# MULTIVARIATE REGRESSION WITH SUBSTITUENT SHIFT INCREMENTS. III. 2,5-DISUBSTITUTED FURANS ${ }^{+}$ 

Miroslav Holík ${ }^{a, *}$, Zdeněk Friedl ${ }^{b}$ and Štefan Marchalín ${ }^{c}$<br>${ }^{a}$ Department of Theoretical and Physical Chemistry, Masaryk University, CZ-611 37 Brno, Czech Republic; e-mail: holik@chemi.muni.cz<br>${ }^{b}$ Faculty of Chemistry, Technical University Brno, CZ-612 00 Brno, Czech Republic; e-mail: friedl@fch.vutbr.cz<br>${ }^{c}$ Department of Organic Chemistry, Slovak Technical University, SK-812 37 Bratislava, Slovak Republic; e-mail: smarchal@chtf.stuba.sk

Received June 18, 1999
Accepted September 13, 1999

Dedicated to Professor Otto Exner on the occasion of his 75th birthday.

Six series of 2-X-5-Y-substituted furans were used for multivariate regression of ${ }^{13} \mathrm{C} N M R$ chemical shifts with four substituent shift increments (SSI). Some of these series were measured in $\mathrm{CDCl}_{3}$ and $D M S O-d_{6}$ on different spectrometers to exclude possible sources of deviation from the general trends. Program ASSIGN was written for visual and numerical estimation of correct assignment of ${ }^{13} \mathrm{C}$ NMR spectra in the series of derivatives. Multivariate regression proved that the corresponding position in benzenes and furans are not affected by the substituents in the same way: in 2-X-5-Y-substituted furans, the "ortho" position requires correction for the different bond order, the "meta" ${ }^{13} \mathrm{C}$ chemical shift is probably affected by change in diamagnetic contribution of the Y group, and "para" position behaves like another "ortho" position in the molecule.
Key words: Substituent shift increments; ${ }^{13} \mathrm{C}$ NMR substituent chemical shifts; Assignment of ${ }^{13} \mathrm{C}$ NMR signals; Linear regression; Furans; NMR spectroscopy; Substituent effects.

For more than twenty years a single-variable regression analysis has been used to explore non-additivity of substituent shift increments (SSI) in calculating chemical shifts of $p$-disubstituted- $X, Y$-benzenes ${ }^{1,2}$. Such an approach is equivalent to the use of the SSI as substituent constants; this kind of linear regression is sometimes called proportionality relationship¹, Eq. (1), or "Lynch

[^0]plot" (ref. ${ }^{3}$ ) according to the author ${ }^{1}$ who introduced it to the ${ }^{13} \mathrm{C}$ NMR spectroscopy.
\[

$$
\begin{equation*}
\delta_{\mathrm{Cj}}=\mathrm{b}_{0}+\mathrm{b}_{\mathrm{j}} \mathrm{z}_{\mathrm{j}}(\mathrm{X})+\varepsilon \quad Y \text { fixed } \tag{1}
\end{equation*}
$$

\]

In this equation, $\delta_{\mathrm{Cj}}$ is the carbon chemical shift for a given position j and substituent $X$ in disubstituted benzene, $b_{0}$ is an estimate of the shift in the parent compound ( $X=H, Y \approx H$ ), $z_{j}$ is the SSI obtained ${ }^{4}$ from the ${ }^{13} \mathrm{C}$ NMR chemical shift of the $X$-monosubstituted benzene $(Y=H), b_{j}$ is the slope of the regression, and $\varepsilon$ is the residual. Ideally, $b_{j}$ should not differ substantially from one: if it does, an estimated parameter is biased and another explanatory variable has to be added to the regression equation. It was shown that in p-disubstituted benzenes with the fixed $Y$ substituent possessing a group with diamagnetic anisotropy effect, a variable expressing variation of the $Y$ group conformation caused by substituent $X$ must be added ${ }^{5}$ in order to properly explain chemical shifts for proximal $\mathrm{C}_{\mathrm{m}}$ carbon atoms. As an additional variable, $\sigma_{p}$ or $\sigma_{R}^{0}$ can be used ${ }^{5}$; another choice is SSI for para position, $z_{p}$, which correlates very well with both $\sigma$ values. Not only the diamagnetic anisotropy effect but also "the electron demand" (ref. ${ }^{6}$ ) of the substituent $Y$ can be accounted for by an additional variable in regression equation ${ }^{7}$, e.g., in p-disubstituted benzenes, an additional explanatory variable for $\delta_{C 1}$ is $z_{m}$ if $Y$ is an electron donor, or $z_{0}$ and $z_{p}$ if it is an acceptor ${ }^{8}$.

Therefore, it seems reasonable to model chemical shifts in disubstituted derivatives ${ }^{7}$ by regression equation of Eq. (2) possessing all SSI values, $z_{j}$, then to test the significance of the slope parameters, $b_{j}$, and finally to recalculate the regression equation only with those variables $z_{j}$ for which the slope $b_{j}$ is statistically significant.

$$
\begin{equation*}
\delta_{\mathrm{Cj}}=b_{0}+b_{1} z_{i}+b_{2} z_{0}+b_{3} z_{m}+b_{4} z_{p}+\varepsilon \quad Y \text { fixed } \tag{2}
\end{equation*}
$$

This approach proved to be usable also in the case of other than benzene aromatic. It is generally accepted, that ortho-SSI, i.e., $\mathrm{z}_{0}$, cannot be used for prediction of ${ }^{13} \mathrm{C}$ NMR chemical shifts in 2-X-naphthalenes ${ }^{9,10}$. Is is attributed to the fact that C2-C1 and C2-C3 $\pi$-bond orders in naphthalene are distinctly non-equivalent while the corresponding $\pi$-bond order in benzene is close to their mean value. However, the ${ }^{13} \mathrm{C}$ chemical shifts for both C 1 and C3 atoms can be predicted ${ }^{7}$ by proper combination of $z_{0}$ and $z_{p}$ values:
addition and subtraction of $z_{p}$ multiples simulates the increase and decrease in the bond order between C2-C1 and C2-C3 atoms, respectively.

In the early days of ${ }^{13} \mathrm{C}$ NMR spectroscopy, it was assumed that the positions 3, 4, and 5 in 2-substituted furans and the ortho, meta, and para positions of a monosubstituted benzenes are affected by the 2-substituent in similar ways ${ }^{11}$. Since our preliminary calculations with Eq. (2) in series of 2-X-5-Y-disubstituted furans gave slightly different results, we have decided to collect more such series to make our predictions statistically more significant. Now we present the results of our calculations on 2 -substituted furans 1 and five series of 2-X-5-Y-disubstituted furans 2-6.


| $\mathbf{1}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | H |
| $\mathbf{b}$ | MeO |
| $\mathbf{c}$ | MeS |
| $\mathbf{d}$ | Cl |
| $\mathbf{e}$ | Br |
| $\mathbf{f}$ | l |
| $\mathbf{g}$ | Me |
| $\mathbf{h}$ | $\mathrm{CH}_{2} \mathrm{OH}$ |
| $\mathbf{i}$ | $\mathrm{CH}=\mathrm{O}$ |
| $\mathbf{j}$ | $\mathrm{MeC=O}$ |
| $\mathbf{k}$ | COOH |
| $\mathbf{l}$ | CN |
| $\mathbf{m}$ | $\mathrm{NO}_{2}$ |



| $\mathbf{2}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ |
| $\mathbf{b}$ | PhS |
| $\mathbf{c}$ | Br |
| $\mathbf{d}$ | l |
| $\mathbf{e}$ | Ph |
| $\mathbf{f}$ | Me |
| $\mathbf{g}$ | H |
| $\mathbf{h}$ | $\mathrm{PhSO}_{2}$ |
| $\mathbf{i}$ | COOMe |
| $\mathbf{j}$ | CN |
| $\mathbf{k}$ | $\mathrm{NO}_{2}$ |


| $\mathbf{3}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ |
| $\mathbf{b}$ | PhO |
| $\mathbf{c}$ | PhS |
| $\mathbf{d}$ | Cl |
| $\mathbf{e}$ | Br |
| $\mathbf{f}$ | I |
| $\mathbf{g}$ | Ph |
| $\mathbf{h}$ | Me |
| $\mathbf{i}$ | $\mathbf{H}$ |
| $\mathbf{j}$ | $\mathrm{PhSO}_{2}$ |
| $\mathbf{k}$ | $\mathrm{COOMe}^{2}$ |
| $\mathbf{l}$ | $\mathrm{CN}^{2}$ |
| $\mathbf{m}$ | $\mathrm{NO}_{2}$ |
| $\mathbf{n}$ | $\mathrm{MeS}^{2}$ |
| $\mathbf{0}$ | MeSO |
| $\mathbf{p}$ | $\mathrm{MeC=O}$ |

## CALCULATIONS

All calculations were carried out with double precision on a PC with Pentium-S/ 150 MHz processor and 32 MB RAM. Standard subrutines were used for linear regression; for assignment and multivariate regression, program ASSIGN was written+ ${ }^{+}$.
++Listing of the ASSIGN program written in Matlab is available from the author (M. H.) on request. Matlab is a commercial product of The MathWorks Inc. Natick, U.S.A. supplied by Humusoft s.r.o., Prague.


| $\mathbf{4}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ |
| $\mathbf{b}$ | PhS |
| $\mathbf{c}$ | Cl |
| $\mathbf{d}$ | Br |
| $\mathbf{e}$ | I |
| $\mathbf{f}$ | Ph |
| $\mathbf{g}$ | Me |
| $\mathbf{h}$ | H |
| $\mathbf{i}$ | $\mathrm{PhSO}_{2}$ |
| $\mathbf{j}$ | COOMe |
| $\mathbf{k}$ | CN |
| $\mathbf{l}$ | $\mathrm{NO}_{2}$ |
| $\mathbf{m}$ | MeS |

m MeS


| $\mathbf{5}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ |
| $\mathbf{b}$ | PhO |
| $\mathbf{c}$ | PhS |
| $\mathbf{d}$ | Cl |
| $\mathbf{e}$ | Br |
| $\mathbf{f}$ | l |
| $\mathbf{g}$ | Ph |
| $\mathbf{h}$ | Me |
| $\mathbf{i}$ | H |
| $\mathbf{j}$ | COOMe |
| $\mathbf{k}$ | $\mathrm{CN}^{2}$ |
| $\mathbf{l}$ | $\mathrm{NO}_{2}$ |

## Program Assign



| $\mathbf{6}$ | X |
| :--- | :--- |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ |
| $\mathbf{b}$ | PhO |
| $\mathbf{c}$ | PhS |
| $\mathbf{d}$ | Cl |
| $\mathbf{e}$ | Br |
| $\mathbf{f}$ | $\mathbf{l}$ |
| $\mathbf{g}$ | Ph |
| $\mathbf{h}$ | Me |
| $\mathbf{i}$ | H |
| $\mathbf{j}$ | COOMe |
| $\mathbf{k}$ | CN |
| $\mathbf{l}$ | PhSO |
| $\mathbf{2}$ |  |

1. Matrix of data, $\mathbf{D}(\mathrm{m}, 4)$, and matrix of variables, $\mathbf{M}(m, 5)$, are introduced; $m$ is number of measurements - $X$ derivatives - in series. Matrix $\mathbf{D}$ is composed from individual column vectors of chemical shifts, e.g., $\delta_{\mathrm{C} 5}, \delta_{\mathrm{C} 4}$, $\delta_{C 3}$, and $\delta_{C 2}$ of furan ring. Matrix $\mathbf{M}$ contains four SSI vectors, i.e., $z_{i}, z_{0}, z_{m}$, and $z_{p}$ with a column of 128.5 s (i.e., $\delta_{0}$ ) added as the first column into the M matrix.
2. Projection matrix $\mathbf{H}=\mathbf{M}\left(\mathbf{M}^{\prime} \mathbf{M}\right)^{-1} \mathbf{M}^{\prime}$ is calculated ( $\mathbf{M}^{\prime}$ being a tranpose of $\mathbf{M}$ matrix) and used for projecting the data matrix $\mathbf{D}$ (experimental values) into the matrix $\mathbf{Q}$ (calculated values): $\mathbf{Q}=\mathbf{H} \mathbf{M}$.
3. Matrices $\mathbf{Q}$ and $\mathbf{D}$ are sorted in the ascending order and data of matrix D (rows are called sets) are plotted against selected column j (called block) of matrix $\mathbf{Q}$ as points together with the line obtained from plotting vector $\mathbf{q}$ against $\mathbf{q}$ (slope $=1$ ). Since the points for different carbon atoms (blocks) are ploted in different marks (e.g., $\bigcirc,{ }^{*}, x_{,}+$), it is easy to judge if the corresponding data lie properly distributed around the line with the slope one (Fig. 1). Coefficients of determination, CD, for each individual columns and also for all data taken as one vector (= total coefficient of determination), TCD, are cal culated.
4. If there is a point of another column which is closer to the line than the point of column corresponding to displayed block, then it can be exchanged (e.g., in Fig. 1, in set 4, +is closer to unity line of block 1 then O).

The improvement caused by this exchange can be checked by an increase in TCD and corresponding CDs. For calculation of new values of TCD and CDs, the program first recalculates the projection matrix $\mathbf{H}$ and then projects unsorted corrected $\mathbf{D}$ matrix to new matrix $\mathbf{Q}$. This procedure is then repeated for all the blocks (columns of $\mathbf{Q}$ ) present.
5. Finally, the matrix of parameters $b_{j}$, B, calculated according to Eq. (3) is presented and all slopes are tested for their statistical significance at the $95 \%$ confidence level dividing them by corresponding standard deviation and comparing the results with critical Student's $t$-value: $t=b_{j} / s_{b_{j}}>t_{(0.05, m-5)}$.

$$
\begin{equation*}
\delta_{\mathrm{Cj}}=\mathrm{b}_{0} \delta_{0}+\mathrm{b}_{1} \mathrm{z}_{\mathrm{i}}+\mathrm{b}_{2} \mathrm{z}_{0}+\mathrm{b}_{3} \mathrm{z}_{\mathrm{m}}+\mathrm{b}_{4} \mathrm{z}_{\mathrm{p}}+\varepsilon \quad \delta_{0}=128.5 \tag{3}
\end{equation*}
$$

6. Then the individual equations are recalculated according to reduced Eq. (3) with only significant variables, e.g.: $\delta_{\mathrm{C} 3}=b_{0} \delta_{0}+b_{3} z_{m}+b_{4} z_{p}+\varepsilon$, $\delta_{0}=128.5$.

## Nmr Chemical Shift Data

The origin of the data used for present calculations is evident from Table I and following comments. ${ }^{13} \mathrm{C}$ NMR chemical shifts of 2-substituted furans 1


Fig. 1
An example of re-assignment procedure with ASSIGN program. $O d_{1} ; * d_{2} ; \times d_{3} ;+d_{4}$
were taken from ref. ${ }^{11 a}$ without any change; they were obtained ${ }^{11 a}$ at 25.142 MHz on a Varian XL-100-15 spectrometer with an accuracy of $\pm 0.1$ ppm; recently published paper ${ }^{11 \mathrm{~b}}$ contains another derivative ( $\mathrm{X}=\mathrm{SPh}$ ), unfortunately, measured in another solvent and therefore, could not be added into the investigated series. 5-X-2-furaldehydes 2, 3-[(5-X-2-furyl)methy-lene]pentane-2,4-diones 4, and 2-[(5-X-2-furyl)methylene]malononitriles 6 were measured at $500 \mathrm{MHz}(125 \mathrm{MHz})$ on a Bruker Avance DRX-500 spectrometer ( $\mathrm{c}=0.1 \mathrm{~mol} \mathrm{dm}^{-3}$ ) and standardized on the central peak of solvent used ${ }^{12}$ : $\mathrm{CDCl}_{3} \delta 76.91$, DM SO-d ${ }_{6} \delta 39.56$. Alternatively, ${ }^{13} \mathrm{C}$ NMR spectra of 2, 4, and 2-acetyl-3-(5-X-2-furyl)prop-2-enenitriles 5 were mearured at 75.426 MHz on a Varian VXR 300 spectrometer under conditions given in ref. ${ }^{13}$. Chemical shifts of 3 -(5-X-2-furyl)-N,N-dimethylprop-2-enamides 3 were taken from Table III in ref. ${ }^{14}$, but in five spectra the signals were reassigned using program ASSIGN and 2D-heterocorrelated NMR experiments; new assignments are collected in Table II. After these changes, the data gave mostly better coefficients of determination: C5 (0.9299 $\rightarrow 0.9412$ ), C4 $(0.9339 \rightarrow 0.9216)$, C3 $(0.2792 \rightarrow 0.7006)$, and C2 $(0.8260 \rightarrow 0.9483)$.

## RESULTS AND DISCUSSION

For the ${ }^{13} \mathrm{C}$ chemical shifts on ipso carbon atom (C2 in 1, C5 in others) almost all measured sets gave a good regression with $z_{i}$ increments only when $z_{m}$ variable is also included. Only in the case of series 6 , the slope $b_{3}$ in the

Table I
Sets of chemical shifts data and measurement conditions used

| Experiment | Series | Solvent | Spectrometer | Reference |
| :---: | :---: | :---: | :---: | :---: |
| A | 1 | acetone $\mathrm{d}_{6}$ | $\checkmark 100$ | ref. ${ }^{11}$ |
| B | 2 | $\mathrm{CDCl}_{3}$ | B 500 | Table III |
| C | 2 | $\mathrm{CDCl}_{3}$ | $\checkmark 300$ | ref. ${ }^{15}$ |
| D | 3 | $\mathrm{CDCl}_{3}$ | J 60 | ref. ${ }^{14}$ |
| E | 4 | $\mathrm{CDCl}_{3}$ | $\checkmark 300$ | ref. ${ }^{15,16}$ |
| F | 4 | DMSO-d ${ }_{6}$ | $\vee 300$ | ref. ${ }^{15,16}$ |
| G | 4 | DMSO-d ${ }_{6}$ | B 500 | Table IV |
| H | 5 | $\mathrm{CDCl}_{3}$ | $\checkmark 300$ | ref. ${ }^{13}$ |
| 1 | 5 | DMSO-d ${ }_{6}$ | $\vee 300$ | ref. ${ }^{15}$ |
| J | 6 | $\mathrm{CDCl}_{3}$ | B 500 | Table V ref. ${ }^{17}$ |

regression equation is only significant at the $90 \%$ confidence level. For series 4 measured on a Bruker spectrometer, also the slope at $z_{p}$ was significant at $95 \%$ but not at $99 \%$ level. These two exceptions from the general trend point out to variations caused probably by sample selection.
In the p-disubstituted ( $\mathrm{X}, \mathrm{Y}$ ) benzene series ${ }^{8}$ significantly negative $\mathrm{b}_{3}$ slope was also observed, but only when $Y$ substituent was an electron donating

Table II
Corrected assignments ${ }^{\text {a }}$ of several derivatives of $\mathbf{3}$ (original data in ref. ${ }^{14}$ )

| $\mathbf{3}$ | X | C 5 | C 4 | C 3 | C 2 | $\mathrm{C}_{\mathrm{a}}$ | $\mathrm{C}_{\mathrm{b}}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{c}$ | PhS | $145.06^{\mathrm{a}}$ | $121.16^{\mathrm{b}}$ | $115.25^{\mathrm{b}}$ | $154.99^{\mathrm{a}}$ | 128.43 | 116.49 |
| $\mathbf{g}$ | Ph | 155.29 | 107.73 | $116.23^{\mathrm{c}}$ | 151.26 | 129.10 | $114.42^{\mathrm{c}}$ |
| $\mathbf{j}$ | $\mathrm{PhSO}_{2}$ | 150.32 | $119.40^{\mathrm{d}}$ | $113.93^{\mathrm{d}}$ | 156.31 | 127.86 | 120.11 |
| $\mathbf{n}$ | $\mathrm{MeS}^{\mathbf{o}}$ | $153.53^{\mathrm{e}}$ | $115.00^{\mathrm{f}}$ | $115.52^{\mathrm{f}}$ | $150.15^{\mathrm{e}}$ | 128.58 | 114.87 |
| $\mathbf{0}$ | $\mathrm{MeSO}_{2}$ | 149.80 | $119.40^{\mathrm{g}}$ | $113.86^{\mathrm{g}}$ | 156.12 | 127.86 | 120.24 |

${ }^{\text {a }}$ Values assigned by the same letter were interchanged.

Table III
Chemical shift data of 5-X-2-furaldehydes 2 (experiment B)

| 2 | X | C5 | C4 | C3 | C2 | $\mathrm{C}_{\alpha}$ | H4 | H3 | $\mathrm{H}_{\alpha}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| a | Me V N | 163.99 | 86.06 | 131.77 | 144.22 | 170.49 | 5.24 | 7.20 | 8.95 |
| b | PhS | 153.70 | 116.99 | 121.65 | 154.50 | 176.89 | 6.57 | 7.21 | 9.57 |
| c | Br | 130.79 | 114.58 | 122.32 | 154.39 | 176.14 | 6.57 | 7.19 | 9.54 |
| d | 1 | 98.92 | 123.07 | 122.16 | 157.88 | 175.98 | 6.78 | 7.11 | 9.51 |
| e | Ph | 159.24 | 107.50 | 123.30 | 151.89 | 177.02 | 6.82 | 7.29 | 9.63 |
| f | Me | 158,24 | 109.24 | 123.61 | 151.62 | 176.58 | 6.25 | 7.19 | 9.50 |
| g | H | 147.93 | 112.40 | 120.88 | 152.80 | 177.72 | 6.62 | 7.27 | 9.67 |
| h | $\mathrm{PhSO}_{2}$ | 153.83 | 117.60 | 118.48 | 154.54 | 178.19 | 7.24 | 7.30 | 9.71 |
| i | COOMe | 153.79 | 118.51 | 118.66 | 158.27 | 178.79 | 7.27 | 7.27 | 9.81 |
| j | CN | 129.29 | 122.80 | 122.48 | 154.44 | 177.49 | 7.27 | 7.32 | 9.78 |
| k | $\mathrm{NO}_{2}$ | 152.93 | 111.62 | 118.90 | 150.82 | 178.15 | 7.45 | 7.39 | 9.84 |

Table IV
Chemical shift data of 3-[(5-X-2-furyl)methylene]pentane-2,4-diones 4 (experiment G)

| $\mathbf{4}$ | X | C 5 | C 4 | C 3 | C 2 | $\mathrm{C}_{\alpha}$ | $\mathrm{C}_{\beta}$ | $\mathrm{CO}-\mathrm{E}$ | $\mathrm{CO}-\mathrm{Z}$ | $\mathrm{Me}-\mathrm{E}$ | $\mathrm{Me}-\mathrm{Z}$ |
| :---: | :---: | ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{a}$ | $\mathrm{Me}_{2} \mathrm{~N}$ | 163.12 | 87.40 | 124.56 | 129.19 | 127.69 | 139.22 | 194.45 | 204.37 | 25.53 | 31.76 |
| $\mathbf{b}$ | PhS | 147.72 | 120.27 | 120.55 | 151.83 | 124.28 | 138.75 | 196.62 | 203.28 | 25.95 | 30.57 |
| $\mathbf{c}$ | Cl | 139.48 | 110.58 | 120.92 | 148.56 | 124.12 | 138.09 | 196.64 | 203.44 | 25.93 | 30.90 |
| $\mathbf{d}$ | Br | 127.08 | 115.36 | 120.99 | 150.70 | 123.99 | 138.11 | 196.63 | 203.44 | 25.92 | 30.69 |
| $\mathbf{e}$ | I | 99.36 | 123.43 | 120.91 | 153.63 | 123.74 | 138.08 | 196.56 | 203.46 | 25.91 | 30.96 |
| $\mathbf{f}$ | Ph | 157.06 | 109.23 | 121.57 | 148.25 | 124.88 | 137.46 | 196.54 | 204.00 | 25.86 | 31.07 |
| $\mathbf{g}$ | Me | 157.37 | 109.96 | 120.65 | 147.28 | 125.12 | 136.53 | 196.44 | 203.86 | 25.82 | 30.96 |
| $\mathbf{h}$ | H | 147.59 | 113.20 | 119.01 | 148.66 | 123.18 | 137.80 | 196.70 | 203.66 | 25.89 | 31.14 |
| $\mathbf{i}$ | $\mathrm{PhSO}_{2}$ | 151.13 | 118.54 | 119.39 | 152.91 | 124.64 | 138.60 | 196.77 | 202.90 | 26.11 | 30.64 |
| $\mathbf{j}$ | COOMe | 145.81 | 119.90 | 118.94 | 151.58 | 125.19 | 141.00 | 196.83 | 203.29 | 26.13 | 30.70 |
| $\mathbf{k}$ | CN | 126.92 | 123.40 | 118.41 | 152.97 | 123.38 | 141.55 | 196.89 | 203.15 | 26.15 | 30.90 |
| $\mathbf{I}$ | NO | 152.66 | 114.16 | 119.77 | 150.39 | 124.30 | 142.46 | 196.76 | 202.96 | 26.26 | 30.65 |
| $\mathbf{m}$ | MeS | 153.70 | 112.93 | 121.15 | 149.66 | 124.38 | 136.99 | 196.44 | 203.77 | 25.84 | 31.13 |

group and the absolute value of $b_{3}$ was less than one. Since the $Y$ group in our 2-X-5-Y-disubstituted furans $\mathbf{1 - 6}$ is rather an electron withdrawing substituent, then the large negative slope at $z_{m}$ probably reflects a specific modification of $Y$ substituent by the furan ring.

Chemical shift of ortho carbon atom in p-disubstituted ( $\mathrm{X}, \mathrm{Y}$ )benzenes significantly ${ }^{8}$ depends only on $z_{0}$. However, in 2-substituted naphthalenes ${ }^{7}$, chemical shifts in both ortho positions can be explained by a regression with $z_{0}$ and $z_{p}$ variables; positive slope +0.52 for $\delta_{\mathrm{C} 1}$ and negative slope -0.49 for $\delta_{C 3}$ follow the increase and decrease in the ab initio bond order between C2-C1 (1.535) and C2-C3 (1.331) atoms, respectively, as compared with benzene CC bond (1.432). In furans 1-6, the ${ }^{13} \mathrm{C}$ chemical shift in ortho position to the X -substituent (C3 in 1, C4 in others) needs, besides $\mathrm{Z}_{0}$, also $z_{p}$ in regression. The values of $b_{4}$ in the range $0.728-1.051$ correspond very well to the increase in the bond order; in furan, the C2-C3 bond order has been calculated as 1.660.

Position meta in furans 1-6 (C4 in 1, C3 in others) substantially differs from that in substituted benzenes; the corresponding ${ }^{13} \mathrm{C}$ chemical shift does not depend on the $z_{m}$ increment but on the $z_{p}$ one. Generally, the

Table V
Chemical shift data of 2-[(5-X-2-furyl)methylene]malononitriles 6 (experiment J)

| $\mathbf{6}$ | X | C 5 | C 4 | C 3 | C 2 | $\mathrm{C}_{\alpha}$ | $\mathrm{C}_{\beta}$ | $\mathrm{CN}-\mathrm{Z}$ | $\mathrm{CN}-\mathrm{E}$ |
| :--- | :--- | ---: | ---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{a}$ | Me N | 164.95 | 91.41 | 133.25 | 140.29 | 134.89 | 59.51 | 116.33 | 117.59 |
| $\mathbf{b}$ | PhO | 164.14 | 91.66 | 127.72 | 139.99 | 141.07 | 72.27 | 113.18 | 114.50 |
| $\mathbf{c}$ | PhS | 157.69 | 116.33 | 124.68 | 149.44 | 141.19 | 76.00 | 112.61 | 113.90 |
| $\mathbf{d}$ | Cl | 147.26 | 111.56 | 124.75 | 145.29 | 141.46 | 77.54 | 112.16 | 113.39 |
| $\mathbf{e}$ | Br | 149.63 | 116.54 | 124.60 | 132.34 | 141.39 | 77.70 | 112.18 | 113.38 |
| $\mathbf{f}$ | I | 100.67 | 124.34 | 125.08 | 153.00 | 141.10 | 77.80 | 112.20 | 113.37 |
| $\mathbf{g}$ | Ph | 161.37 | 109.36 | 126.24 | 147.21 | 141.18 | 75.43 | 113.28 | 114.27 |
| $\mathbf{h}$ | Me | 161.80 | 111.64 | 125.51 | 146.91 | 142.06 | 74.62 | 112.90 | 114.21 |
| $\mathbf{i}$ | H | 149.40 | 114.27 | 123.33 | 147.91 | 142.91 | 77.36 | 112.43 | 113.63 |
| $\mathbf{j}$ | COOMe | 148.52 | 119.61 | 121.83 | 149.41 | 143.13 | 82.05 | 111.77 | 112.76 |
| $\mathbf{k}$ | CN | 130.20 | 120.88 | 123.52 | 150.32 | 141.73 | 83.80 | 111.31 | 112.30 |
| $\mathbf{I}$ | $\mathrm{PhSO}_{2}$ | 155.01 | 118.04 | 121.89 | 150.25 | 142.11 | 82.86 | 111.40 | 112.54 |

Table VI
Regression parameters, correlation coefficients, and standard deviation of data for $\mathrm{C}_{\mathrm{ipso}}$ carbon atoms

| $\mathrm{C}_{\text {ipso }}$ | Experi- <br> ment | $\mathrm{b}_{0}$ | $\mathrm{~b}_{1}$ | $\mathrm{~b}_{3}$ | $\mathrm{~b}_{4}$ | r | $\mathrm{s}_{\mathrm{y}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{1}$ | A | 1.120 | 1.008 | -10.339 |  | 0.9858 | 19.12 |
| $\mathbf{2}$ | B | 1.168 | 0.939 | -10.426 |  | 0.9893 | 19.08 |
| $\mathbf{2}$ | C | 1.139 | 0.905 | -6.578 |  | 0.9657 | 18.05 |
| $\mathbf{3}$ | D | 1.108 | 0.960 | -7.392 |  | 0.9691 | 17.51 |
| $\mathbf{4}$ | E | 1.126 | 0.940 | -7.145 |  | 0.9684 | 18.19 |
| $\mathbf{4}$ | F | 1.124 | 0.880 | -6.980 |  | 0.9703 | 17.12 |
| $\mathbf{4}$ | G | 1.142 | 0.915 | -9.478 | $(-0.341)^{\mathrm{a}}$ | 0.9963 | 17.21 |
| $\mathbf{5}$ | H | 1.144 | 0.879 | -7.395 |  | 0.9564 | 18.83 |
| $\mathbf{5}$ | I | 1.140 | 0.850 | -6.363 |  | 0.9640 | 17.58 |
| $\mathbf{6}$ | J | 1.145 | 0.961 | $(-5.329)^{\mathrm{b}}$ |  | 0.9679 | 18.75 |

Parameters b calculated according to reduced Eq. (3) which are significant by Student t-test at the $95 \%$ confidence level. ${ }^{\text {a }}$ Not significant at the $99 \%$ level; ${ }^{\text {b }}$ significant at the $90 \%$ level.

Table VII
Regression parameters, correlation coefficients, and standard deviation of data for $\mathrm{C}_{\text {ortho }}$ carbon atoms

| $\mathrm{C}_{\text {ortho }}$ | Experiment | $\mathrm{b}_{0}$ | $\mathrm{~b}_{2}$ | $\mathrm{~b}_{4}$ | r | $\mathrm{s}_{\mathrm{y}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{1}$ | A | 0.870 | 1.309 | 1.381 | 0.9576 | 10.98 |
| $\mathbf{2}$ | B | 0.880 | 1.177 | 0.886 | 0.9859 | 10.19 |
| $\mathbf{2}$ | C | 0.875 | 1.248 | 0.882 | 0.9597 | 11.38 |
| $\mathbf{3}$ | D | 0.887 | 1.206 | 1.051 | 0.9598 | 10.94 |
| $\mathbf{4}$ | E | 0.885 | 1.232 | 0.858 | 0.9736 | 10.70 |
| $\mathbf{4}$ | F | 0.890 | 1.198 | 0.942 | 0.9742 | 10.83 |
| $\mathbf{4}$ | G | 0.891 | 1.110 | 0.901 | 0.9699 | 9.28 |
| $\mathbf{5}$ | H | 0.888 | 1.188 | 0.728 | 0.9756 | 10.65 |
| $\mathbf{5}$ | I | 0.891 | 1.218 | 0.810 | 0.9855 | 11.06 |
| $\mathbf{6}$ | J | 0.889 | 1.159 | 0.804 | 0.9797 | 10.81 |

Parameters b calculated according to reduced Eq. (3) which are significant by Student t-test at the $95 \%$ confidence level.

Table VIII
Regression parameters, correlation coefficients, and standard deviation of data for $\mathrm{C}_{\text {meta }}$ carbon atoms

| $\mathrm{C}_{\text {meta }}$ | Experiment | $\mathrm{b}_{0}$ | $\mathrm{~b}_{3}$ | $\mathrm{~b}_{4}$ | r | $\mathrm{s}_{\mathrm{y}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{1}$ | A | 0.870 | 0.734 | 0.147 | 0.6878 | 0.96 |
| $\mathbf{2}$ | B | 0.951 |  | -0.680 | 0.9183 | 3.66 |
| $\mathbf{2}$ | C | 0.953 |  | -0.635 | 0.9334 | 3.19 |
| $\mathbf{3}$ | D | 0.896 |  | -0.220 | 0.7204 | 1.49 |
| $\mathbf{4}$ | E | 0.930 |  | -0.383 | 0.9345 | 1.96 |
| $\mathbf{4}$ | F | 0.937 |  | -0.315 | 0.8973 | 1.68 |
| $\mathbf{4}$ | G | 0.937 |  | -0.301 | 0.8933 | 1.57 |
| $\mathbf{5}$ | H | 0.960 |  | -0.616 | 0.9642 | 3.09 |
| $\mathbf{5}$ | I | 0.965 | $(1.734)^{\mathrm{a}}$ | $(-0.196)^{\mathrm{b}}$ | 0.6218 | 2.41 |
| $\mathbf{6}$ | J | 0.965 |  | -0.674 | 0.9629 | 3.27 |

Parameters b calculated according to reduced Eq. (3) which are significant by Student t-test at the $95 \%$ confidence level. ${ }^{\text {a }}$ Significant only at the $90 \%$ level; ${ }^{\text {b }}$ significant only at the $80 \%$ level.

Table IX
Regression parameters, correlation coefficients, and standard deviation of data for $\mathrm{C}_{\text {para }}$ carbon atoms

| $\mathrm{C}_{\text {para }}$ | Experi- <br> ment | $\mathrm{b}_{0}$ | $\mathrm{~b}_{1}$ | $\mathrm{~b}_{2}$ | $\mathrm{~b}_{4}$ | r | $\mathrm{s}_{\mathrm{y}}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{1}$ | A | 1.128 |  | 0.440 | 0.587 | 0.9504 | 4.17 |
| $\mathbf{2}$ | B | 1.193 |  | 0.558 |  | 0.8968 | 3.80 |
| $\mathbf{2}$ | C | 1.187 |  | 0.574 |  | 0.9635 | 3.80 |
| $\mathbf{3}$ | D | 1.191 |  | 0.402 | 0.445 | 0.9643 | 4.03 |
| $\mathbf{4}$ | E | 1.166 |  | 0.437 | 0.372 | 0.9593 | 4.13 |
| $\mathbf{4}$ | F | 1.166 |  | 0.436 | 0.403 | 0.9620 | 4.24 |
| $\mathbf{4}$ | G | 1.159 | $(0.122)^{\mathrm{a}}$ | 1.014 | 0.558 | 0.9803 | 6.25 |
| $\mathbf{5}$ | H | 1.162 |  | 0.471 | 0.243 | 0.9778 | 4.04 |
| $\mathbf{5}$ | I | 1.160 |  | 0.459 | 0.302 | 0.9770 | 4.20 |
| $\mathbf{6}$ | J | 1.154 |  | 0.459 | $(0.217)^{\mathrm{b}}$ | 0.9489 | 3.92 |

Parameters b calculated according to reduced Eq. (3) which are significant by Student t-test at the $95 \%$ confidence level. ${ }^{\text {a }}$ Not significant at the $99 \%$ level; ${ }^{\text {b }}$ significant only at the $90 \%$ level.
variance of the shift data is relatively small (standard deviation, $\mathrm{s}_{\mathrm{y}}$, is below two); these results can be questioned, but in a few cases with higher variance, the trend is obvious. In p-disubstituted ( $\mathrm{X}, \mathrm{Y}$ )benzenes, a negative slope at the variable for para position either confirms the donor character ${ }^{8}$ of the $Y$ substituent or reflects a contribution of diamagnetic anisotropy due to changing the position of the multiple bond in $Y$ group ${ }^{5 b}$ relative to the plane of the ring. Since chemical shifts of $\beta$ carbon in the $Y$ substituent of 3-6 correlate very well with $z_{p}$ increments, different polarization of the $\mathrm{C}_{\alpha}=\mathrm{C}_{\beta}$ bond can be expected and the diamagnetic anisotropy field around it can be changed with variation of the $X$ substituent.

Not less interesting results were found in regression of ${ }^{13} \mathrm{C}$ chemical shifts of C5 in 1 and C2 in others furans. They correlate both with $z_{0}$ and $z_{p}$ making from this position another "ortho" position to substituent $X$, however, with slopes about half of those in Table VII.

## CONCLUSIONS

Multivariate regression proved that the corresponding position in benzenes and furans are not affected by substituents in the same way: in $2-X-5-Y-$ substituted furans, the "ortho" position requires correction on the different bond order, the "meta" ${ }^{13} \mathrm{C}$ chemical shift is probably affected by change in diamagnetic contribution of the Y group, and "para" position behaves like another ortho position in the molecule. These general findings are not substantially affected either by change of NMR solvent ( $\mathrm{CDCl}_{3}$ for DM SO-d ${ }_{6}$ ) or by the spectrometer frequency and quality. Program ASSIGN based on the multivariate regression can be used to check and eventually improve the assignment of ${ }^{13} \mathrm{C}$ NMR chemical shifts in a series of compounds with one variable and one fixed substituent.

The authors wish to thank to Dr O. Humpa for measurements on a Bruker DRX-500 spectrometer and to MrV. Proks for calculation of ab initio bond order values.

## REFERENCES

1. Lynch B. M.: Can. J. Chem. 1977, 55, 541.
2. Holík M.: Org. Magn. Reson. 1977, 9, 491.
3. a) Membrey F., Ancian B., Doucet J.-P.: J. Chem. Soc., Perkin Trans. 2 1980, 1391; b) Chandrasekaran R., Perumal S., Wilson D. A.: Magn. Reson. Chem. 1987, 25, 1001.
4. Ewing D. F.: Org. Magn. Reson. 1979, 12, 499.
5. a) Holík M., Matějková B.: Collect. Czech. Chem. Commun. 1990, 55, 261; b) Holík M.: Chemom. Intell. Lab. Syst. 1993, 19, 225.
6. Bromilow J., Brownlee R. T. C., Craik D. J., Sadek M., Taft R. W.: J. Org. Chem. 1980, 45, 2429.
7. Holík M.: Collect. Czech. Chem. Commun. 1996, 61, 713.
8. Holík M.: J. Mol. Struct. 1999, 482-3, 345.
9. Craig D. J., Ternai B.: Org. Magn. Reson. 1981, 15, 268.
10. Johnels D., Edlund U., Johansson E., Wold S.: J. Magn. Reson. 1983, 55, 316.
11. a) Gronowitz S., Johnson I., Hörnfeldt A.-B.: Chem. Scr. 1975, 7, 211; b) Alvarez-Ibarra C., Quiroga-Feijoo M. L., Toledano E.: J. Chem. Soc., Perkin Trans. 2 1998, 679.
12. a) Levy G. C., Cargioli J. D.: J. Magn. Reson. 1972, 6, 143; b) Günther H.: NMR Spectroscopy, Basic Principles, Concenpts, and Applications in Chemistry, 2nd ed., p. 532. Wiley, New York 1995.
13. Marchalín Š., Mamani L. N. H., Ilavský D., Prónayová N., Leško J.: Collect. Czech. Chem. Соттии. 1993, 58, 1388.
14. a) Friedl Z., Böhm S., Goljer I., Piklerová A., Poórová D., Ríčková A., Kováč J.: Collect. Czech. Chem. Commun. 1987, 52, 409; b) Holík M., Friedl Z.: Proceedings of CHEMOMETRICS IV Conference, p. 91. Masaryk University, Brno 1996.
15. Marchalín Š.: Unpublished ${ }^{13} \mathrm{C}$ NMR data are available on request.
16. Marchalín Š., Ilavský D., Kováč J.: Z. Chem. 1987, 27, 406.
17. a) Musumarra G., Ballistreri F. P.: Chem. Scr. 1981, 18, 209; b) Ilavský D., Goljer I., Kováč J.: Z. Chem. 1990, 30, 99.

[^0]:    + For Parts I and II of this series see refs ${ }^{7,8}$.

