Structures: General Design and Structural Building Blocks. *Anal. Chim. Acta* **1987**, *184*, 1−19.

(3) Dubois, J. E.; Bonnet, J. C. The DARC Pluridata System: The $^{13}$C-N.M.R. Data Bank. *Anal. Chim. Acta* **1979**, *112*, 245−252.
(4) Brown, D. W. A short set of $^{13}$C-NMR Correlation Tables. *J. Chem. Ed.* **1985**, *62*, 209−212.
(5) Buchanan, B. G.; Sutherland, G. L.; Fiegenbaum, E. U. In *Machine Intelligence 4*; Meltzer, B., Michie, D., Eds.; Edinburgh University Press: Edinburgh, 1969; pp 209−254.
(6) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Mag. Res. Chem.* **1985**, *23*, 1056−1071.
(7) Robein, W. Computer-Assisted Structure Elucidation of Organic Compounds III*: Automatic Fragment Generation From $^{13}$C-NMR Spectra. *Mikrochim Acta [Wien]* **1986**, *II*, 271−279.
(8) Lah, L.; Tusar, M.; Zupan, J. Simulation of carbon-13 spectra. *Tetrahedron Comput. Methodol.* **1989**, *2*, 5−15.
(9) Furst, A.; Pretsch, E. A Computer Program for the Prediction of $^{13}$C-NMR Chemical Shifts of Organic Compounds. *Anal. Chim. Acta* **1990**, *229*, 17−25.
(10) Pretsch, E.; Furst, A.; Badertscher, M.; Burgin, R. C13 Shift: A Computer Program for the Prediction of 13C NMR Spectra Based on an Open Set of Additivity Rules. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 291−295.
(11) Chen, L.; Robien, W. OPSI: A Universal Method for Prediction of 13C-NMR Spectra Based on Optimised Additivity Models. *Anal. Chem.* **1993**, *65*, 2282−2287.
(12) Grant, D. M.; Paul, E. G. Carbon-13 Magnetic Resonance II. Chemical Shift Data for the Alkanes. *J. Am. Chem. Soc.* **1964**, *86*, 2984−2990.
(13) Lindeman, L. P.; Adams, J. Q. Carbon-13 Nuclear Magnetic Resonance Spectroscopy: Chemical Shifts for the Paraffins through $C_9$. *Anal. Chem.* **1971**, *43*, 1245−1252.
(14) *Computer Supported Spectroscopic Databases*; Zupan, J., Ed.; Ellis Horwood Ltd.: Chichester, UK, 1987.
(15) West, G. M. J. Predicting Phosphorus NMR Shifts Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 577−589.
(16) Doucet, J. P.; Panaye, A.; Feuilleaubois, E.; Ladd, P. Neural Networks and 13C NMR Shift Prediction. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 320−324.
(17) Kvasnicka, V. An Application of Neural Networks in Chemistry. Prediction of 13C NMR Shifts. *J. Meth. Chem.* **1991**, *6*, 63−76.
(18) Kvasnicka, V.; Skelenak, S.; Pospichal, J. Application of Recurrent Neural Networks in Chemistry. Prediction and Classification of C13 NMR Shifts in a Series of Monosubstituted Benzenes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 742−747.
(19) Anker, L. S.; Jurs, P. C. Prediction of Carbon 13 Nuclear Magnetic Resonance Chemical Shifts by Artificial Neural Networks. *Anal. Chem.* **1992**, *64*, 10, 217−219.
(20) IUPAC Physical Chemistry Division; Commission on Molecular Structure and Spectroscopy; Presentation of NMR data for Publication in Chemical Journals-B: Conventions Relating to Spectra from Nuclei Other than Proton. Recommendations 1975; *Pure Appl. Chem.* **1976**, *45*, 217−219.
(21) Tebby, J. C. *Handbook of Phosphorus NMR Data*; CRC Press: U.S.A., 1991.
(22) Kolmogorov, A. N. On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition. *Dokl. Akad. Nauk. USSR* **1957**, *114*, 953−956.
(23) Cybenko, G. Approximation by Superpositions of a Sigmoidal Function. *Math. Control, Signals Syst.* **1989**, *2*, 337−341.
(24) Denker, J.; Schwartz, D.; Wittner, B.; Solla, S.; Howard, R.; Jackel, L.; Hopfield, J. Large Automatic Learning, Rule Extraction and Generalisation. *Complex Systems* **1987**, *1*, 877−922.
(25) Minsky, M. L.; Papert, S. *Perceptrons*; The MIT Press: Cambridge, MA, 1969.
(26) Pretsch, E.; Clerc, J. T.; Seibl, J.; Simon, W. Tables of Spectral Data for Structural Elucidation of Organic Compounds, 2nd ed.; Springer-Valag: Berlin, 1989.
(27) Becker, S.; Le Cun, Y. Improving the Convergence of Back-Propagation Learning with Second Order Methods. In *Proceedings of the 1988 Connectionist Models Summer School*; Touretzky, D. S., Hinton, G. E., Sejnowski, T. J., Eds.; Morgan Kaufmann: San Mateo, 1990; pp 29−37.
(28) Fahlman, S. E. The Cascade-Correlation Learning Architecture. In *Advances in Neural Information Processing Systems II*; Touretzky, D. S., Ed.; Morgan Kaufmann: San Mateo, 1990; pp 524−532.
(29) Goldberg, D. E. *Genetic Algorithms in Search Optimization, and Machine Learning*; Addison-Wesley: New York, 1989.
(30) Morgan, H. L. Generation of Unique Machine Description for Chemical Structures, a Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.
(31) Randic, M. On Unique Numbering of Atoms and Uniques Codes for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 105−117.
(32) Jochum, C.; Gasteiger, J. Canonical Numbering and Constitutional Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *1*, 113−117.
(33) Golender, V. E.; Drboglav, V. V.; Rosenblit, A. B. Graph Potentials Method and Its Application to Chemical Information Processing. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 196−204.
(34) Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Heirarchically Ordered Extended Connectivities (HOC procedures). III Topological Chemical and Stereochemical Coding of Molecular Structure. *J. Comput. Chem.* **1985**, *6*, 562−569.
(35) Liu, X.; Balasubramanian, K.; Munk, M. E. Computational Techniques For Vertex Partitioning of Graphs. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 263−269.

CI930273W