Towards Theoretical Spectroscopy with Error Bars: Systematic Quantification of the Structural Sensitivity of Calculated Spectra

Tobias G. Bergmann, Michael O. Welzel, Christoph R. Jacob[1]

Technische Universität Braunschweig, Institute of Physical and Theoretical Chemistry,
Gaußstraße 17, 38106 Braunschweig, Germany

Supporting Information

Date: December 4, 2019

[1]E-Mail: c.jacob@tu-braunschweig.de

S1 Comparison of Sensitivity Modes and Normal Modes

Because of the conceptual similarity, it might be instructive to compare the sensitivity mode coordinates obtained here,

$$\boldsymbol{q}_k^{\text{sens}} = \sum_{I\alpha} U_{I\alpha,k}\, \boldsymbol{e}_{I\alpha} \tag{S1}$$

to the well-known normal mode coordinates. In the calculation of vibrational spectra in the harmonic approximation one obtains the normal mode coordinates,

$$\boldsymbol{q}_k^{\text{NM}} = \sum_{I\alpha} L_{I\alpha,k}\, M_I^{-1/2} \boldsymbol{e}_{I\alpha} = \sum_{I\alpha} L_{I\alpha,k}^{(\text{cart})} \boldsymbol{e}_{I\alpha}, \tag{S2}$$

where $L_{I\alpha,k}$ are the components of the k-the eigenvector of the mass-weighted Hessian matrix and M_I is the mass of the I-th nucleus. With respect to not mass-weighted, Cartesian displacements, the normal coordinates are defined by the normal mode vector $\boldsymbol{L}_k^{(\text{cart})}$ with $L_{I\alpha,k}^{(\text{cart})} = M_I^{-1/2} L_{I\alpha,k}$. Note that while both the sensitivity mode vectors $\boldsymbol{U}_k$ (columns of the transformation matrix $\boldsymbol{U}$) and the mass-weighted normal mode vectors $\boldsymbol{L}_k$ form an orthogonal basis, this is not the case for Cartesian normal mode vectors $\boldsymbol{L}_k^{(\text{cart})}$.

For a quantitative comparison, we decompose the sensitivity mode vectors $\boldsymbol{U}_k$ into contributions of the re-normalized Cartesian normal mode vectors, i.e.,

$$\boldsymbol{U}_k = \sum_i c_i^{(k)} \frac{\boldsymbol{L}_i^{(\text{cart})}}{\left|\boldsymbol{L}_i^{(\text{cart})}\right|} \tag{S3}$$

with

$$c_i^{(k)} = \left\langle \frac{\boldsymbol{L}_i^{(\text{cart})}}{\left|\boldsymbol{L}_i^{(\text{cart})}\right|} \middle| \boldsymbol{U}_k \right\rangle = \frac{1}{\left|\boldsymbol{L}_i^{(\text{cart})}\right|} \sum_{I\alpha} L_{I\alpha,i}^{(\text{cart})} U_{I\alpha,k} \tag{S4}$$

The squared coefficients $\left(c_i^{(k)}\right)^2$ indicate the contribution of the i-th Cartesian normal mode vector to the k-th sensitivity mode. Here, $\sum_k \left(c_i^{(k)}\right)^2 = 1$ because of the orthogonality of the sensitivity modes, but $\sum_i \left(c_i^{(k)}\right)^2 < 1$.

Table S1 presents the decomposition of the four most influential sensitivity modes for the XES spectrum of $Fe(CO)_5$ into contributions of the Cartesian normal modes obtained

Table S1: Decomposition of the most influential sensitivity modes for the calculated XES spectrum of $Fe(CO)_5$ into the normalized Cartesian normal modes. For each sensitivity mode, the columns give the squared coefficients $\left(c_i^{(k)}\right)^2$.

			sensitivity modes			
No.	ν/cm^{-1}	symm.	1	2	3	4
1	49	E'_1	0.00	0.00	0.00	0.00
2	49	E'_1	0.00	0.00	0.00	0.00
3	91	E''_1	0.00	0.00	0.00	0.00
4	91	E''_1	0.00	0.00	0.00	0.00
5	95	A''_2	0.00	0.00	0.00	0.00
6	100	E'_1	0.00	0.00	0.00	0.00
7	100	E'_1	0.00	0.00	0.00	0.00
8	352	A'_2	0.00	0.00	0.00	0.00
9	372	E''_1	0.00	0.00	0.00	0.00
10	372	E''_1	0.00	0.00	0.00	0.00
11	430	E'_1	0.00	0.01	0.00	0.00
12	430	E'_1	0.00	0.00	0.00	0.00
13	432	A'_1	0.00	**0.92**	0.03	0.03
14	454	A'_1	0.00	0.05	0.13	**0.81**
15	481	A''_2	0.00	0.00	0.00	0.00
16	488	E'_1	0.00	0.00	0.00	0.00
17	488	E'_1	0.00	0.00	0.00	0.00
18	543	E''_1	0.00	0.00	0.00	0.00
19	543	E''_1	0.00	0.00	0.00	0.00
20	613	A''_2	0.00	0.00	0.00	0.00
21	647	E'_1	0.00	0.00	0.00	0.00
22	647	E'_1	0.00	0.00	0.00	0.00
23	1931	E'_1	0.00	0.00	0.00	0.00
24	1931	E'_1	0.00	0.00	0.00	0.00
25	1960	A''_2	0.00	0.00	0.00	0.00
26	1968	A'_1	0.00	0.01	**0.92**	0.07
27	2080	A'_1	**0.98**	0.01	0.00	0.00

using the same computational methodology. In the case of the XES spectrum of $Fe(CO)_5$, the four most influential sensitivity modes are all totally symmetric. This could possibly be understood by a detailed analysis of the orbital symmetry of the allowed XES transitions in combination with the symmetry of the possible symmetry-adapted displacements.

As the four most influential sensitivity modes are totally symmetric, they only overlap with the totally symmetric normal modes (irreducible representation A_1'). Each sensitivity mode shows a rather large overlap ($c_i^2 > 0.8$) with one the totally symmetric normal modes. Thus, the four most influential sensitivity modes are rather similar to the totally symmetric normal modes. However, this similarity is to a large extent dictated by symmetry requirements.

There seems to be no correlation between the vibrational frequencies (related to the energy increase for a displacement along the normal mode) and the singular values (related to the magnitude of the change in the calculated XES spectrum for a displacement along the sensitivity mode). The most-influential sensitivity mode corresponds to the totally-symmetric normal mode with the highest wavenumber, whereas the second-most influential sensitivity mode corresponds to the totally-symmetric normal mode with the lowest wavenumber.

For the XES spectrum of $Fe(CO)_3$(cod), the decomposition of the sensitivity modes and the normal modes (calculated using the same computational methodology as for the XES spectrum) is presented in Table S2. As there are fewer symmetry requirements, the correspondence between the most influential sensitivity modes and the normal modes is not as clear as for $Fe(CO)_5$. The most influential sensitivity mode shows a large overlap of 0.79 with the symmetric C=O stretching vibration (normal mode 68 at 1953 cm^{-1}). However, the second-most influential sensitivity mode contains significant contributions of at least four normal modes (normal modes 22 at 516 cm^{-1}, 28 at 616 cm^{-1}, 67 at 1869 cm^{-1}, and 68 at 1953 cm^{-1}). Altogether, while there are a number of normal modes

Table S2: Decomposition of the most influential sensitivity modes for the calculated XES spectrum of $Fe(CO)_3$(cod) into the normalized Cartesian normal modes. For each sensitivity mode, the columns give the squared coefficients $\left(c_i^{(k)}\right)^2$. Normal modes that do not overlap with any of the included sensitivity modes are not shown.

		sensitivity modes										
No.	ν/cm^{-1}	1	2	3	4	5	6	7	8	9	10	11
6	98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00
7	112	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01
10	231	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
12	286	0.00	0.01	0.00	0.04	0.00	0.01	0.04	0.00	0.02	0.01	**0.07**
17	404	0.00	0.00	0.01	0.09	0.02	0.08	0.01	0.00	**0.05**	0.00	**0.08**
18	438	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02
20	495	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
21	497	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	**0.05**	0.00	0.01
22	516	0.01	**0.24**	**0.10**	0.11	**0.23**	0.11	0.01	0.05	0.00	0.00	0.00
24	532	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.01	0.05	0.01
25	533	0.00	0.00	0.00	0.07	0.02	0.00	0.00	**0.07**	0.02	**0.11**	0.03
27	596	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00
28	616	0.04	0.10	0.00	0.05	**0.27**	0.00	0.03	0.01	0.01	0.00	0.01
29	640	0.00	0.00	0.06	0.02	0.01	0.00	**0.07**	0.02	0.00	0.07	0.01
33	730	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
35	777	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
38	863	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01
39	875	0.00	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.08	0.09
42	1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.01	0.01
44	1017	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00
47	1155	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.03	0.00
50	1204	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
53	1271	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.00
54	1284	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.00
58	1366	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
63	1443	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.01
64	1457	0.01	0.00	0.03	0.00	0.03	0.00	0.00	0.01	0.03	0.00	0.00
67	1869	0.00	0.10	0.01	**0.21**	0.02	**0.56**	0.00	0.02	0.00	0.02	0.00
68	1953	**0.74**	0.09	**0.09**	0.05	0.00	0.00	0.00	0.00	0.00	0.01	0.00
71	2911	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
75	2997	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
76	2999	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.01	0.02
78	3048	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02
80	3068	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.02	0.06

that to not contribute to any of the sensitivity modes, there is no clear correspondence between normal modes and sensitivity modes anymore.

For the UV/Vis spectrum of aminocoumarin C151 and for the IR spectrum of alanine, the decomposition of the most influential sensitivity modes is shown in Table S3 and S4, respectively. Again, while there is some similarity between the normal modes and the sensitivity modes, no clear correspondence can be found. We also note that the total overlap of the sensitivity modes with the (non-orthogonal) Cartesian normal modes decreases for these two molecules containing hydrogen atoms. Here, the differences between the mass-weighted coordinates used in vibrational spectroscopy and the not mass-weighted coordinates used in our sensitivity analysis becomes more obvious.

Table S3: Decomposition of the most influential sensitivity modes for the calculated UV/Vis spectrum of aminocoumarin C151 into the normalized Cartesian normal modes. For each sensitivity mode, the columns give the squared coefficients $\left(c_i^{(k)}\right)^2$. Normal modes that do not overlap with any of the included sensitivity modes are not shown.

	sensitivity modes					
ν/cm^{-1}	1	2	3	4	5	
14	436	0.00	0.00	0.00	0.00	0.02
16	463	0.00	0.00	0.01	0.00	0.01
19	508	0.00	0.00	0.01	0.00	0.01
25	696	0.00	0.00	0.00	0.00	0.02
28	777	0.00	0.01	0.00	0.00	0.03
30	824	0.00	0.00	0.02	0.00	**0.05**
33	920	0.00	0.00	0.01	0.00	0.04
38	1086	0.04	0.01	0.02	0.00	**0.06**
39	1109	0.00	0.00	0.00	0.02	0.01
42	1176	0.00	0.00	0.00	0.00	0.01
44	1234	0.02	0.00	0.00	0.01	0.00
45	1259	0.00	0.00	0.01	0.00	0.00
46	1322	0.00	0.00	0.01	0.00	0.01
47	1358	0.00	0.00	0.02	0.02	0.01
48	1389	0.05	0.01	**0.04**	0.05	0.00
49	1441	0.00	0.01	0.00	0.03	0.01
50	1507	0.04	0.05	0.00	0.02	0.00
51	1532	0.03	0.11	0.01	**0.19**	0.01
52	1596	0.02	**0.15**	0.01	0.01	0.00
53	1610	0.01	0.00	0.02	0.00	0.00
54	1625	0.02	0.00	**0.04**	0.00	0.04
55	1740	**0.15**	**0.19**	**0.05**	0.04	0.00
60	3478	0.00	0.00	0.01	0.00	0.01

Table S4: Decomposition of the most influential sensitivity modes for the calculated IR spectrum of alanine into the normalized Cartesian normal modes. For each sensitivity mode, the columns give the squared coefficients $\left(c_i^{(k)}\right)^2$. Normal modes that do not overlap with any of the included sensitivity modes are not shown.

		sensitivity modes												
No.	ν/cm^{-1}	1	2	3	4	5	6	7	8	9	10	11	12	13
3	249	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	259	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	319	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	393	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.02	0.00
7	507	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.01	0.01
8	526	0.02	0.00	0.00	0.04	0.01	0.06	0.00	0.00	0.02	0.03	**0.07**	0.02	0.02
9	542	0.02	0.00	0.00	0.00	0.00	**0.29**	0.00	0.00	0.00	0.00	0.03	0.00	0.01
10	582	0.00	0.00	0.00	0.02	0.00	0.13	0.00	0.00	0.01	0.01	0.01	0.00	0.00
11	692	0.00	0.00	0.00	0.00	0.07	0.16	0.00	0.00	0.00	0.00	0.00	0.02	0.03
12	767	0.00	0.00	0.01	0.02	0.07	0.01	0.00	0.01	0.00	0.00	0.00	0.02	0.01
13	934	0.00	0.00	0.00	0.01	0.04	0.01	0.00	0.01	0.01	0.00	0.05	0.00	**0.11**
14	1010	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.01	0.01	0.00	0.00	0.01
15	1019	0.00	0.00	0.00	0.03	0.00	0.05	0.00	0.00	0.00	0.00	0.02	0.00	0.02
16	1102	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.01	0.07	0.00
17	1190	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.03	0.00
18	1249	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.00
19	1308	0.00	0.00	0.00	0.00	0.03	0.06	0.00	0.00	0.00	0.00	0.01	0.01	0.01
20	1317	0.00	0.00	0.00	0.00	0.02	0.08	0.00	0.00	0.00	0.01	0.01	0.00	0.00
21	1391	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.03	0.00	0.01	0.01
22	1416	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
24	1479	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
25	1637	0.02	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.04	0.06	0.02	0.02	0.01
26	1652	0.00	**0.32**	0.02	0.00	0.00	0.01	0.00	0.00	0.03	0.06	0.05	0.00	0.00
27	2895	0.00	0.00	0.00	0.00	0.01	0.00	0.06	**0.24**	0.09	0.04	0.00	**0.15**	0.08
28	2927	0.00	0.00	0.00	0.01	0.00	0.01	**0.60**	0.09	0.00	0.02	0.00	0.01	0.09
29	3008	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	**0.17**	**0.17**	0.00	0.00	0.08
30	3043	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.00
31	3389	0.06	0.02	**0.38**	0.02	**0.10**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
32	3424	**0.55**	0.00	0.08	**0.12**	0.03	0.01	0.00	0.01	0.02	0.00	0.06	0.01	0.02
33	3547	0.00	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.13	0.11	0.10	0.01	0.02

S2 Assessment of reduced-space surrogate models

As discussed in Section 2 of the manuscript, by using a principal component analysis it is possible to identify the sensitivity modes that are most influential within a linearized model. The construction of non-linear surrogate models of the structural sensitivity within the reduced space of the most influential sensitivity modes (see Section 3 of the manuscript) is based on the assumption that the sensitivity modes that are non-influential in the linearized model also only have a small influence when considering the full structural sensitivity.

To verify this assumption, we consider the change in the calculated spectrum for the maximum relevant distortions along each sensitivity mode $\Delta Q_{\max}$, i.e.,

$$\Delta\sigma_k^{(1)}(Q_k = \pm\Delta Q_{\max}). \tag{S5}$$

Within the linearized model, this change in the calculated spectrum is approximated by

$$\Delta\sigma_k^{(1)}(Q_k) \approx \delta\sigma_k^{\mathrm{PC}} \cdot Q_k, \tag{S6}$$

and the norm of $\Delta\sigma_k^{(1)}$ within the linearized model is proportional to the corresponding singular values.

For the calculated XES spectrum of $Fe(CO)_5$, Fig. S1 compares the changes in the calculated spectrum for distortions of $Q_k = \pm 4$ pm along all sensitivity modes. Note that for each considered distortion, an additional quantum-chemical calculation of the spectrum is required. Fig. S1a shows the change for the four most influential sensitivity modes. These are dominated by the linear contribution and thus (except for the scaling of the vertical axis) visually agree with those shown in Fig. 1c. To allow for a quantitative comparison, Table S5 lists the corresponding norms of the changes in the calculated spectrum.

Of the remaining sensitivity modes, the largest changes in the calculated spectrum for distorted structures are found for q_5, q_6, q_{10}, and q_{12}. These are shown in Fig. S1b (note the

different scaling of the vertical axis compared to Fig. S1a). Overall, the observed changes in the calculated spectrum are significantly smaller than for the two most influential sensitivity modes and are comparable to those found for q_4. This is also apparent from the norms of the changes in the calculated spectrum given in Table S5. For these four sensitivity modes, $\left|\Delta\sigma_k^{(1)}(Q_k)\right|$ is between 2.51 and 4.95, compared to 3.97 for the fourth most-influential sensitivity mode q_4 and 38.47 for the most influential sensitivity mode q_1.

The sensitivity modes q_5, q_6, q_{10}, and q_{12}, for which distortions of $Q_i = \pm 4$ pm show the largest changes in the calculated spectrum among those that are non-influential in the linearized models, are visualized in Fig. S1d. Modes q_5, q_6 correspond to changes in the angles between the equatorial CO ligands, mode q_{10} corresponds to a distortion of the angles between the axial CO ligands, and mode q_{12} is an out-of-phase combination of bond-length distortions of the two axial CO ligands. For these distortions, the linear change in the calculated spectrum approximately vanishes because of the molecular symmetry, but there is a quadratic contribution (see also Fig. S1b, where it is obvious that positive and negative distortions result in the same change in the spectrum).

The change in the calculated spectrum for all remaining sensitivity modes (see Fig. S1c and Table S5) is by at least a factor of two smaller than for q_4. Thus, neglecting these sensitivity modes in the construction of a non-linear surrogate model of the structural sensitivity seems justified.

For the further test cases considered here, Tables S6 – S8, list the norms of the change in the calculated spectra for the largest considered distortions along the sensitivity modes. For $Fe(CO)_3(cod)$, we find that for all sensitivity modes that are non-influential in the linearized model, the change in the calculated spectrum is also negligible when explicitly distorting the structures. The largest change of 0.67 is obtained for sensitivity mode q_{13}, compared to 23.22 for the most influential sensitivity mode and 1.15 for q_{12}. The norm of the changes in the spectrum mostly correlate with the magnitude of the singular

Figure S1: Change in the calculated XES spectrum of $Fe(CO)_5$ for distortions of $Q_i = \pm 4$ pm along the different sensitivity modes (solid lines correspond to distortions of $Q_i = +4$ pm, dotted lines to distortions of $Q_i = -4$ pm), (a) of the four sensitivity modes that are most influential in the linearized model, (b) for the remaining four sensitivity that give the largest change in the calculated spectrum, and (c) for all remaining sensitivity modes. (d) Sensitivity modes q_5, q_6, q_{10}, and q_{12}.

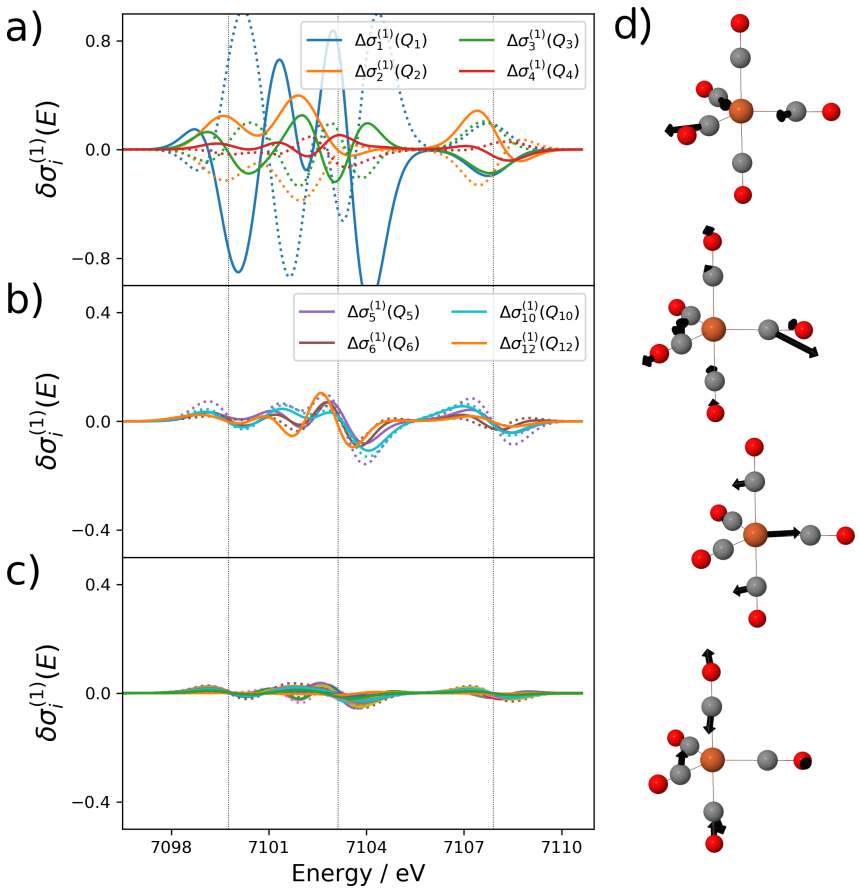


Table S5: Maximum norms of the change in calculated XES spectrum of $Fe(CO)_5$ for distortions of $Q_i = \pm 4$ pm along the different sensitivity modes, both within a linearized model and when explicitly distorting the structure.

k	$\left\|\delta\sigma_k^{\mathrm{PC}} \cdot Q_k\right\|$	$\left\|\Delta\sigma_k^{(1)}(Q_k)\right\|$
1	37.75	38.47
2	14.30	14.00
3	10.14	10.21
4	3.50	3.97
5	0.90	4.95
6	0.14	2.51
10	0.05	4.10
12	0.02	2.88
max. remaining	0.11	2.00
min. remaining	0.00	0.20

value, i.e., the change obtained in the linearized model. Only for a few sensitivity modes (e.g., q_{47}, q_{48}, q_{49}, q_{50}, and q_{52}) there is a small change in the spectrum even though the corresponding singular value is zero.

Similarly, for the calculated UV/Vis spectrum of aminocoumarin C151 (see Table S7), the norms of the changes in the spectrum obtained for structures distorted along the sensitivity modes with $Q_i = \pm 1$ pm correlate with the magnitudes of the corresponding singular values. Thus, the change in the calculated spectrum is small also for these larger distortions for all sensitivity modes that are non-influential in the linearized model.

Finally, for the calculated IR spectrum of alanine (see Tables S8, there are larger differences between the norms of the changes in the spectrum obtained within the linearized model and those obtained for structures distorted by $Q_i = \pm 0.5$ pm along the sensitivity modes. This is in line with the observation that for this test case the higher-order contributions in a Taylor expansion of the one-mode contributions of the full structural sensitivity become more important. Nevertheless, for all sensitivity modes that are ne-

Table S6: Maximum norms of the change in calculated XES spectrum of $Fe(CO)_3$(cod) for distortions of $Q_i = \pm 4$ pm along the different sensitivity modes, both within a linearized model and when explicitly distorting the structure.

k	$\left\|\delta\sigma_k^{\mathrm{PC}} \cdot Q_k\right\|$	$\left\|\Delta\sigma_k^{(1)}(Q_k)\right\|$
1	22.05	23.22
2	13.94	14.22
3	12.30	12.84
4	7.58	7.72
5	5.74	5.86
6	4.95	5.78
7	4.01	4.00
8	3.05	3.18
9	2.56	2.65
10	1.90	1.89
11	1.78	1.79
12	1.12	1.15
13	0.63	0.67
14	0.46	0.54
15	0.38	0.43
47	0.00	0.42
48	0.00	0.49
49	0.00	0.47
50	0.00	0.61
52	0.00	0.40
max. remaining	0.22	0.38
min. remaining	0.00	0.04

Table S7: Maximum norms of the change in calculated UV/Vis spectrum of aminocoumarin C151 for distortions of $Q_i = \pm 1$ pm along the different sensitivity modes, both within a linearized model and when explicitly distorting the structure.

k	$\left\lvert \delta\sigma_k^{\mathrm{PC}} \cdot Q_k \right\rvert$	$\left\lvert \Delta\sigma_k^{(1)}(Q_k) \right\rvert$
1	46.47	40.46
2	15.14	14.01
3	7.49	7.29
4	3.62	3.80
5	2.81	2.76
6	0.92	0.93
7	0.61	0.60
8	0.51	0.67
max. remaining	0.18	0.24
min. remaining	0.00	0.05

glected in our surrogate model, the change in the spectrum is significantly smaller than for the most influential sensitivity modes. The largest change of 5.24 is found for q_{18}, compared to 110.56 for the most influential sensitivity mode and 3.80 for q_{13}.

Altogether, the additional tests presented here confirm the assumption that the sensitivity modes that are most influential in a linearized model are also the most influential ones when considering the full structural dependence. However, in some cases including additional modes that have a large quadratic (or higher-order) contributions might become necessary, in particular in cases where symmetry breaking leads to large changes in the spectrum. Efficient strategies to this end that do not require calculations along all sensitivity modes corresponding to small singular values, will be explored in our future work.

Table S8: Maximum norms of the change in calculated infrared spectrum of alanine for distortions of $Q_i = \pm 0.5$ pm along the different sensitivity modes, both within a linearized model and when explicitly distorting the structure.

k	$\left\|\delta\sigma_k^{\mathrm{PC}} \cdot Q_k\right\|$	$\left\|\Delta\sigma_k^{(1)}(Q_k)\right\|$
1	80.23	110.56
2	70.24	75.73
3	54.32	81.14
4	18.22	26.53
5	14.94	22.09
6	10.68	10.74
7	7.71	8.81
8	4.42	6.41
9	4.15	14.68
10	3.93	10.60
11	3.21	12.14
12	2.88	3.83
13	2.60	3.80
14	2.21	4.34
15	1.98	4.56
16	1.52	4.38
17	1.24	3.63
18	1.04	5.24
24	0.39	3.51
25	0.38	3.24
26	0.35	3.69
28	0.29	3.14
32	0.14	3.08
34	0.13	3.15
max. remaining	0.75	2.76
min. remaining	0.04	0.38

S3 Calculation of Variance for Surrogate Models Including Two-Mode Contributions

For the surrogate mode given in Eq. (16) with two-mode contributions the straightforward calculation of the variance would become rather intricate and possibly inefficient. A more convenient approach is to convert our Cut-HDMR expansion into a random sampling (RS) HDMR expansion [1, 2]

$$\sigma(E;\boldsymbol{R}) = \big\langle \sigma(E) \big\rangle + \sum_{k=1}^{k_{\max}} \Delta\tilde{\sigma}_k^{(1)}(E;Q_k) + \sum_{k<l}^{k_{\max}} \Delta\tilde{\sigma}_{kl}^{(2)}(E;Q_k,Q_l) + \cdots . \tag{S7}$$

Here, the tilde is used to distinguish the different terms from those in the Cut-HDMR expansion of Eq. (16). The one-mode contributions in the RS-HDMR expansion are given by

$$\begin{aligned}
\Delta\tilde{\sigma}_k^{(1)}(E;Q_k) &= \big\langle \sigma(E;\boldsymbol{R}) \big\rangle_k - \big\langle \sigma(E) \big\rangle \\
&= \Delta\sigma_k^{(1)}(E;Q_k) + \sum_{l\neq k} \int \Delta\sigma_{lk}^{(2)}(E;Q_l,Q_k) p_l(Q_l)\,\mathrm{d}Q_l \\
&\quad - \Big\langle \Delta\sigma_k^{(1)} \Big\rangle - \sum_{l\neq k} \Big\langle \Delta\sigma_{kl}^{(2)} \Big\rangle
\end{aligned} \tag{S8}$$

and the two-mode contributions by

$$\begin{aligned}
\Delta\tilde{\sigma}_{kl}^{(2)}(E;Q_k,Q_l) &= \big\langle \sigma(E;\boldsymbol{R}) \big\rangle_{kl} - \Delta\tilde{\sigma}_k^{(1)}(E;Q_k) - \Delta\tilde{\sigma}_l^{(1)}(E;Q_l) - \big\langle \sigma(E) \big\rangle \\
&= \Delta\sigma_{kl}^{(2)}(E;Q_k,Q_l) - \int \Delta\sigma_{kl}^{(2)}(E;Q_k,Q_l) p_l(Q_l)\,\mathrm{d}Q_l \\
&\quad - \int \Delta\sigma_{kl}^{(2)}(E;Q_k,Q_l) p_k(Q_k)\,\mathrm{d}Q_k + \Big\langle \Delta\sigma_{kl}^{(2)} \Big\rangle .
\end{aligned} \tag{S9}$$

Here, $\big\langle \sigma(E;\boldsymbol{R}) \big\rangle_k$ indicates that the mean is taken over all coordinates but Q_k and $\big\langle \sigma(E;\boldsymbol{R}) \big\rangle_{kl}$ is the mean taken over all coordinates by Q_k and Q_l.

With the RS-HDMR expansion, the variance can now be calculated as [2–4]

$$\mathrm{Var}\big[\sigma(E)\big] = \sum_{k=1}^{k_{\max}} \mathrm{Var}\Big[\Delta\tilde{\sigma}_k^{(1)}\Big] + \sum_{k<l}^{k_{\max}} \mathrm{Var}\Big[\Delta\tilde{\sigma}_{kl}^{(2)}\Big] \tag{S10}$$

with

$$\mathrm{Var}\left[\Delta\tilde{\sigma}_k^{(1)}\right] = \int \left(\Delta\tilde{\sigma}_k^{(1)}(Q_k)\right)^2 p_k(Q_k)\,\mathrm{d}Q_k \tag{S11}$$

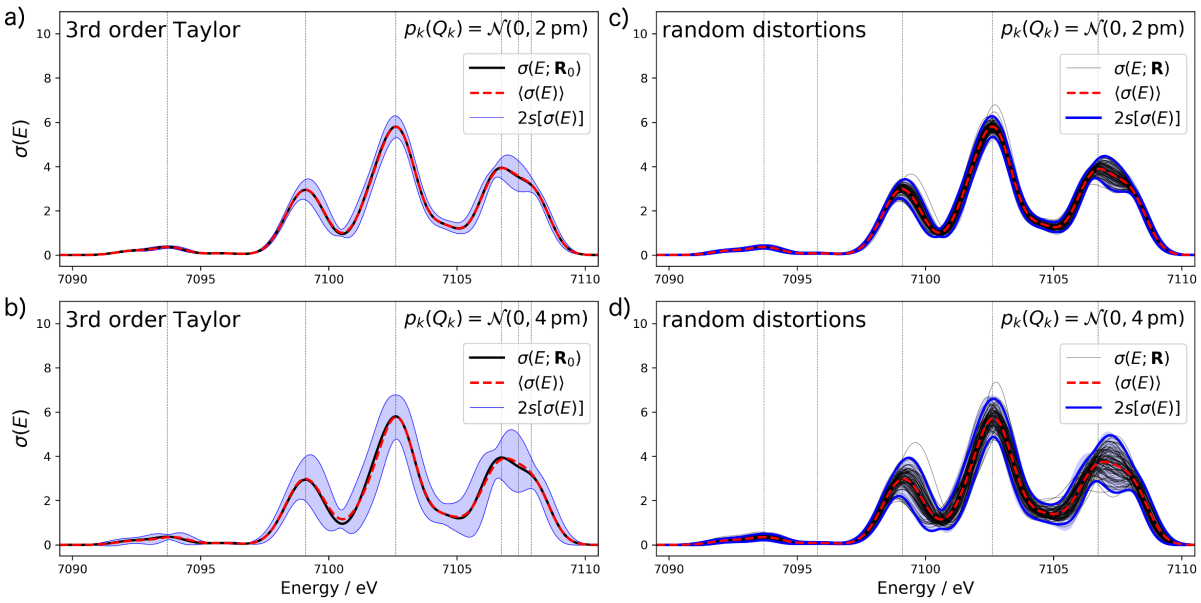
and

$$\mathrm{Var}\left[\Delta\tilde{\sigma}_{kl}^{(2)}\right] = \int \left(\Delta\tilde{\sigma}_{kl}^{(2)}(Q_k, Q_l)\right)^2 p_k(Q_k) p_k(Q_l)\,\mathrm{d}Q_k \mathrm{d}Q_l \tag{S12}$$

S4 Comparison with Random Distortions

To assess the accuracy of the error bars obtained from the non-linear surrogate models, we compare to the spectra calculated for 100 random distortions. The results are included and discussed in the manuscript for XES of $Fe(CO)_5$, the UV/Vis spectrum of aminocoumarin C151, and the IR spectrum of alanine. For completeness, a similar comparison for the XES spectrum of $Fe(CO)_3$(cod) is shown in Fig. S2.

Figure S2: Analysis of the structural sensitivity of the calculated XES spectrum of $Fe(CO)_3$(cod). (a,b) calculated spectrum including error bars corresponding to two standard deviations when assuming a normal distribution with standard deviation (a) $s_Q = 2$ pm and (b) $s_Q = 4$ pm for the distortions of the molecular structure. The error bars are obtained using the non-linear surrogate model based on a 3rd order Taylor expansion for the one-mode contributions and neglecting two-mode and higher-order contributions (reproduced from Fig. 5c,d). (c,d) Spectra calculated for 100 random distortions sampled from independent normal distributions with (d) $s_Q = 2$ pm and (e) $s_Q = 4$ pm (black lines) as well as the error bars corresponding to two standard deviations (blue lines). For comparison, the error bars from (a) and (b), respectively, are included as blue shaded area.



S5 Quantitative Statistical Metrics

To further analyze the uncertainty in the calculated spectra, we consider quantitative statistical metrics for the intensity at relevant energies, E_j (see also Tables I and II in the manuscript).

Results for $Fe(CO)_3$(cod) are collected in Tab. S9. Overall, the quantitative statistical metrics confirm the observations already expected from the visual inspection of the spectra with error bars. The smallest absolute uncertainty is found for the low-energy feature at 7093.7 eV, while significantly larger absolute uncertainties of comparable magnitude are found for the remaining peaks. The largest absolute uncertainty is found at the minimum in between the two high-energy peaks at 7106.7 eV and 7107.9 eV. On the other hand, the relative uncertainty as measured by the COV is similar for all peaks, a larger COV is only found at the minimum between the two high-energy peaks. As for the XES spectrum of $Fe(CO)_5$, when increasing the standard deviation of the normal distribution that is assumed for the structural distortions by a factor of two, both the standard deviation and the COV roughly double for all considered peaks.

Quantitative statistical metrics referring to the intensities at the maxima of the peaks in the calculated UV/Vis spectrum of aminocoumarin C151 are collected in Tab. S10. The largest absolute uncertainty is found for the first peak, which also has the highest intensity, while the lowest absolute uncertainty is found for the third peak, which also has the lowest intensity. When looking at the relative uncertainties, the COVs of the first, second, and fourth peak are roughly comparable, whereas a significantly smaller COV is found for the third peak. However, we note that the metrics collected in the Table only refer to the intensities at the peak maxima, and thus do not reflect the different uncertainties in the position of the peaks.

For the calculated IR spectrum of alanine, the quantitative statistical metrics for the

Table S9: Quantitative statistical metrics for the uncertainty of the calculated XES spectrum of $Fe(CO)_3$(cod) at the maxima of the peaks (E_j, indicated by vertical lines in the spectra in Fig. 5) in the considered region of the spectrum. In addition, the minimum between the two highest-energy peaks is also considered. The statistical analysis assumes a normal distribution with standard deviation $s_Q = 2$ pm and $s_Q = 4$ pm for the distortions of the underlying molecular structure. Listed are the intensity for the undistorted structure $\sigma(E_j; \boldsymbol{R}_0)$, the mean of the intensity $\langle\sigma(E_j)\rangle$, its variance $\mathrm{Var}\big[\sigma(E_j)\big]$, its standard deviation $s\big[\sigma(E_j)\big]$, and the coefficient of variance $\mathrm{COV}\big[\sigma(E_j)\big]$. All metrics refer to the non-linear surrogate model based on a 3rd order Taylor expansion for the one-mode contributions and neglecting two-mode and higher-order contributions.

E_j/eV		$\sigma(E_j; \boldsymbol{R}_0)$	$\langle\sigma(E_j)\rangle$	$\mathrm{Var}\big[\sigma(E_j)\big]$	$s\big[\sigma(E_j)\big]$	$\mathrm{COV}\big[\sigma(E_j)\big]$
$p_k(Q_k) = \mathcal{N}(0, 2\text{ pm})$						
7093.7	max.	0.37	0.36	0.00	0.03	0.08
7099.1	max.	2.95	2.95	0.05	0.23	0.08
7102.6	max.	5.80	5.80	0.06	0.25	0.04
7106.7	max.	3.94	3.92	0.06	0.24	0.06
7107.4	min.	3.53	3.57	0.12	0.34	0.10
7107.9	max.	3.16	3.18	0.03	0.17	0.05
$p_k(Q_k) = \mathcal{N}(0, 4\text{ pm})$						
7093.7	max.	0.37	0.34	0.00	0.06	0.18
7099.1	max.	2.95	2.98	0.25	0.50	0.17
7102.6	max.	5.80	5.77	0.25	0.50	0.09
7106.7	max.	3.94	3.87	0.21	0.46	0.12
7107.4	min.	3.53	3.69	0.43	0.65	0.18
7107.9	max.	3.16	3.23	0.12	0.34	0.11

Table S10: Quantitative statistical metrics for the uncertainty of the calculated UV/Vis spectrum of aminocoumarin C151 at the maxima of the four considered peaks (E_j, indicated by vertical lines in the spectra in Fig. 6) assuming a normal distribution with standard deviation $s_Q = 0.5$ pm for the distortions of the underlying molecular structure. Listed are the intensity for the undistorted structure $\sigma(E_j; \boldsymbol{R}_0)$, the mean of the intensity $\langle\sigma(E_j)\rangle$, its variance $\mathrm{Var}\big[\sigma(E_j)\big]$, its standard deviation $s\big[\sigma(E_j)\big]$, and the coefficient of variance $\mathrm{COV}\big[\sigma(E_j)\big]$. All metrics refer to the non-linear surrogate model based on a 4th order Taylor expansion for the one-mode contributions and neglecting two-mode and higher-order contributions.

E_j/eV	$\sigma(E_j; \boldsymbol{R}_0)$	$\langle\sigma(E_j)\rangle$	$\mathrm{Var}\big[\sigma(E_j)\big]$	$s\big[\sigma(E_j)\big]$	$\mathrm{COV}\big[\sigma(E_j)\big]$
3.2	2.35	1.89	0.48	0.69	0.37
3.6	0.60	0.40	0.03	0.18	0.45
4.5	0.23	0.19	0.00	0.02	0.13
4.8	0.84	0.64	0.04	0.19	0.30

intensities at the peak maxima when assuming a normal distribution with a standard deviation of $s_Q = 0.25$ pm for the structural distortions are shown in Table II and are discussed in the manuscript. For completeness, in Table S11, these quantitative statistical metrics are also given for the case that a standard deviation of $s_Q = 0.5$ pm is assumed (cf. also Fig. 7f).

When increasing the standard deviation from $s_Q = 0.25$ pm to $s_Q = 0.5$ pm, the absolute and relative uncertainties for the intensities of the low intensity peaks roughly double, as for the other test cases considered above. However, a significantly larger, nonlinear increase is found for the peaks with a larger intensity. Because the employed approximation of the one-mode contributions breaks down for $s_Q = 0.5$ pm, a further discussion in not warranted here.

Table S11: Quantitative statistical metrics for the uncertainty of the calculated IR spectrum of alanine at the maxima of selected peaks (E_j, indicated by vertical lines in the spectra in Fig. 7) assuming a normal distribution with standard deviation $s_Q = 0.25$ pm and $s_Q = 0.5$ pm for the distortions of the underlying molecular structure. Listed are the intensity for the undistorted structure $\sigma(E_j; \boldsymbol{R}_0)$, the mean of the intensity $\langle\sigma(E_j)\rangle$, its variance $\mathrm{Var}\big[\sigma(E_j)\big]$, its standard deviation $s\big[\sigma(E_j)\big]$, and the coefficient of variance $\mathrm{COV}\big[\sigma(E_j)\big]$. All metrics refer to the non-linear surrogate model based on a 4th order Taylor expansion for the one-mode contributions and neglecting two-mode and higher-order contributions.

E_j/cm^{-1}	assignment	$\sigma(E_j; \boldsymbol{R}_0)$	$\langle\sigma(E_j)\rangle$	$\mathrm{Var}\big[\sigma(E_j)\big]$	$s\big[\sigma(E_j)\big]$	$\mathrm{COV}\big[\sigma(E_j)\big]$
$p_k(Q_k) = \mathcal{N}(0, 0.25\ \mathrm{pm})$						
534.7	fingerprint	11.12	8.66	8.77	2.96	0.34
1015.6	X–H bend	7.70	7.22	0.48	0.69	0.10
1191.3	C^{α}–N stretch	0.88	0.86	0.00	0.04	0.05
1313.1	C^{α}–H bend	0.99	0.96	0.02	0.15	0.16
1412.5	symm. CH_3 bend	1.08	1.07	0.00	0.06	0.06
1650.9	C=O stretch	8.31	7.18	3.52	1.88	0.26
2922.2	C–H stretch	0.79	0.66	0.05	0.22	0.34
3014.7	C–H stretch	0.59	0.54	0.02	0.13	0.24
3390.2	O–H stretch	2.07	1.10	0.88	0.94	0.85
3547.1	N–H stretch	0.58	0.37	0.05	0.23	0.62
$p_k(Q_k) = \mathcal{N}(0, 0.5\ \mathrm{pm})$						
534.7	fingerprint	11.12	5.19	74.32	8.62	1.66
1015.6	X–H bend	7.70	5.98	4.20	2.05	0.34
1191.3	C^{α}–N stretch	0.88	0.79	0.01	0.10	0.13
1313.1	C^{α}–H bend	0.99	0.92	0.09	0.29	0.32
1412.5	symm. CH_3 bend	1.08	1.04	0.02	0.13	0.12
1650.9	C=O stretch	8.31	6.07	30.75	5.54	0.91
2922.2	C–H stretch	0.79	0.45	0.35	0.59	1.32
3014.7	C–H stretch	0.59	0.43	0.18	0.42	0.97
3390.2	O–H stretch	2.07	0.08	9.67	3.11	36.98
3547.1	N–H stretch	0.58	0.21	1.24	1.11	5.35

References

[1] G. Li, C. Rosenthal, and H. Rabitz, *J. Phys. Chem. A*, 2001, **105**, 7765–7777.

[2] G. Li, S.-W. Wang, H. Rabitz, S. Wang, and P. Jaffé, *Chem. Eng. Sci.*, 2002, **57**, 4445–4460.

[3] I. M. Sobol, *Math. Modeling Comput. Experiment*, 1995, **1**, 407–414.

[4] H. Rabitz and O. F. Aliş, *J. Math. Chem.*, 1999, **25**, 197–233.