The prediction of the ^{19}F NMR spectra of fluoroarenes using 'statistical' substituent chemical shift values

C.A.L. Mahaffy*
Department of Chemistry, Auburn University at Montgomery, Montgomery, AL 36117-3596 (USA)
and J.R. Nanney
Department of Mathematics, Auburn University at Montgomery, Montgomery, AL 36117-3596 (USA)

(Received February 1, 1993; accepted April 8, 1993)

Abstract

^{19}F NMR data for fluoroarenes from a previous publication (Nanney *et al., J. Fluorine Chem., 64* (1993) 217) together with additional available literature data and some new data have been analyzed using a new method which determines substituent chemical shifts (SCS) statistically.

This new technique and over 400 of these 'statistical' SCS values (SSCS values) are reported, having been computed from a data base of 1961 signals. When these values are used to predict the signal position they are found to give a correlation coefficient of predicted versus observed values of 0.995 with a standard deviation of 2.46.

Introduction

We recently reported a number of mathematical models which could predict the ^{19}F NMR spectra of fluoroarenes [1] and fluoroarenetricarbonylchromium complexes [2, 3], including compounds which contained *ortho* substitution. While these models were very useful in the prediction of the ^{19}F NMR spectra, they suffer from the disadvantage that the various substituent parameters used in the prediction equations must be known. The field, $\mathscr{F}$, and resonance, $\mathscr{R}$, parameters together with the molar refractivity (*MR*) are known for the majority of common groups [4], but Charton's steric parameter, ν, is only known for a limited number of groups and is known only as a range for others [5].

Another approach to the prediction of the ^{19}F NMR spectra of arenes and arenemetal complexes is to use substituent chemical shifts (SCS) [6] but this requires that the chemical shift for the fluoroarene containing the group in question in the desired position be determined. The SCS value for that group in that position is then determined by subtracting the chemical shift for the parent fluorobenzene.

$$SCS_{group\ X} = \Delta^{F}_{FC_6H_4X} - \Delta^{F}_{C_6H_5F} \quad (1)$$

In the statistical approach which we report here, it is not necessary to determine the spectrum of an arene which contains only the group in question along with the fluoro substituent in order to determine the SCS value. It is only necessary that the desired group appears somewhere in the data base in that position in order for the 'statistical' SCS (SSCS) value to be determined.

In this paper we report the SSCS values for 414 groups which were determined from a data base of 1961 signals. When these values are used to predict the ^{19}F NMR signal positions, we find a correlation coefficient of observed against predicted signal positions of 0.995 with a standard deviation of 2.46.

The 'statistical' SCS method

The assumptions underlying the SCS method for predicting the ^{19}F NMR signal positions in fluoroarenes are:

(1) The effect of a group attached to the ring depends on whether the group is in the *ortho*, *meta* or *para* position, and so each group has three SCS values corresponding to the three different positions.

(2) The effect on the NMR signal of a group attached to the ring is independent of any other groups attached to the ring.

(3) The effect on the NMR signal of a group attached to the ring is solvent-dependent. Thus, each group has three SCS values for each solvent.

*Author to whom correspondence should be addressed.

SSDI 0022-1139(93)02934-7

Using eqn. (1), the number of compounds needed to compile a complete table of SCS values is $3kl$ where k is the number of groups and l is the number of solvents considered. However, studies by Fivolt *et al.* [6] and ourselves [1] have shown that, in most cases, the solvent effects are small. Ignoring the solvent would reduce the number of compounds required for a complete study to $3k$. We have found it necessary to ignore the solvent effects in this work because of the lack of variety in solvents in the data base, and there are also many spectra in the literature for which the solvent data are not given. Because of the statistical nature of this new method, an averaging effect corrects somewhat for solvent effects, and so essentially the SSCS values computed are for 'an average solvent'.

The SSCS method is based on assumptions (1) and (2) above, but uses a different procedure to compute the SSCS values. One of its advantages is that it can be applied to data bases and does not require the acquisition of the three isomeric fluorobenzenes containing only the group in question. The SSCS value for a group in the *ortho*, *meta* or *para* position can be computed from existing data bases provided the group appears in that position in some compound in the data base. It is not necessary that the other substituents of that compound be hydrogen.

The SSCS method is easily understood once the standard SCS method is put into a proper algebraic form. The number of variables used in the algebraic form is three times the number of groups, k, to be considered. The variables x_{ij}, $i=1, k$, $j=o, m, p$ (*ortho*, *metal*, *para*) are defined for a given compound by the formula:

$$x_{ij}=\begin{cases}0 \text{ if group } i \text{ is not in position } j\\ 1 \text{ if group } i \text{ is in one of the } j \text{ positions}\\ 2 \text{ if group } i \text{ is in both } j \text{ positions}\end{cases} \qquad (2)$$

Clearly the third case cannot occur with x_{ip}. Let S_{ij} be the SCS value of group i in position j, which for the moment we assume to be known. Then for a given compound, computing Δ^F_{pred} for that compound by means of formula (3) is equivalent to the usual SCS method.

$$\Delta^F_{pred.}=\sum\sum S_{ij}x_{ij}-113.9 \qquad (i=1, k;\ j=o, m, p) \qquad (3)$$

Clearly, at most, five of the x_{ij} values are non-zero. Although this formula is needlessly complex for computing $\Delta^F_{pred.}$ by the usual SCS method, it is given because it forms the basis for the SSCS method.

The SSCS method consists in realizing that, for each compound in the data base, $\Delta^F_{obs.}$ as well as the values of the variables x_{ij} are known for that compound, and thus S_{ij} can be computed as regression coefficients when eqn. (3) is used as the model in the multiple regression with x_{ij} as independent variables and $\Delta^F_{obs.}$ as the dependent variable. This has the effect of determining the values of S_{ij} that minimize the sum in eqn. (4) over all compounds in the data base. When S_{ij} is computed in this manner, we refer to it as the SSCS value of group i in position j.

$$\sum_{\text{data base}}(\Delta^F_{obs.}-\Delta^F_{pred.})^2$$

$$=\sum_{\text{data base}}\left(\Delta^F_{obs.}-\sum_{\text{groups}}\sum_{\text{positions}}S_{ij}x_{ij}+113.9\right)^2 \qquad (4)$$

As well as the advantages already noted, one additional advantage of this new method is that standard statistical information which measures how well the procedure works is given by the SAS software. The margins of error in computing the S_{ij} values can therefore be found. It would be expected that the margins of error are higher for groups with low frequency in the data base.

The available data base puts certain limitations on the SSCS method. If group i does not appear in position j anywhere in the data base, obviously it cannot be determined. This is also, of course, true in the traditional SCS method as well. There are also other conditions which will, in some cases, prohibit determination of S_{ij}. For instance, if group i occurs only twice, once in *ortho* and once in *meta*, whether or not $S_{i\,ortho}$ and $S_{i\,meta}$ can be determined depends on whether both occurrences of group i are in the same compound or not.

The lack of variety in solvents in the data base prohibits the further breaking down of S_{ij} into S_{ijh}, which would represent the SSCS value for group i in position j in solvent h. Ignoring solvent effects results in an SSCS value which is the average for the solvents represented in the data base. It will also be noted that neither the SCS nor SSCS methods as presently developed can account for interaction effects between groups. It is theoretically possible to add variables to eqn. (3) which measure these effects, though the number of possible interactions is so large that an enormous data base, much larger than is presently available, would be needed to evaluate all the coefficients.

Results and discussion

The data base

The same literature search was used as in the previous study [1]. Unlike the previous study, however, we did not require to know the solvent and were not restricted to compounds where we knew the parameters $\mathscr{F}$, $\mathscr{R}$, MR and ν. We did not include any derivatives of biphenyl because this method cannot take account of systems containing restricted rotation, but did include some systems containing two rings which were not directly

bonded to each other. We also included several systems containing σ-bonded metal fragments. Systems containing heteroatoms, π-bonded metal moieties or fused ring systems were excluded*.

The data base consists of 1961 resonances some of which are for different fluorine atoms in the same molecule and some of which are the same signal observed in different solvents.

*Separate studies are currently being undertaken on pyridine, biphenyl and fused ring systems.

The sort code

In order that compounds could be sorted logically so that a particular structure could be found quickly and easily, a sort code was established in which a compound is identified in the computer programs by its twelve letter sort code, the key to which is given in Table 1. The letters of the sort code tell the program which group is in each position. The fluorine atom which is being studied is always in position 1. Fluoroarenes with more than one signal are encoded separately for each signal.

TABLE 1. Sort codes, groups, SSCS values and standard error of estimate of SSCS values for the *ortho*, *meta* and *para* positions

Sort code	Group	SSCS value			Standard error of estimate		
		Ortho	*Meta*	*Para*	*Ortho*	*Meta*	*Para*
AA	H	0.00	0.00	0.00	0.0	0.0	0.0
Ac	$COCH_3$	2.49	1.79	7.62	0.7	0.6	0.8
BA	BF_2	11.92	4.49	13.37	2.5	2.5	2.5
BB	BCl_2	10.52	3.89	11.37	2.5	2.5	2.5
BO	$B(OH)_2$	6.82	0.79	2.09	2.5	2.5	1.7
Br	Br	7.61	3.52	0.14	0.3	0.3	0.5
CA	$C(CH_3)_3$	1.07	0.20	−4.80	2.5	2.5	2.5
CB	CH_2Br	−3.18	2.33	1.64	1.8	1.4	1.2
CC	$CH(CH_3)_2$	[a]	0.10	−4.10	[a]	2.5	2.5
CD	$CH=CH_2$	−4.41	0.73	−0.55	1.4	1.4	1.4
CE	CH_2COCl	−3.88	2.69	2.77	2.5	2.5	2.5
CF	CF_3	0.42	3.08	5.84	0.4	0.3	0.5
CG	CF_2CF_3	3.82	4.15	8.34	2.5	1.7	1.7
CH	CH_2OH	−5.90	1.49	−0.65	1.3	1.1	1.2
CI	CHF_2	−3.48	3.79	6.77	2.5	2.5	2.5
CJ	CF_2CH_3	0.92	4.19	5.67	2.5	2.5	2.5
CK	CF_2CF_3	−0.58	2.89	3.07	2.5	2.5	2.5
CL	$CHFCF_3$	−0.08	4.49	7.67	2.5	2.5	2.5
CM	$CH[COOEt]_2$	0.06	2.44	1.57	1.7	1.7	2.5
CN	CN	6.87	4.11	10.13	0.4	0.4	0.5
CP	CH_2CN	−3.78	2.20	0.00	2.5	1.7	1.2
CQ	CO_2Br	−10.62	14.81	[a]	2.5	1.8	[a]
CR	$CHCl_2$	−4.30	2.20	3.30	2.5	2.5	2.5
CS	CH_2SOMe	[a]	[a]	−0.60	[a]	[a]	2.5
CT	CH_2Cl	−7.70	−4.36	−0.43	1.7	1.4	1.7
CU	$CH_2N=C(NH_2)_2$	[a]	−0.50	−3.40	[a]	2.5	2.5
CV	$CH_2CH_2NH_2$	−0.90	5.82	4.23	1.8	1.8	2.5
CW	$COCH_2Ac$	−0.78	3.09	4.87	2.5	2.5	2.5
CY	CH_2COCH_3	[a]	[a]	−2.60	[a]	[a]	2.5
CZ	CH_2CO_2Me	[a]	−0.10	−2.50	[a]	2.5	1.7
Ca	CH_2CONMe_2	[a]	[a]	−6.51	[a]	[a]	2.5
Cb	COOMe	3.25	3.84	7.07	2.5	1.2	0.9
Cc	COCl	3.36	3.45	12.85	1.1	0.9	1.1
Cd	COCN	[a]	4.10	[a]	[a]	2.5	[a]
Ce	$COCF_3$	[a]	3.40	13.10	[a]	2.5	2.5
Cf	COF	−14.38	3.00	6.24	2.5	1.7	1.7
Cg	$C[CN]_3$	[a]	7.20	7.50	[a]	2.5	2.5
Ci	C≡CH	[a]	[a]	3.30	[a]	[a]	2.5
Cj	$CF[CF_3]_2$	[a]	3.60	4.80	[a]	2.5	2.5
Ck	COEt	[a]	[a]	7.40	[a]	[a]	2.5
Cl	Cl	−0.34	3.45	−0.68	0.2	0.2	0.3
Cm	CHO	−7.39	2.08	10.34	1.2	1.1	1.1
Cn	$CONH_2$	0.45	−0.78	3.40	1.7	1.4	1.1

(*continued*)

TABLE 1. (*continued*)

Sort code	Group	SSCS value			Standard error of estimate		
		Ortho	*Meta*	*Para*	*Ortho*	*Meta*	*Para*
Co	COOH	2.30	1.11	6.46	0.4	0.4	0.5
Cq	CH_2COOH	[a]	−0.10	−2.61	[a]	2.5	1.4
Cs	CH_2SiMe_3	[a]	[a]	−6.20	[a]	[a]	2.5
Ct	$CH_2CH(NH_2)COOH$	0.72	[a]	[a]	2.5	[a]	[a]
Cu	$CH(OH)CH_3$	−3.78	1.05	−1.66	2.5	1.7	1.7
Cv	$CHBr_2$	[a]	1.30	[a]	[a]	2.5	[a]
Cw	$C[CF_3]_2OH$	[a]	2.10	2.80	[a]	2.5	2.5
Cx	n-C_4F_9	[a]	3.10	[a]	[a]	2.5	[a]
Cy	COC_3F_7	[a]	[a]	13.10	[a]	[a]	2.5
Cz	$COSCH_3$	[a]	[a]	6.80	[a]	[a]	2.5
DA	CH_2NH_2	[a]	[a]	−2.80	[a]	[a]	2.5
DB	CCl_3	9.90	1.90	3.40	2.5	2.5	2.5
DC	$CF=CF_2$	2.32	3.29	6.67	2.5	2.5	2.5
DD	$CCl=CCl_2$	1.72	3.89	5.77	2.5	2.5	2.5
DE	CH_2F	−12.58	2.39	12.77	2.5	2.5	2.5
DF	COOEt	1.20	1.17	6.43	0.8	0.7	1.0
Et	CH_2CH_3	[a]	−0.20	−4.29	3.5	2.5	1.7
Fe	$Fe(CO)_2Cp$	32.59	−2.88	−9.32	0.5	0.6	1.3
Fl	F	−23.20	1.96	−6.55	0.1	0.1	0.2
GM	$GeMe_3$	11.82	1.75	1.64	2.5	1.7	1.7
Hg	HgMe	15.22	2.19	0.27	2.5	2.5	2.5
IC	ICl_2	[a]	[a]	7.90	[a]	[a]	2.5
ID	IF_2	[a]	[a]	16.90	[a]	[a]	2.5
Io	I	19.93	3.55	1.39	0.6	0.5	0.7
Me	CH_3	−3.87	−0.36	−3.61	0.4	0.4	0.4
Mg	MgBr	25.22	1.79	−3.23	2.5	2.5	2.5
Mn	$Mn(CO)_5$	36.53	−0.59	−1.13	1.1	1.4	2.5
NA	$NHCH_2CH_2OH$	−18.18	−1.11	−9.93	2.5	2.5	2.5
NB	$N[CH_2CH_2OH]_2$	−13.18	−4.75	[a]	1.8	1.8	[a]
NC	$NHCOCH_3$	−11.16	1.76	−5.54	0.8	0.8	1.0
ND	NHOH	[a]	0.30	−11.40	[a]	2.5	1.7
NF	NO_2	−5.63	3.76	9.61	0.3	0.3	0.4
NH	NH_2	−22.88	−1.27	−17.40	0.3	0.3	0.5
NJ	$NHCONH_2$	[a]	0.90	−8.10	[a]	2.5	2.5
NL	$NHCONHCONH_2$	[a]	2.20	−6.30	[a]	2.5	2.5
NM	$N(CH_3)_2$	−13.51	1.21	−7.52	0.6	0.6	0.9
NN	$NHNH_2$	−16.13	−1.28	−11.66	1.1	0.8	1.5
NO	NO	[a]	2.60	11.75	[a]	1.7	1.7
NQ	$NHCOCF_3$	[a]	[a]	−2.30	[a]	[a]	2.5
NR	$NHCONHCH_3$	[a]	0.60	−9.70	[a]	1.7	1.7
NS	NHCN	[a]	[a]	−7.30	[a]	[a]	2.5
NT	$N[SnMe_3]_2$	−12.28	−1.11	−16.63	2.5	2.5	2.5
NU	$NHSO_2CH_3$	[a]	[a]	−5.20	[a]	[a]	2.5
NV	$NHCONMe_2$	[a]	[a]	−7.90	[a]	[a]	2.5
NW	$N=PCl_3$	−11.42	3.30	[a]	2.5	2.5	[a]
NX	NHOMe	[a]	[a]	−10.30	[a]	[a]	2.5
NZ	N_3	−11.43	2.84	−0.31	0.9	0.8	1.5
Nc	NCO	−9.20	2.34	−2.15	1.7	1.4	1.7
Nd	NCS	[a]	3.20	0.20	[a]	2.5	2.5
Ne	NHCOOMe	[a]	0.10	−7.05	[a]	2.5	1.7
Nf	$N[COF]_2$	[a]	4.30	4.00	[a]	2.5	2.5
Ng	$N[CF_3]_2$	[a]	3.70	4.00	[a]	2.5	2.5
Nh	$NHCH_2CH_2NH_2$	−21.38	−1.21	−18.23	2.5	2.5	2.5
Ni	$N=CCl_2$	−10.38	2.19	−2.73	2.5	2.5	2.5
Nj	$NHNO_2$	[a]	[a]	−0.50	[a]	[a]	2.5
Nk	$NHNHCOCH_3$	[a]	[a]	−12.80	[a]	[a]	2.5
Nm	$NHCH_3$	−21.27	0.73	−17.91	0.6	0.5	0.8
Ns	$N[SiMe_3]_2$	−8.18	2.19	−9.03	2.5	2.5	2.5

(*continued*)

TABLE 1. (*continued*)

Sort code	Group	SSCS value			Standard error of estimate		
		Ortho	*Meta*	*Para*	*Ortho*	*Meta*	*Para*
Nt	$NHSnMe_3$	−25.28	−1.11	−24.63	2.5	2.5	2.5
OA	OCH_2CH_2OH	−21.38	−1.71	−17.93	2.5	2.5	2.5
OB	OCH_2COOH	−17.48	−0.71	−8.83	2.5	2.5	2.5
OE	OCH_2CH_3	−17.47	−0.11	−8.58	0.6	0.5	0.9
OF	OCH_2CF_3	−18.54	−0.59	−6.50	0.8	0.9	1.5
OG	$OCCl_3$	−8.02	0.71	−1.03	1.7	1.7	2.5
OH	OH	−23.52	−0.04	−13.27	0.4	0.4	0.6
OJ	$OCOCF_3$	[a]	[a]	−0.80	[a]	[a]	1.7
OK	OCO_2Me	[a]	[a]	−2.80	[a]	[a]	2.5
OL	OCOF	[a]	[a]	−1.10	[a]	[a]	2.5
OM	$OCONMe_2$	[a]	[a]	−4.70	[a]	[a]	2.5
ON	ONO	−18.09	3.64	13.07	1.7	1.6	2.5
OO	OCOMe	[a]	[a]	−3.66	[a]	[a]	1.4
OP	OSOF	−11.28	2.89	0.67	2.5	2.5	2.5
OQ	$OC(Me)_2CH{=}CH_2$	−13.68	−0.81	−11.43	2.5	2.5	2.5
OS	OSO_2F	−12.08	4.29	3.07	2.5	2.5	2.5
OT	OCF_3	15.18	1.61	0.19	1.2	1.1	1.4
OU	O-Bu^n	−17.47	−1.07	[a]	2.5	2.5	[a]
OV	OSO_2Me	[a]	[a]	−1.00	[a]	[a]	2.5
OX	OCH_3	−18.94	0.78	−9.02	0.4	0.3	0.5
PA	PH_2	9.62	1.89	[a]	2.5	2.5	[a]
PB	PMe_2	6.32	0.59	0.57	2.5	2.5	2.5
PC	PEt_2	7.42	0.59	1.47	2.5	2.5	2.5
PD	$P[NMe_2]_2$	−0.48	0.79	−0.73	2.5	2.5	2.5
PE	PF_2	−0.38	1.39	7.57	2.5	2.5	2.5
PF	PCl_2	8.12	4.09	9.87	2.5	2.5	2.5
PG	$P[NMe_2]Cl$	10.72	2.39	5.37	2.5	2.5	2.5
PH	$P[NBu^t]_2$	2.02	2.79	0.67	2.5	2.5	2.5
PI	$POCl_2$	[a]	[a]	12.10	[a]	[a]	2.5
PM	$PbMe_3$	[a]	0.40	−0.40	[a]	2.5	2.5
RB	$CH_2C_6H_5$	[a]	[a]	−3.40	[a]	[a]	2.5
RD	COC_6H_5	[a]	[a]	7.10	[a]	[a]	1.2
RE	COC_6F_5	−1.74	1.48	9.97	1.4	1.4	2.5
RF	COC_6F_4H-*o*	3.29	4.11	10.64	2.5	1.7	2.5
RG	COC_6F_4OMe-*p*	−2.99	2.73	[a]	2.5	2.5	[a]
RW	$SC_6H_4SCH_3$-*p*	7.95	[a]	[a]	2.5	[a]	[a]
RX	COC_6F_4COOH-*p*	−2.10	1.69	9.57	1.8	1.8	2.5
RZ	OC_6H_4Br-*m*	−2.36	[a]	[a]	1.1	[a]	[a]
Ra	NHC_6H_5	−14.02	−3.18	[a]	2.5	2.5	[a]
Re	$Re(CO)_5$	39.50	−1.93	[a]	1.3	1.8	[a]
Ri	cyclo-C_5H_9	17.39	2.31	1.90	1.5	1.3	2.5
Rj	cyclo-C_6H_{11}	−8.78	−4.61	−8.63	2.5	2.5	2.5
Rk	SC_6H_5	8.82	3.42	2.18	0.8	0.8	1.0
Rl	$OCOC_6H_5$	[a]	[a]	−3.00	[a]	[a]	2.5
Rm	$NHCOC_6H_5$	[a]	[a]	−4.90	[a]	[a]	2.5
Rn	$CH_2SC_6H_5$	[a]	[a]	−1.50	[a]	[a]	2.5
Ro	$CH_2COC_6H_5$	[a]	[a]	−2.60	[a]	[a]	2.5
Rq	OC_6H_5	[a]	2.70	−6.70	[a]	2.5	2.5
Rr	OC_6H_4F-*p*	[a]	0.30	−5.90	[a]	2.5	2.5
SA	SOF_3	[a]	[a]	12.40	[a]	[a]	2.5
SB	SO_2F	7.46	5.76	13.75	2.5	2.5	1.7
SC	SCF_3	8.60	3.10	−3.90	2.5	2.5	2.5
SD	SO_2CF_3	9.50	5.50	−14.30	2.5	2.5	2.5
SF	SOMe	[a]	[a]	3.80	[a]	[a]	2.5
SG	SOOMe	2.24	5.89	10.01	1.3	0.7	1.3
SH	SH	9.95	0.90	−3.50	3.5	2.5	1.7
SI	SOF	[a]	[a]	11.40	[a]	[a]	2.5
SJ	SF_5	[a]	3.90	6.30	[a]	2.5	2.5

(*continued*)

TABLE 1. (*continued*)

Sort code	Group	SSCS value			Standard error of estimate		
		Ortho	*Meta*	*Para*	*Ortho*	*Meta*	*Para*
SK	SO_3Et	[a]	3.70	9.10	[a]	2.5	2.5
SL	SO_2NH_2	[a]	[a]	8.30	[a]	[a]	2.5
SM	SCH_3	6.50	1.24	−4.53	0.4	0.4	0.9
SN	SO_2Cl	[a]	[a]	13.00	[a]	[a]	2.5
TC	$SnCl_3$	[a]	1.85	2.14	[a]	1.7	1.7
TM	$SnMe_3$	16.32	2.20	2.12	1.8	1.7	1.4
Tl	TlBr	20.42	−2.11	−1.83	2.5	2.5	2.5
UW	$COOC_6H_5$	[a]	[a]	8.50	[a]	[a]	2.5
UY	*N*-morpholino	−12.78	0.80	−2.46	1.2	0.8	1.4
VH	C≡CPh	3.48	−1.18	0.94	0.8	0.7	1.8
VI	*N*-piperidino	−11.52	−2.18	−2.38	1.0	0.9	1.8
VJ	*N*-pyrrolidino	−13.97	−2.27	−0.43	1.2	1.2	2.9
VK	*N*-homopiperidino	−9.96	−3.22	[a]	1.6	1.6	[a]
sA	$SiHMe_2$	14.86	0.73	2.11	1.7	1.0	1.9
sB	$SiMe_2OEt$	15.89	−0.60	−0.09	1.1	1.1	2.7
sC	$SiMe_2OMe$	15.59	−0.40	−0.41	1.1	1.1	2.7
sD	$SiMe_2OH$	15.81	−0.72	−1.64	1.2	1.2	2.8
sF	SiF_3	14.42	5.09	12.57	2.5	2.5	2.5
sM	$SiMe_3$	13.76	0.27	1.60	0.5	0.8	0.8
sf	SeF_3	[a]	[a]	9.40	[a]	[a]	2.5

[a]Not computable from the data base.

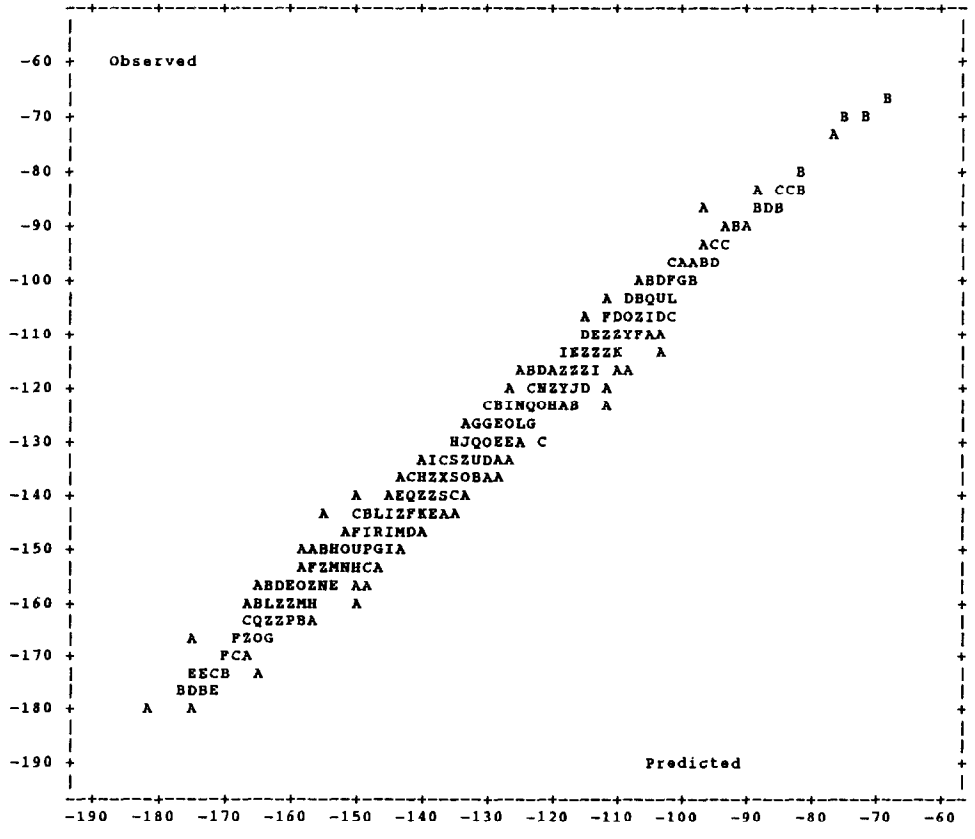

Fig. 1. Scatter plot of observed ^{19}F NMR signal position versus predicted position. B = two coincident data points, C = three coincident data points, etc.

The data were sorted alphabetically according to the sort code. Note that lower case letters are treated as different from capital letters. The computer treats all capital letters as coming before all lower case letters. Common groups are given sort codes which seemed logical, e.g. the sort code for methyl is Me, ethyl is Et and so on, but with a large number of groups it is not possible to do this in most cases. However, generally groups that have a carbon attached directly to the ring begin with C or D, groups attached through the nitrogen begin with N, oxygen O, phosphorus and lead P, sulfur S, silicon and selenium s and tin T. Metal fragments use the symbol for the metal in question.

Many compounds have two sort codes, one clockwise and one counterclockwise. For those compounds with two sort codes, the sort code which comes first in alphabetical order was listed. This data base may be requested from the authors.

Statistical computations

The statistical analysis consisted of running a 414 variable multiple regression using eqn. (3) as the model, thus obtaining the 414 SSCS values given in Table 1. The run time for this program was approximately 30 min. The correlation coefficient between the observed and predicted ^{19}F NMR signal positions was $r=0.995$ and the standard deviation was $s=2.46$. The average error of prediction was about 1.6 ppm. A plot of the observed versus predicted values is given in Fig. 1.

The standard estimates of error for the SSCS values are also given in Table 1. They range from low values, *c.* 0.2 ppm for groups with frequent occurrence in the data base to about 2.5 ppm for groups which only rarely appear in the data base. The SSCS values for the rare groups are therefore not quite as reliable as those for common groups. The SCS values for the 20 or so groups given in ref. 6 compare closely to our SSCS values with iodine in DMSO-d_6 as the exception. Iodine is known to exhibit very large solvent effects [1]. An exact statistical comparison is difficult because the values in ref. 6 were determined in either acetone-d_6 or DMSO-d_6, whereas the SSCS values are for an 'average' solvent.

We have used the SSCS values listed in Table 1 to interpret the signals of some previously uninterpreted literature compounds and these are given in Table 2.

TABLE 2. Previously uninterpreted literature data interpreted using the SSCS values listed in Table 1

No.	Sort code	^{19}F NMR signal position		Ref.
		Obs.	Pred.	
1	Fl Fl Cb OX Fl OX	−159.0	−159.3	7
	Fl Cb OX Fl OX Fl	−144.0	−138.8	
	Fl OX Fl Fl Cb OX	−152.0	−152.5	
2	Fl Fl Co OH Fl AA	−147.7	−147.3	8
	Fl Co OH Fl AA Fl	−138.1	−141.4	
	Fl AA Fl Fl Co OH	−138.1	−140.9	
3	Fl Fl OH Fl Fl CT	−144.9	−149.4	9
	Fl OH Fl Fl CT Fl	−162.0	−169.6	
4	Fl Fl Fl Fl AA CH	−146.3	−147.6	10
	Fl Fl Fl AA CH Fl	−157.0[a]	−156.9	
	Fl Fl AA CH Fl Fl	−157.9[a]	−159.0	
	Fl AA CH Fl Fl Fl	−139.8	−140.2	
5	Fl Fl OH Fl Fl CH	−148.6	−147.6	9
	Fl OH Fl Fl CH Fl	−164.0	−163.7	
6	Fl Fl Fl Fl AA Cu	−146.7	−145.5	10
	Fl Fl Fl AA Cu Fl	−157.0	−157.3	
	Fl Fl AA Cu Fl Fl	−158.0	−160.0	
	Fl AA Cu Fl Fl Fl	−139.4	−140.6	
7	Fl Fl Fl Fl AA RF	−140.6	−138.4	10
	Fl Fl Fl AA RF Fl	−156.4	−154.2	
	Fl Fl AA RF Fl Fl	−149.7	−147.7	
	Fl AA RF Fl Fl Fl	−139.3	−137.6	
8	Fl NO NW Fl CF Fl	−141.3	−142.9	11
	Fl CF Fl Fl NO NW	−122.2	−125.7	

[a]May be interchanged.

Experimental

The new data reported here were determined using a Bruker AC250 spectrometer operating at 235.4 MHz using Freon-11 as the external standard and DMSO-d_6 as the solvent. All fluoroarenes were obtained from commercial sources.

Statistical computations were done using programs written in FORTRAN and SAS (Statistical Analysis System) running on a Digital Equipment Corporation Microvax model 3800.

Conclusions

(1) This work both agrees with and extends the work of Fivolt *et al.* [6] by determining large numbers of SSCS values for fluoroarenes. Both the SCS and the SSCS values give excellent prediction capability and can be applied to a large variety of compounds.

(2) Unlike traditional SCS values, it is possible to provide an estimate of error for the SSCS value. As more ^{19}F NMR data for these compounds become available, the SSCS values can presumably be made more precise.

(3) SCS values for several metal moieties have been determined along with a number for groups containing rings.

Acknowledgements

We thank the Auburn University at Montgomery Grant-in-Aid program and the Chemistry and Math-

ematics Departments for support. Thanks are also due to Debra West and Carolyn Johnson (AUM Library) for their assistance in the literature search and to Dr K.P. Sarathy (Auburn University) for NMR spectra.

References

1 J.R. Nanney and C.A.L. Mahaffy, *J. Fluorine Chem., 64* (1993) 217.
2 B. Armstrong, A. Grier, J.B. Hamilton, H. Khuu, C.A.L. Mahaffy, J. Rawlings and J.R. Nanney, *Inorg. Chim. Acta, 191* (1992) 189.
3 J.R. Nanney and C.A.L. Mahaffy, *Inorg. Chim. Acta, 201* (1992) 55.
4 (a) C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani and E.J. Lien, *J. Med. Chem., 16* (1973) 1207; (b) C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology,* Wiley, New York, 1979; (c) C. Hansch, A. Leo and R.W. Taft, *Chem. Rev., 91* (1991) 165.
5 (a) M. Charton, *J. Am. Chem. Soc., 97* (1975) 1552; (b) *ibid., 97* (1975) 3691; (c) *ibid., 97* (1975) 3694; (d) *J. Org. Chem., 41* (1976) 2217; (e) in E. B. Roche (ed.), *Design of Biopharmaceutical Properties through Prodrugs and Analogues,* Am. Pharm. Ass., Washington, DC, 1977.
6 M.J. Fivolt, S.A. Sojka, R.A. Wolfe, D.A. Hohnicki, J.F. Bieron and F.J. Dinan, *J. Org. Chem., 54* (1989) 3019.
7 E.G. Lubenets, T.N. Gerasimova, V.V. Furov and V.A. Backhash, *J. Org. Chem. USSR, 6* (1970) 352.
8 T.D. Petrova, L.I. Kann, V.A. Barkhash and G.G. Yakobson, *Khim. Geterotsikl. Soedin.,* (1969) 778.
9 A.G. Budnik, S.L. Rusakov, E.I. Khomenko and G.G. Yakobson, *J. Org. Chem. USSR, 14* (1978) 1185.
10 T.N. Gerasimova, N.V. Semikolenova and E.P. Fokin, *J. Org. Chem. USSR, 14* (1978) 91.
11 A.I. Rezvukhin and G.G. Furin, *J. Fluorine Chem., 17* (1981) 103.