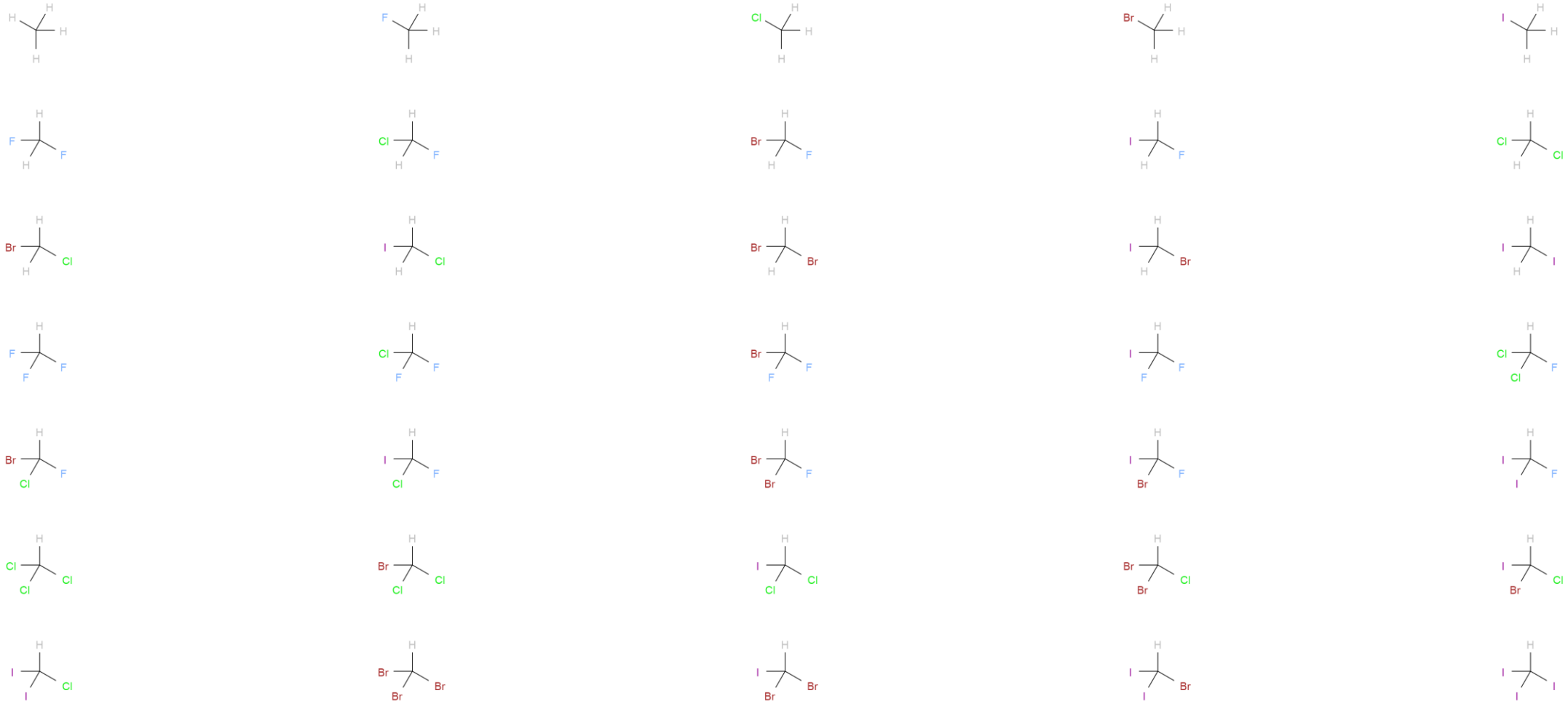
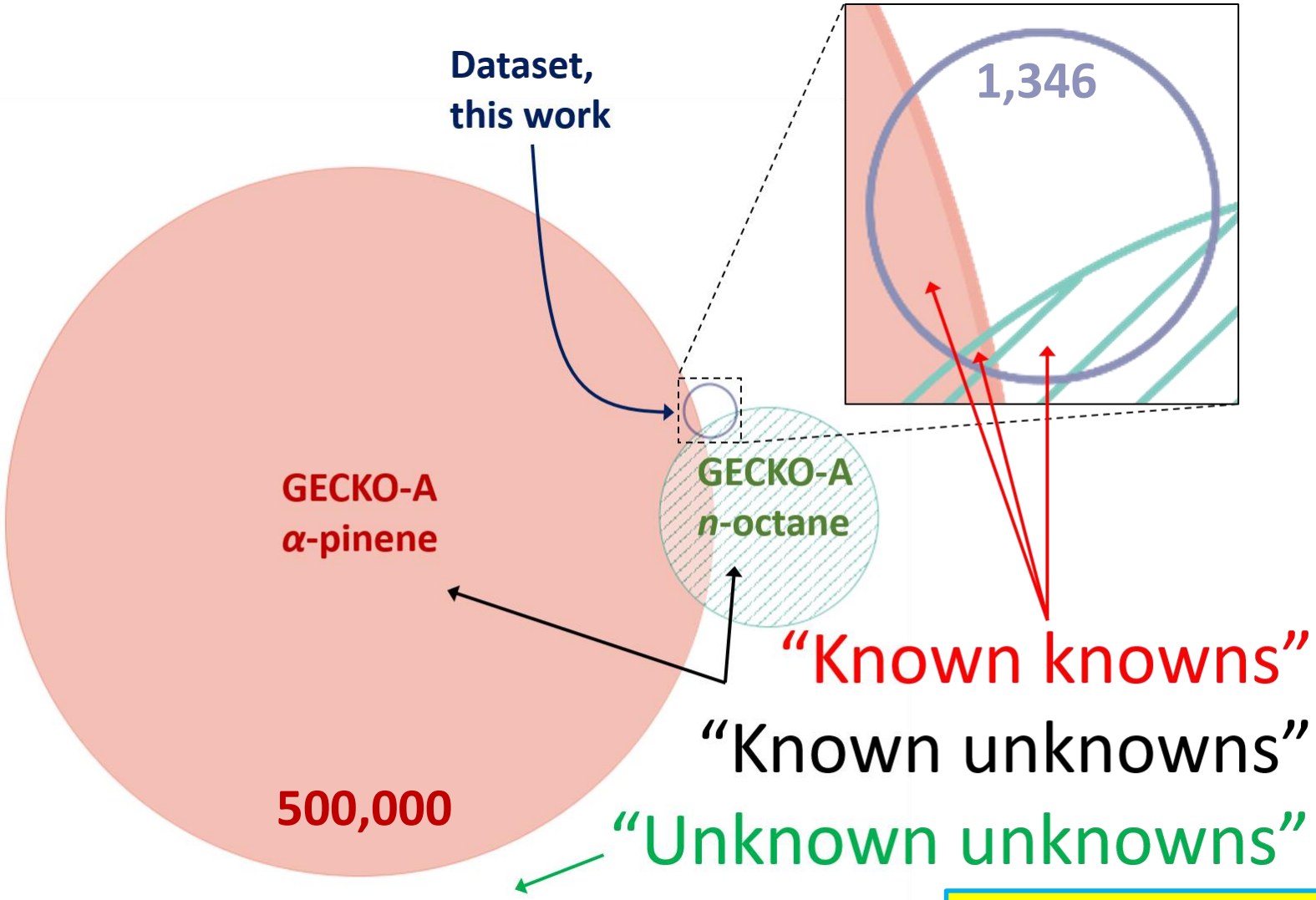
A Broad View of Structure-Activity Relationship Performance for Alkanes and Haloalkanes



The 26th International Symposium on Gas Kinetics and Related Phenomena, Rennes, 2022

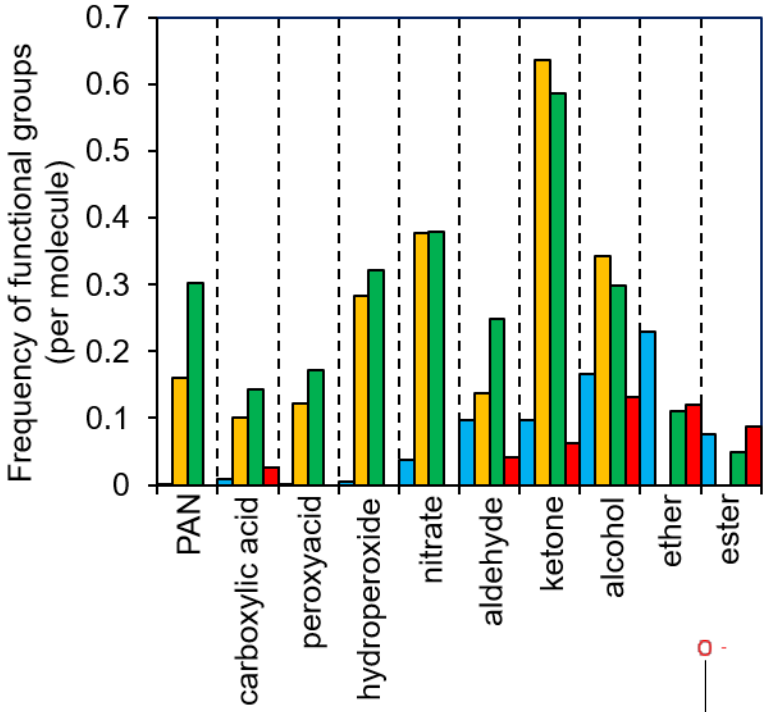
Max McGillen (CNRS-Orléans, France)
✉ max.mcgillen@cnrs-orleans.fr

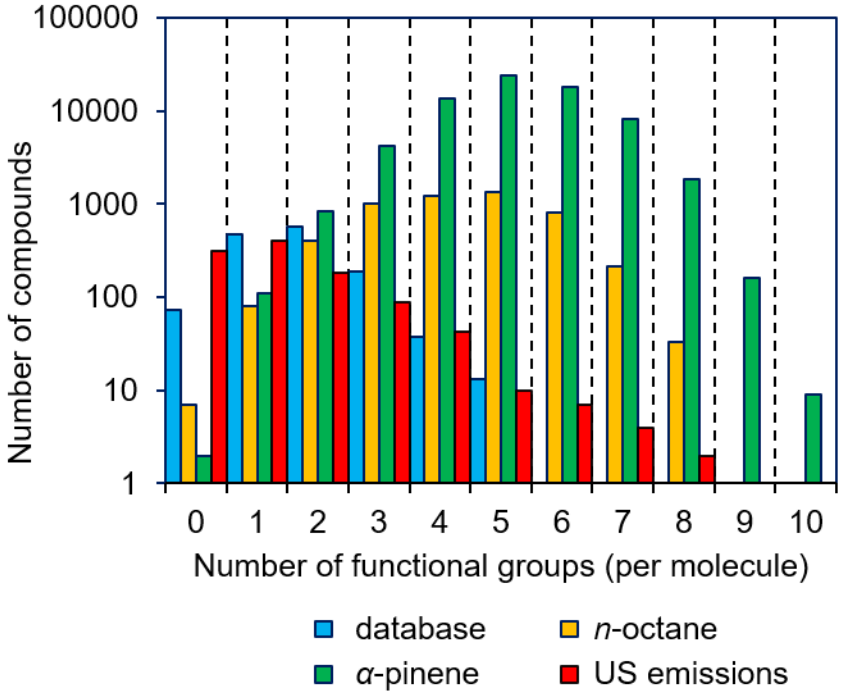
The combinatorial explosion – A justification for SARs –

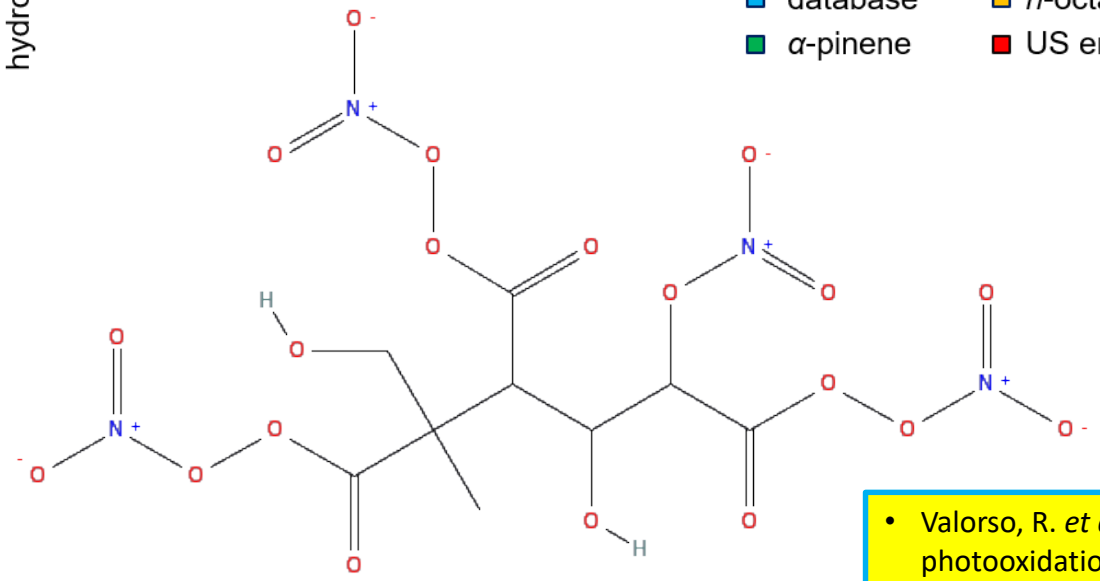


- The atmosphere is a complicated place
- Our best view of complexity comes from automated mechanism generation software (AMG)
- AMG produces a large number of molecules
- **About which we don't know very much…**

McGillen, M. R. *et al.* Database for the kinetics of the gas-phase atmospheric reactions of organic compounds. *Earth System Science Data* **12**, 1203–1216 (2020).







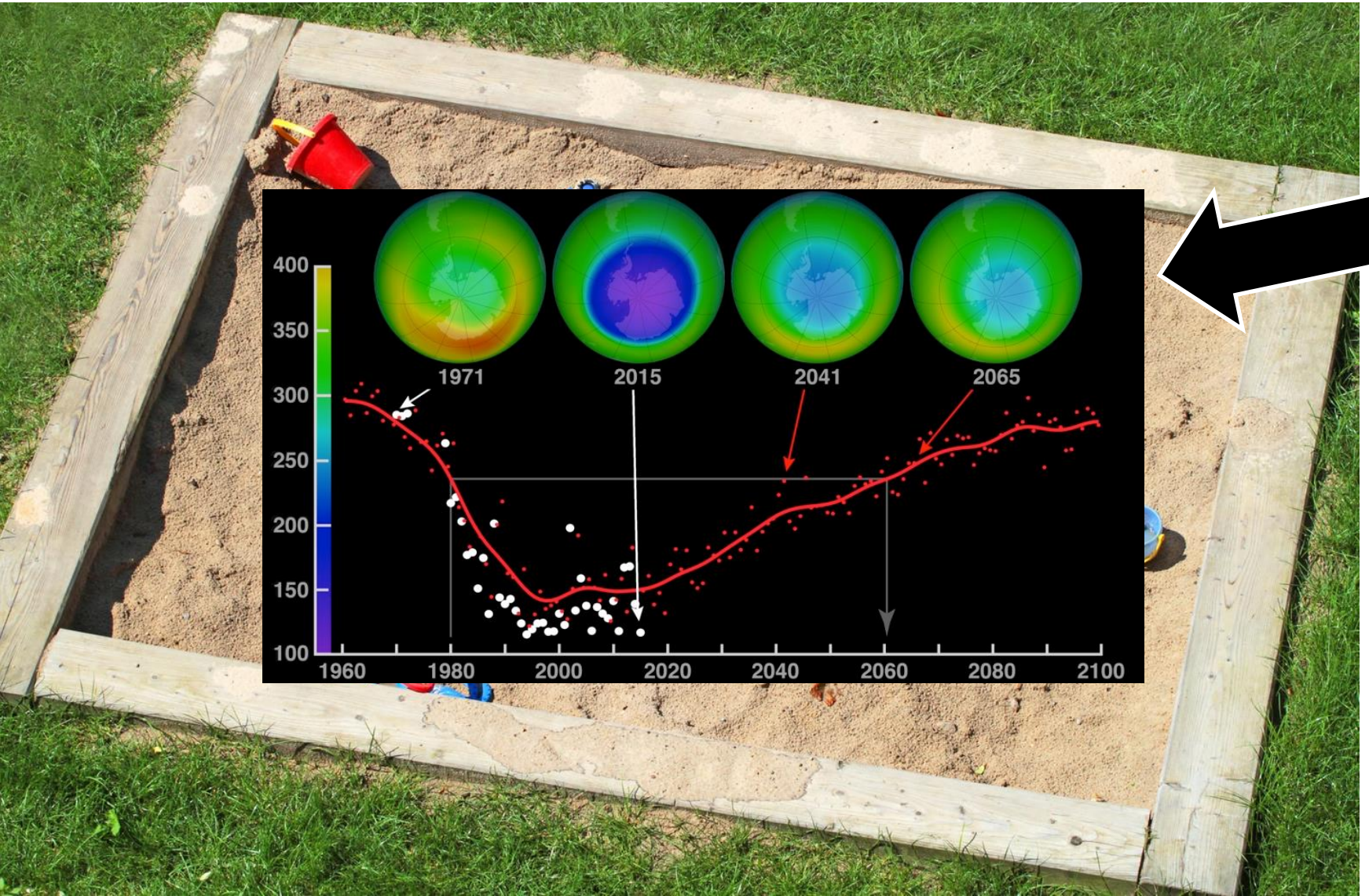
- A high degree of multifunctionality
- Especially for functional groups that aren't well studied
- Prediction of these rate coefficients is surely a gross extrapolation
- **Clearly this is too much for now…**

- Valorso, R. *et al.* Explicit modelling of SOA formation from α-pinene photooxidation: sensitivity to vapour pressure estimation. *Atmos. Chem. Phys.* **11**, 6895–6910 (2011).
- McGillen, M. R. *et al.* Database for the kinetics of the gas-phase atmospheric reactions of organic compounds. *Earth System Science Data* **12**, 1203–1216 (2020).

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons

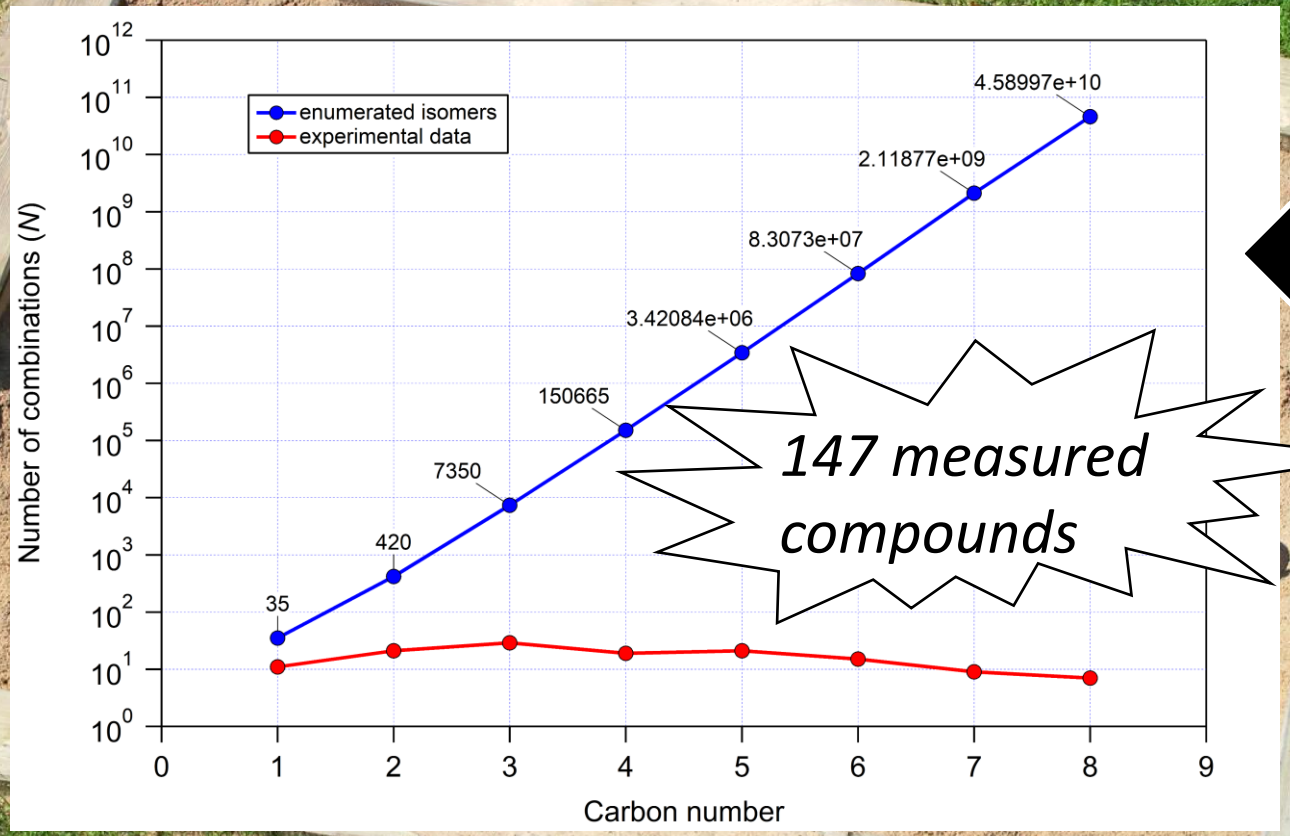
- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information
- Highly functionalized
- Different methods for estimation

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons



- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information
- Highly functionalized
- Different methods for estimation

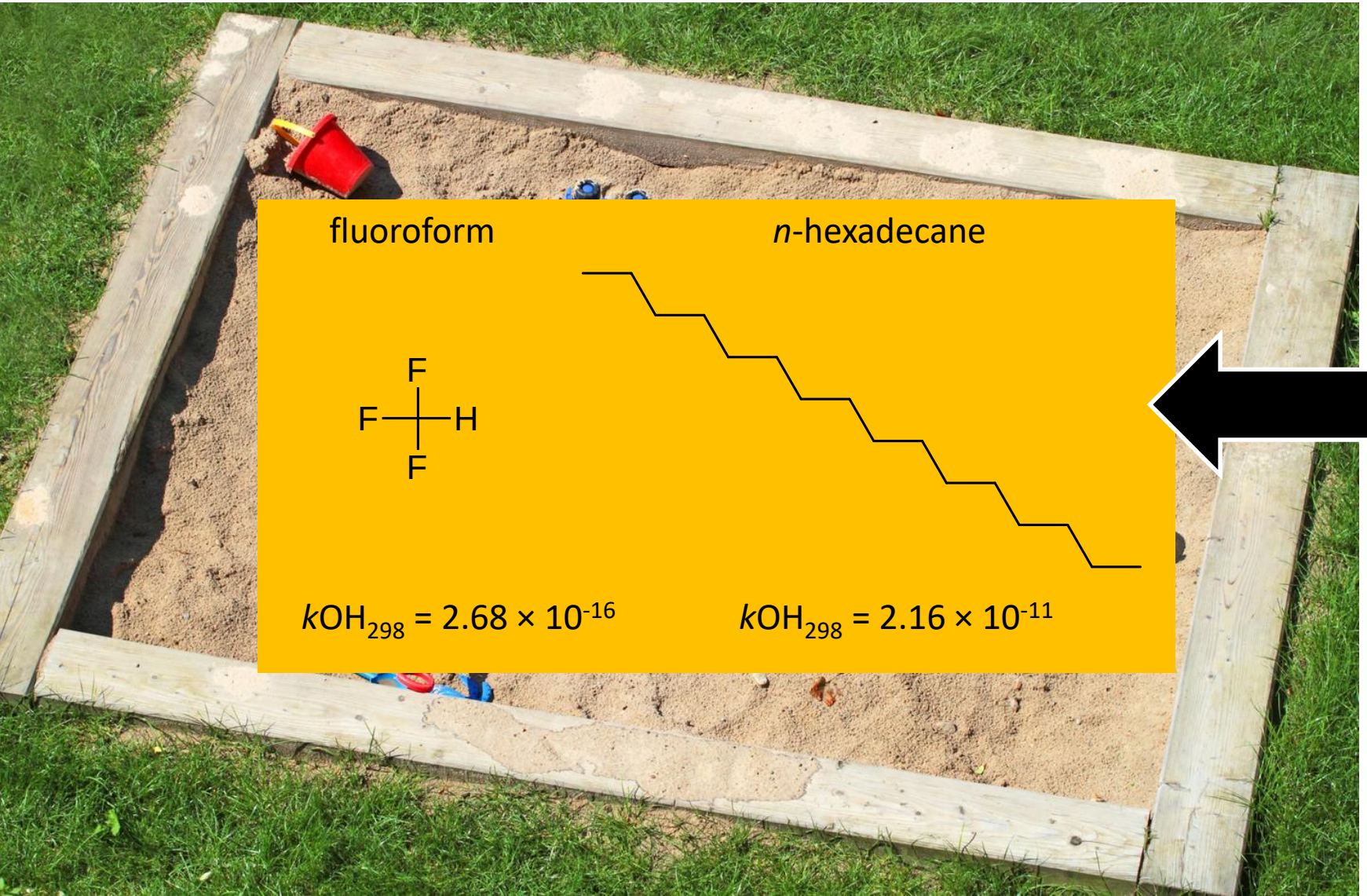
The Sandbox: a unique testing environment containing hydrocarbons and halocarbons



McGillen, M. R. *et al.* Database for the kinetics of the gas-phase atmospheric reactions of organic compounds. *Earth System Science Data* **12**, 1203–1216 (2020).

- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information
- Highly functionalized
- Different methods for estimation

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons



- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage

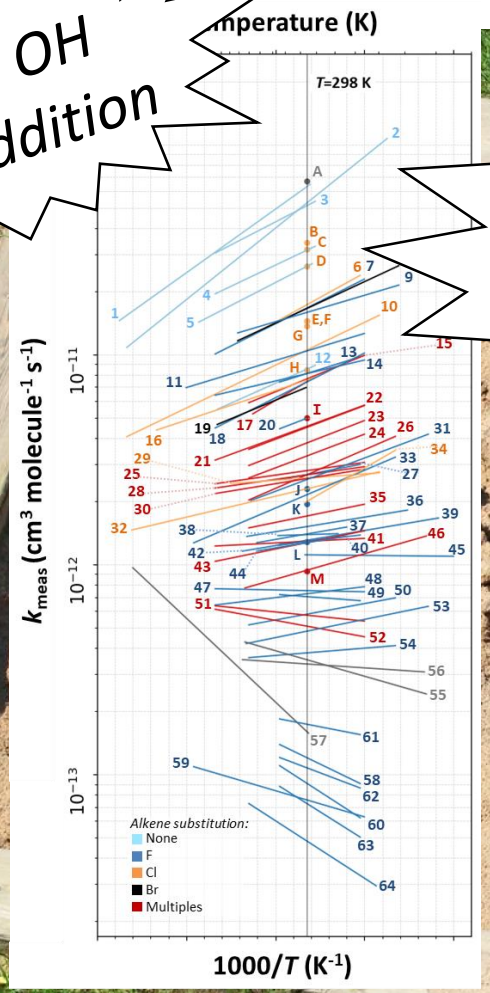
Variation in reactivity

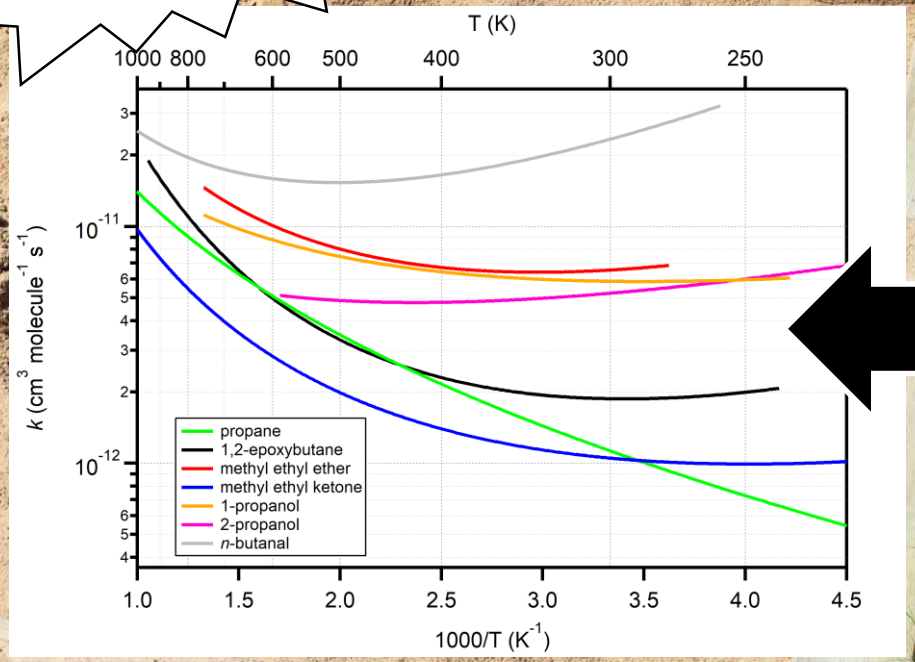
- Mechanistically simpler
- Complementary information
- Highly functionalized
- Different methods for estimation

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons

No OH addition

No H-bonded complexes

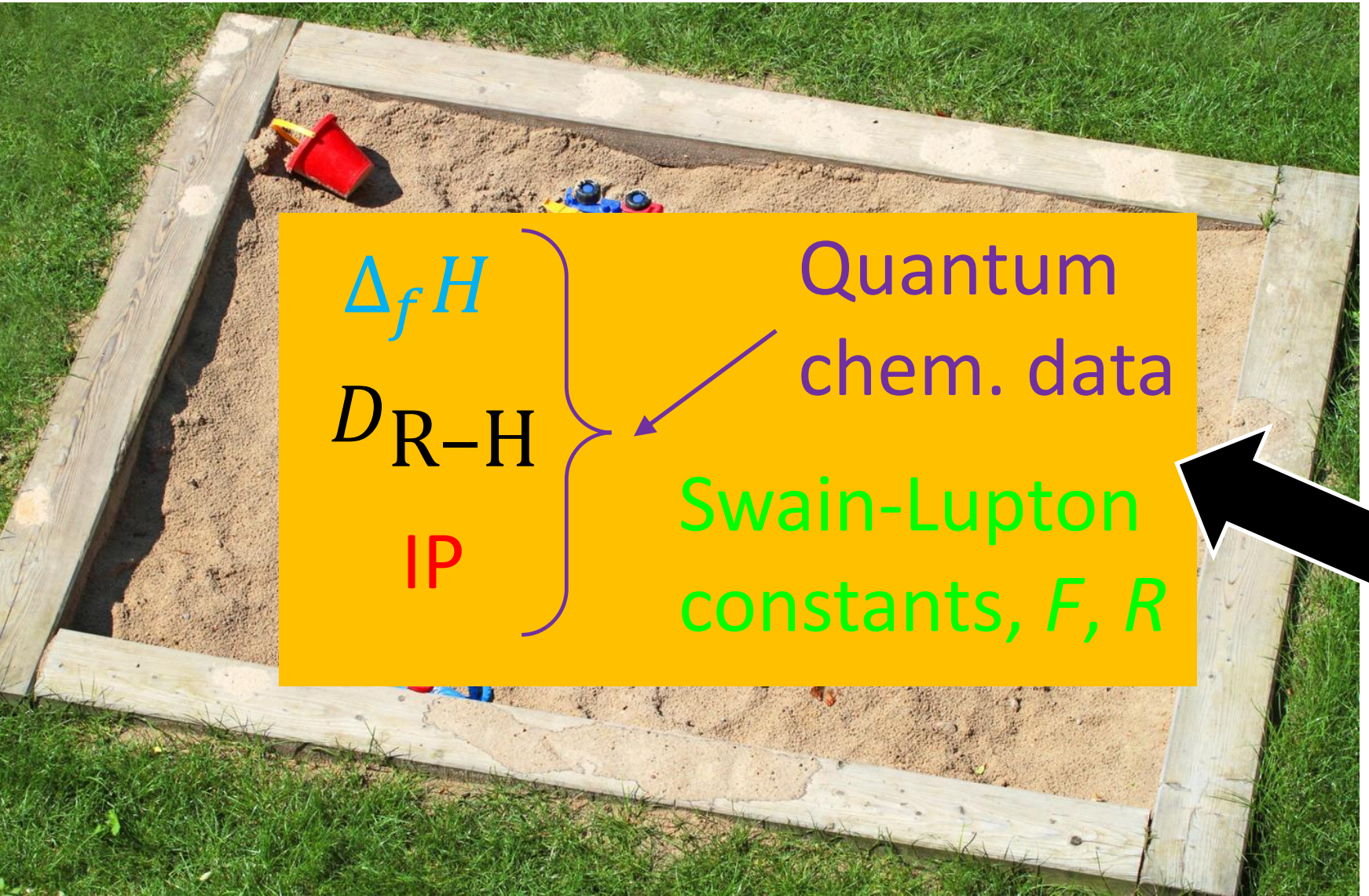




- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information
- Highly functionalized
- Different methods for estimation

- Michelat et al. *in preparation*.
- El Othmani, H. *et al.* Gas-Phase Rate Coefficient of OH + 1,2-Epoxybutane Determined between 220 and 950 K. *ACS Earth Space Chem.* **5**, 960–968 (2021).

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons

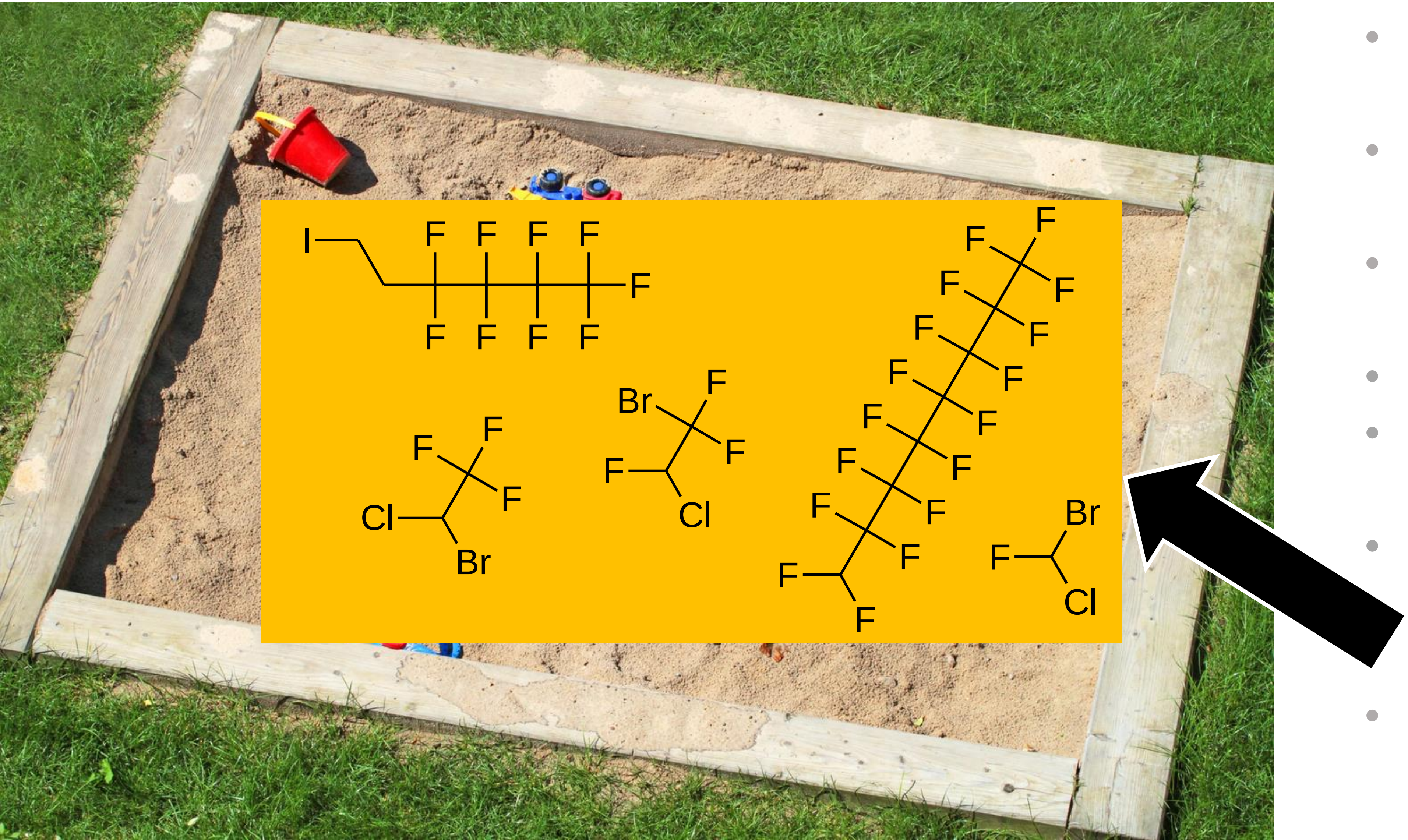


- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler

Complementary information

- Highly functionalized
- Different methods for estimation

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons



- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information

Highly functionalized

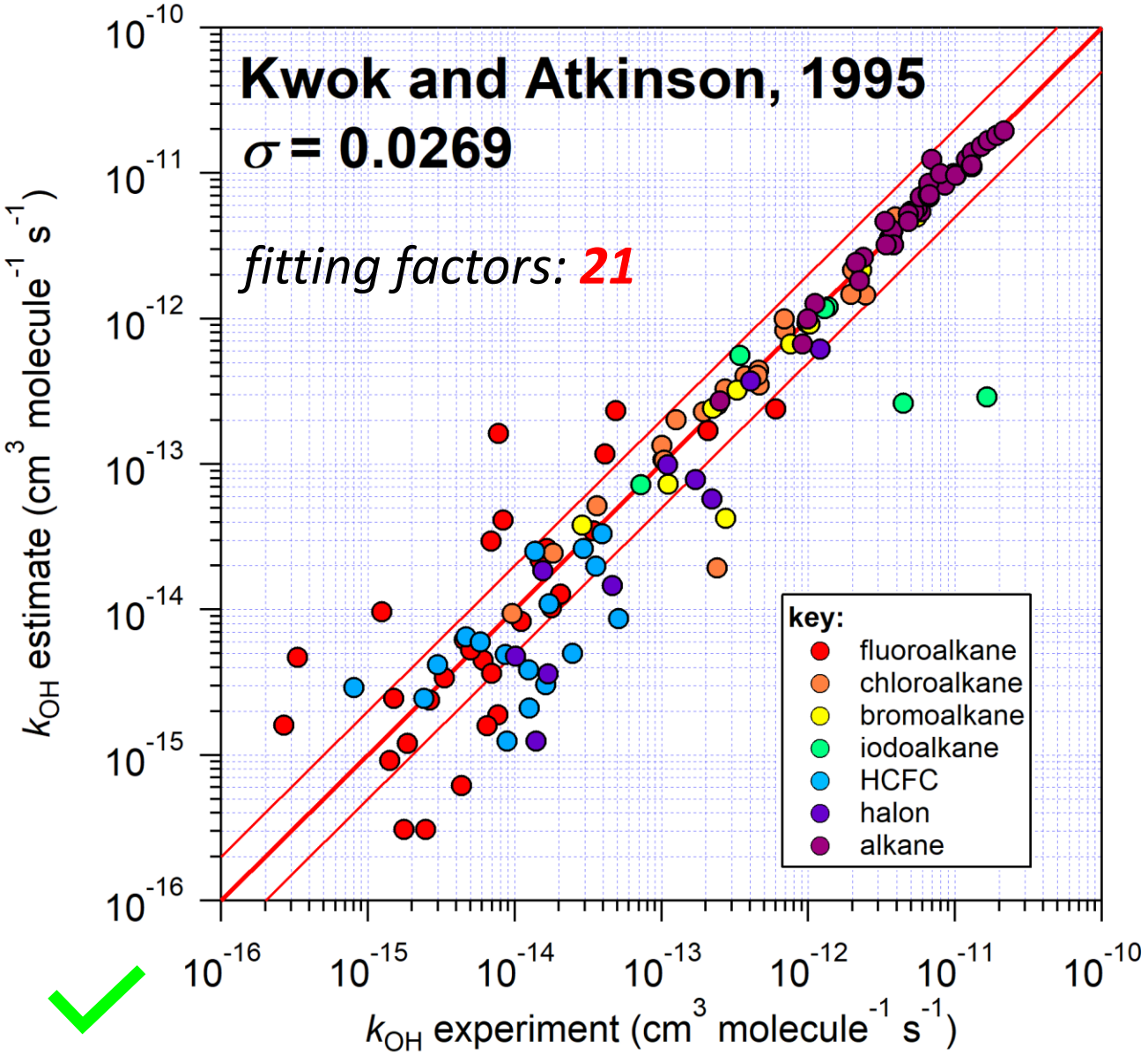
- Different methods for estimation

The Sandbox: a unique testing environment containing hydrocarbons and halocarbons

- Only consider alkanes and haloalkanes
- Atmospherically important
- (relatively) high data coverage
- Variation in reactivity
- Mechanistically simpler
- Complementary information
- Highly functionalized

Different methods for estimation

Group-additivity methods



- This method (and its derivatives) is the go-to group-additivity method for atmospheric chemistry
- **Despite this, it's not very effective for haloalkanes**

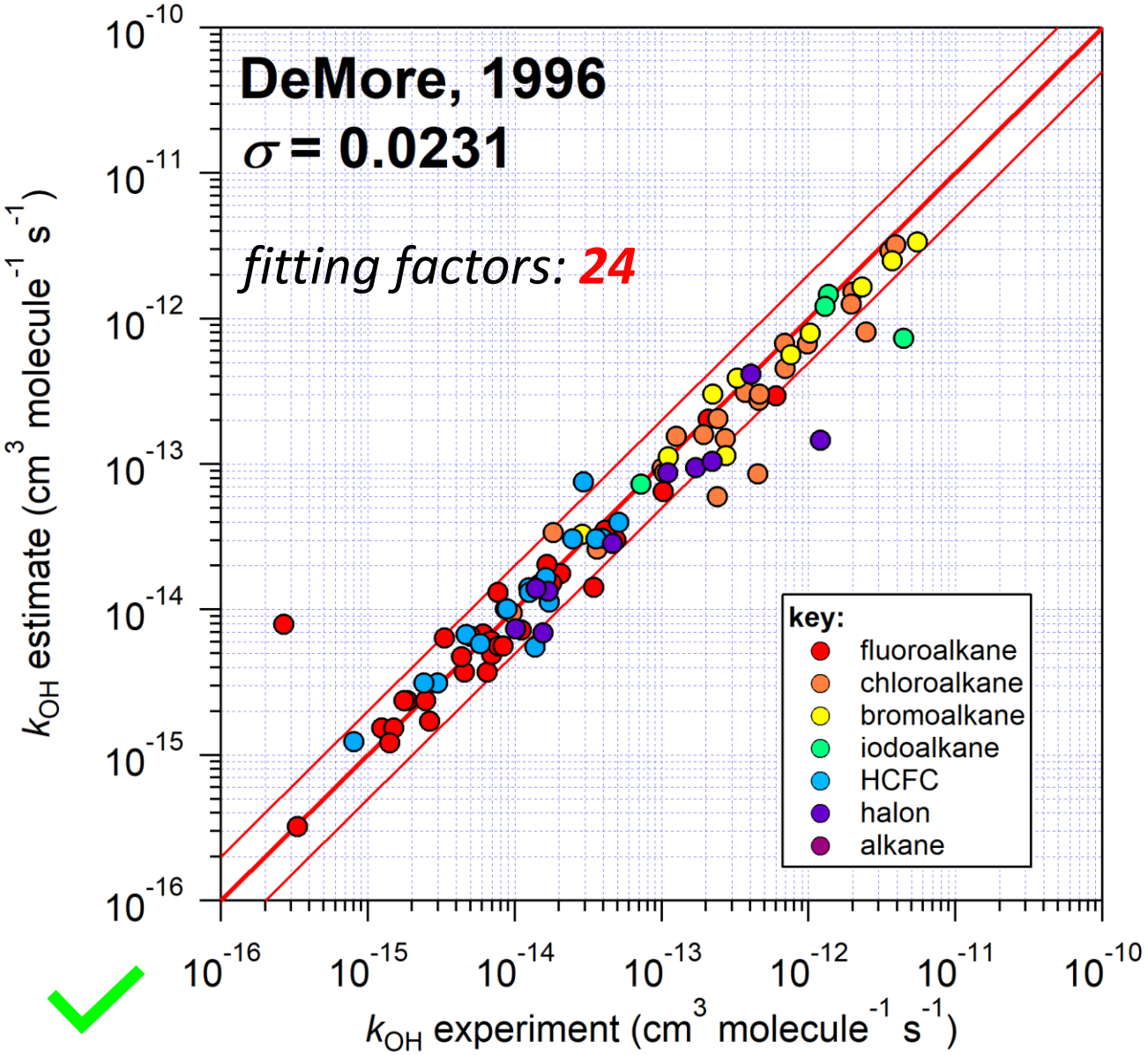
$$k(\mathrm{CH_3–X}) = k_{\mathrm{prim}}F(\mathrm{X})$$

$$k(\mathrm{X–CH_2–Y}) = k_{\mathrm{sec}}F(\mathrm{X})F(\mathrm{Y})$$

$$k\left(\mathrm{X–CH}\begin{matrix}\diagup\mathrm{Y}\\ \diagdown\mathrm{X}\end{matrix}\right) = k_{\mathrm{tert}}F(\mathrm{X})F(\mathrm{Y})F(\mathrm{Z})$$

- Kwok, E. & Atkinson, R. Estimation of hydroxyl radical reaction rate constants for gas-phase organic compounds using a structure-reactivity relationship: An update. *Atmos. Environ.* **29**, 1685–1695 (1995).

Group-additivity methods

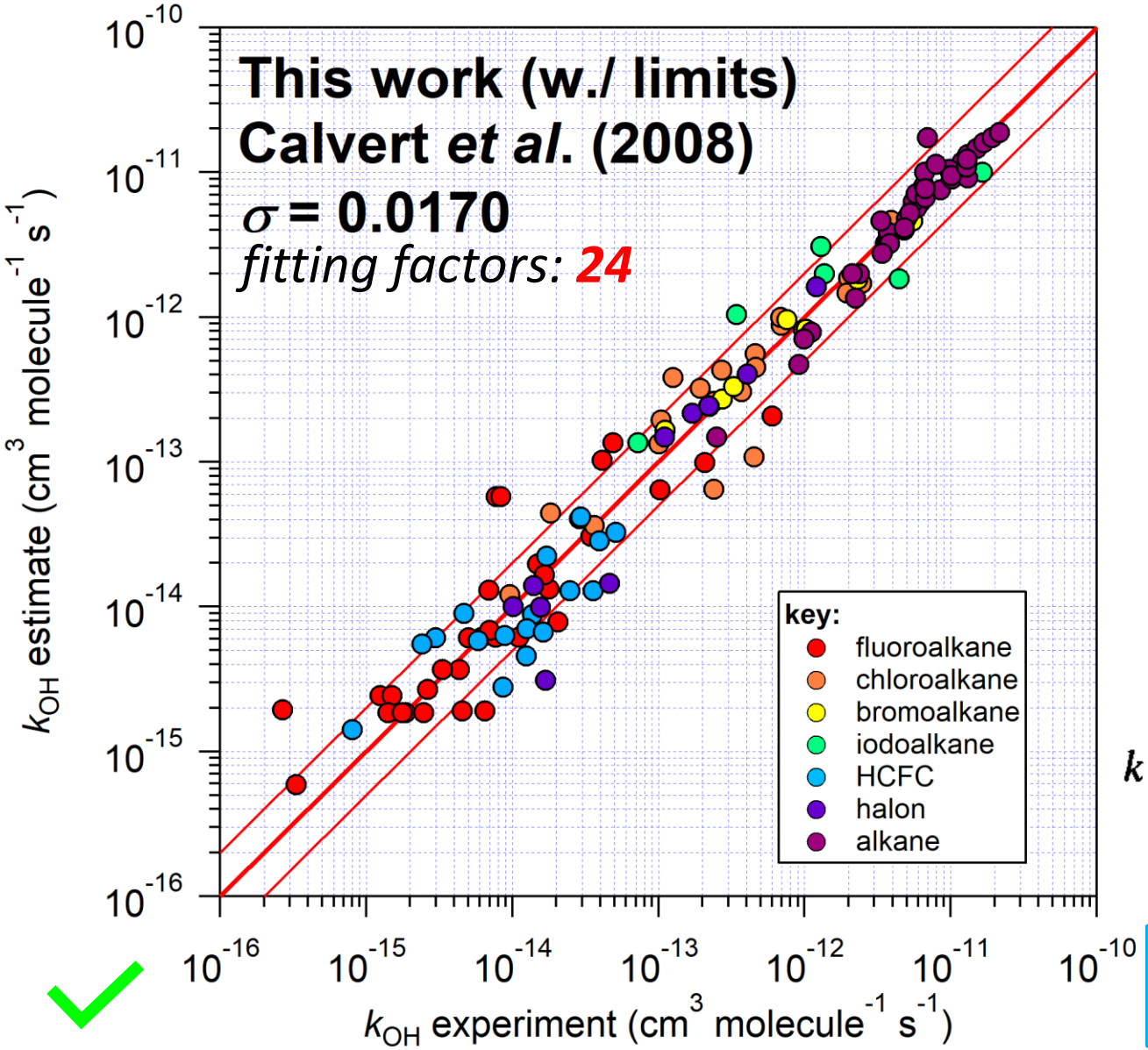


- This approach is less commonly used than the Atkinson approach
- **However, it is clearly superior for haloalkanes, which may relate to it's "third-group multiplier"**

$$\log k = \log k(\mathrm{CH_4}) + G_1 \ldots G_3$$

- DeMore, W. B. Experimental and Estimated Rate Constants for the Reactions of Hydroxyl Radicals with Several Halocarbons. *J. Phys. Chem.* **100**, 5813–5820 (1996).
- **Update of fitting parameters: this work**

Group-additivity methods



- We updated an update of the Atkinson method (Calvert) to impose lower limits on individual reaction sites
- **This mimics the "third-group multiplier" and appears to improve the algorithm**

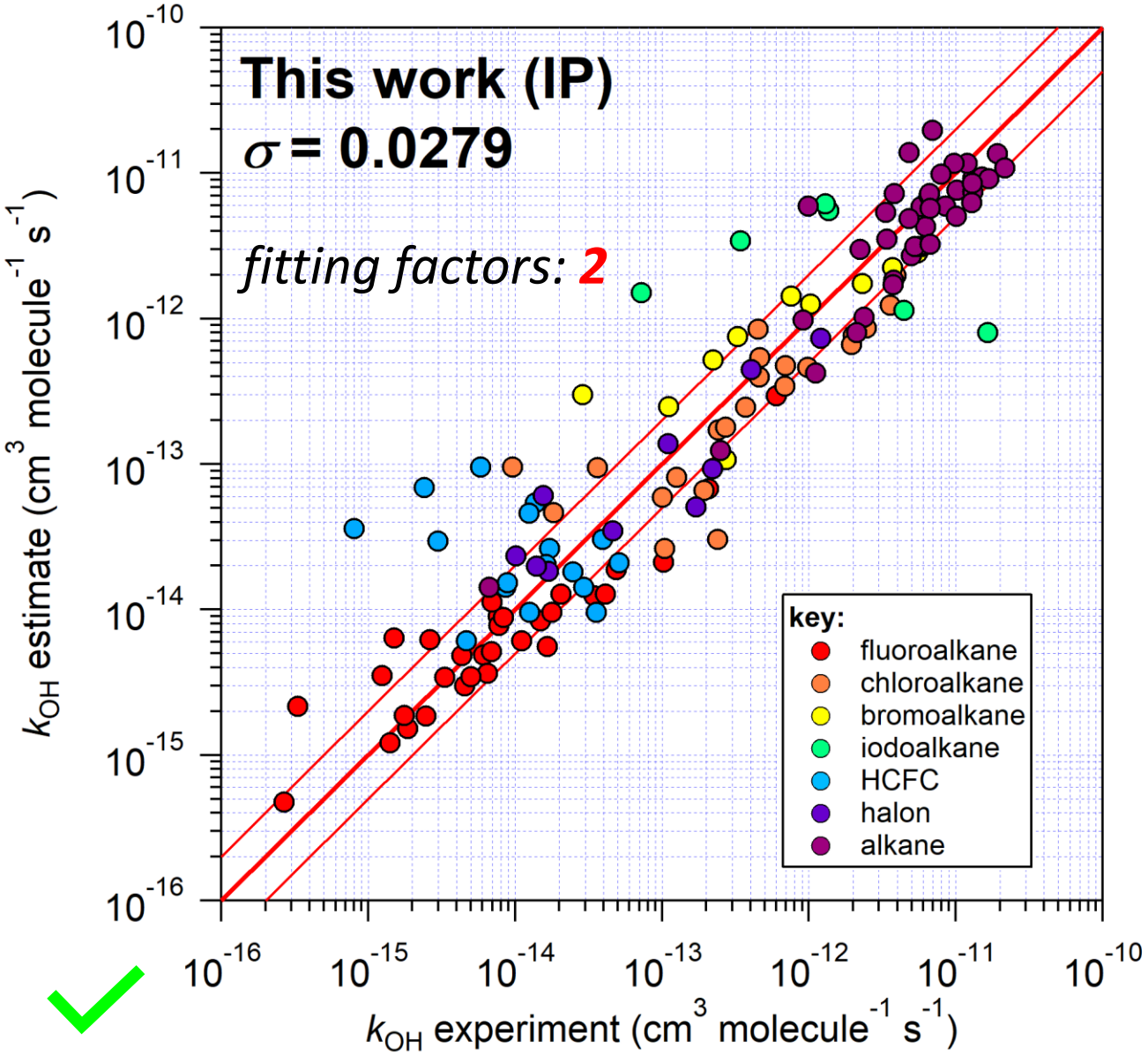
$$k(\mathrm{CH_3–X}) = k_{\mathrm{prim}}F(\mathrm{X}) \qquad \text{if } k(\mathrm{CH_3\text{-}X}) < a = a$$

$$k(\mathrm{X–CH_2–Y}) = k_{\mathrm{sec}}F(\mathrm{X})F(\mathrm{Y}) \qquad \text{if } k(\mathrm{X\text{-}CH_2\text{-}Y}) < b = b$$

$$k\left(\mathrm{X–CH}\begin{matrix}\diagup\,\mathrm{Y}\\ \diagdown\,\mathrm{X}\end{matrix}\right) = k_{\mathrm{tert}}F(\mathrm{X})F(\mathrm{Y})F(\mathrm{Z}) \qquad \text{if } k(\mathrm{X\text{-}CH_2\text{<}}\begin{matrix}\mathrm{Y}\\ \mathrm{Z}\end{matrix}) < c = c$$

- Calvert, J. G., Derwent, R. G., Orlando, J. J., Tyndall, G. S. & Wallington, T. J. *Mechanisms of the atmospheric oxidations of the alkanes*. (Oxford University Press, 2008).
- **Algorithm update: this work**

Linear free-energy relationship methods

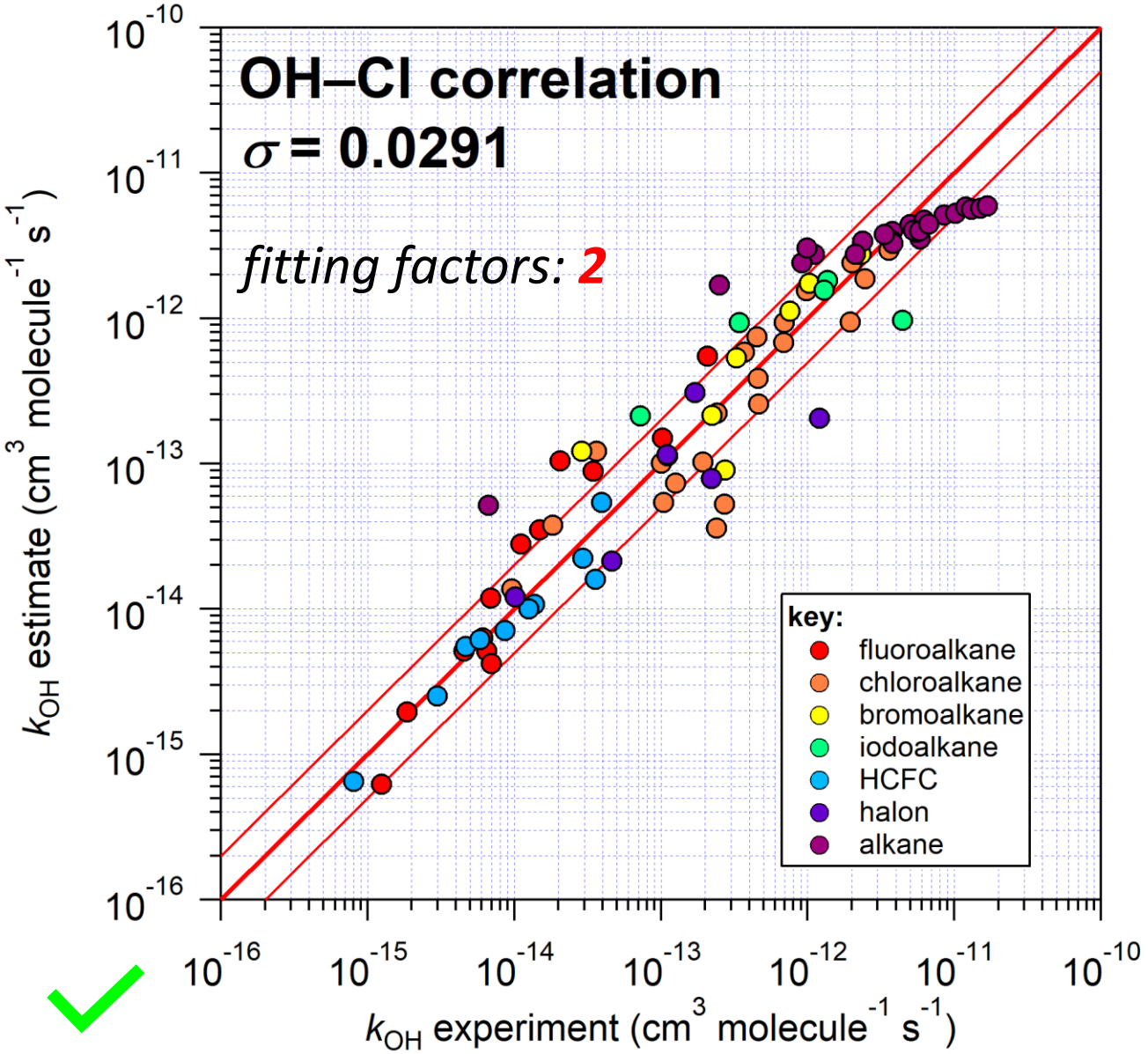


- Ionization potential (IP) has been used previously to correlate with k_{OH}
- **IP calculated using PM6, and corrected against experimental values (where available)**

$$\frac{\log k}{n_H} = m \cdot \mathrm{IP} + c$$

- Percival, C. J., Marston, G. & Wayne, R. P. Correlations between rate parameters and calculated molecular properties in the reactions of the hydroxyl radical with hydrofluorocarbons. *Atmospheric Environment* **29**, 305–311 (1995).
- Stewart, J. J. P. MOPAC2016.
- **Algorithm update, IP calculations: this work**

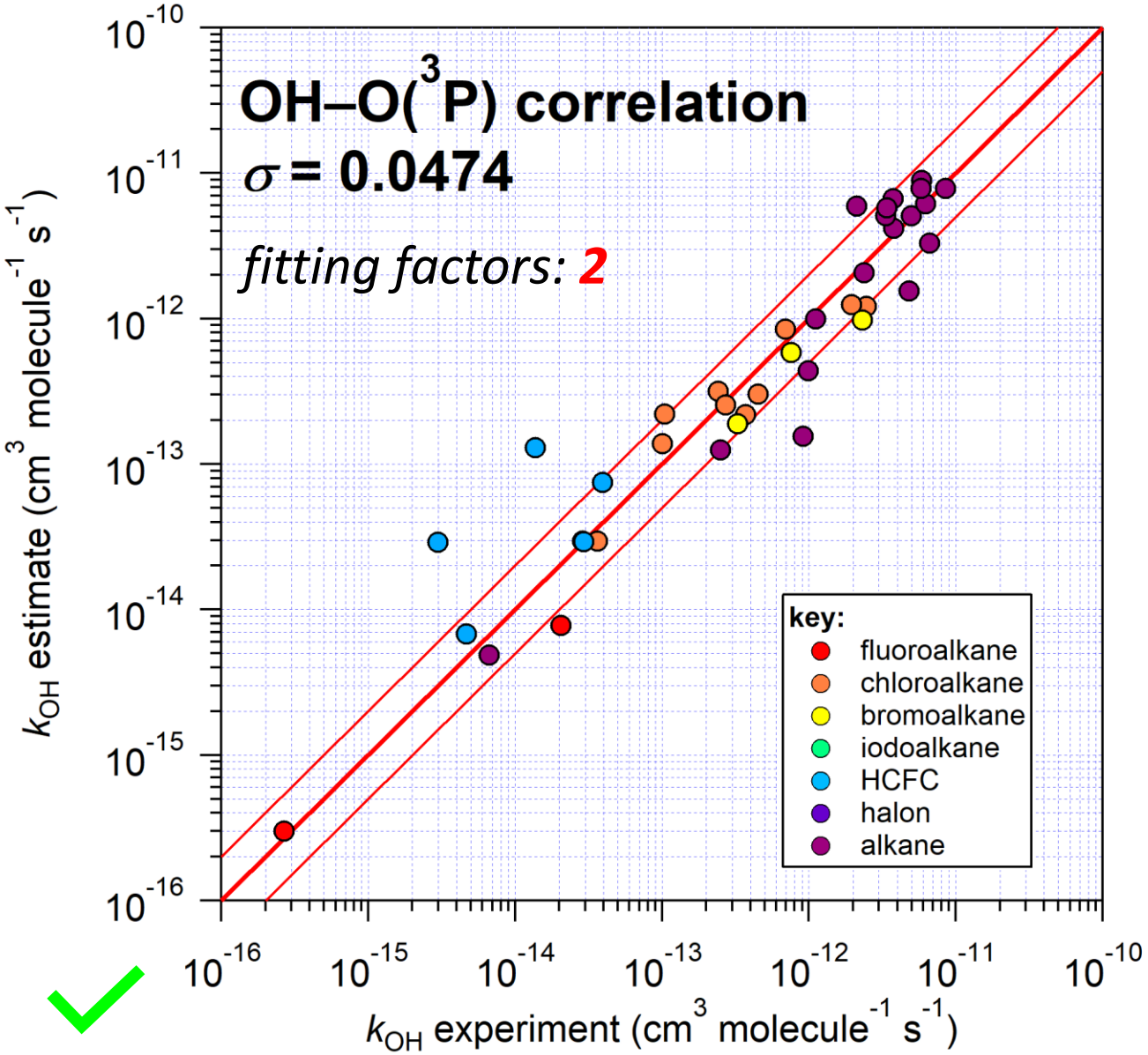
Linear free-energy relationship methods



- The OH–Cl relationship has been observed to work well before
- It implies a similar mechanism, probably C–H bond strength
- **Although it is a reasonably good predictor, there are some problems with higher *k* values (which can be corrected)**

- T.J. Wallington and co-workers, various works, applied mainly to HFEs
- **This work, focussing on a more complete alkanes and haloalkane dataset**

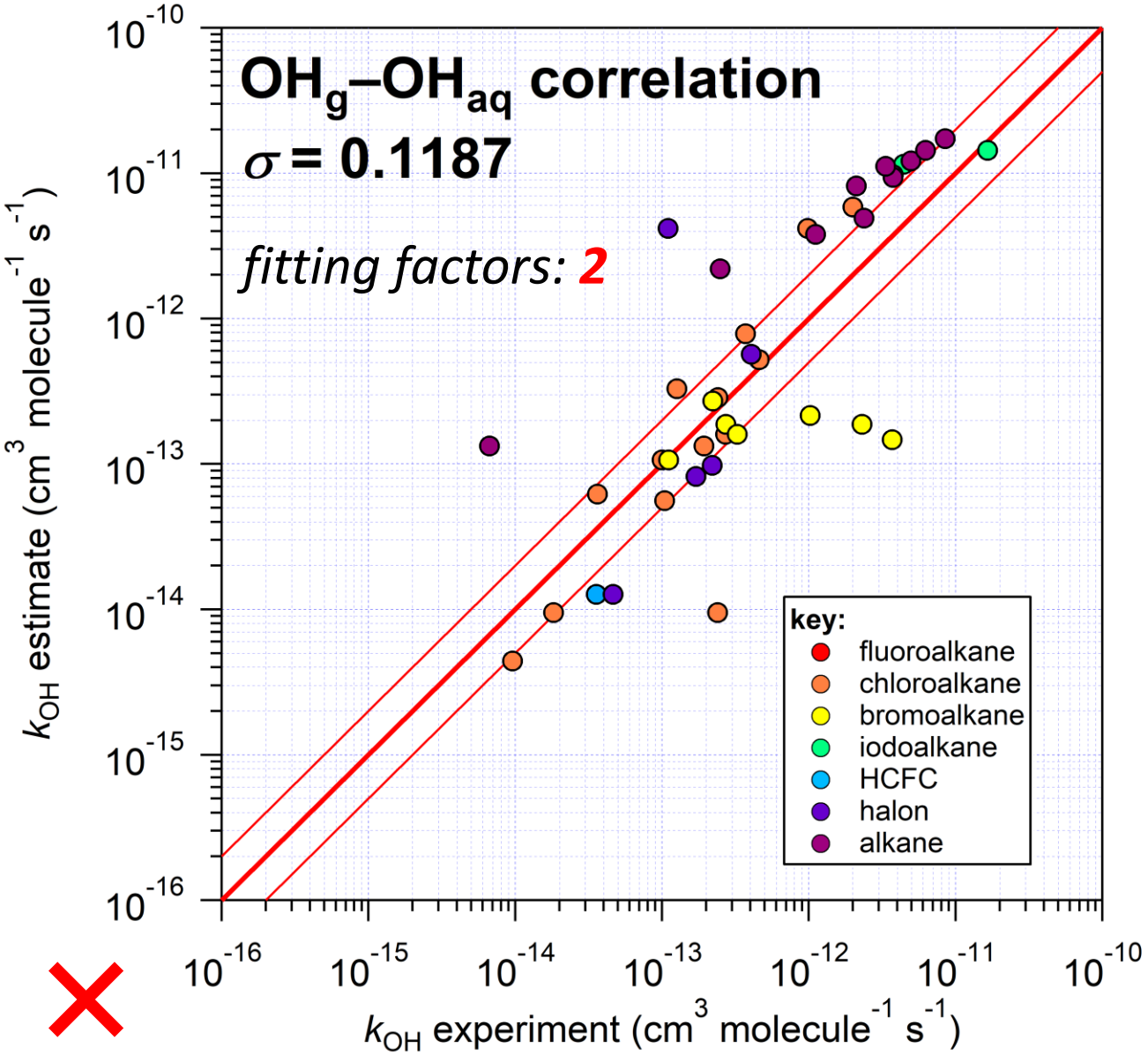
Linear free-energy relationship methods



- The OH–O(^{3}P) relationship has also been observed to work well before
- As with Cl, it implies a similar mechanism, probably C–H bond strength
- **These types of measurements seem to have gone out of favour, so there is very little new data to correlate**

- Gaffney, J. S. & Levine, S. Z. Predicting gas phase organic molecule reaction rates using linear free-energy correlations. I. O(^{3}P) and OH addition and abstraction reactions. *Int. J. Chem. Kinet.* **11**, 1197–1209 (1979).
- **This work, focussing on a more complete alkanes and haloalkane dataset**

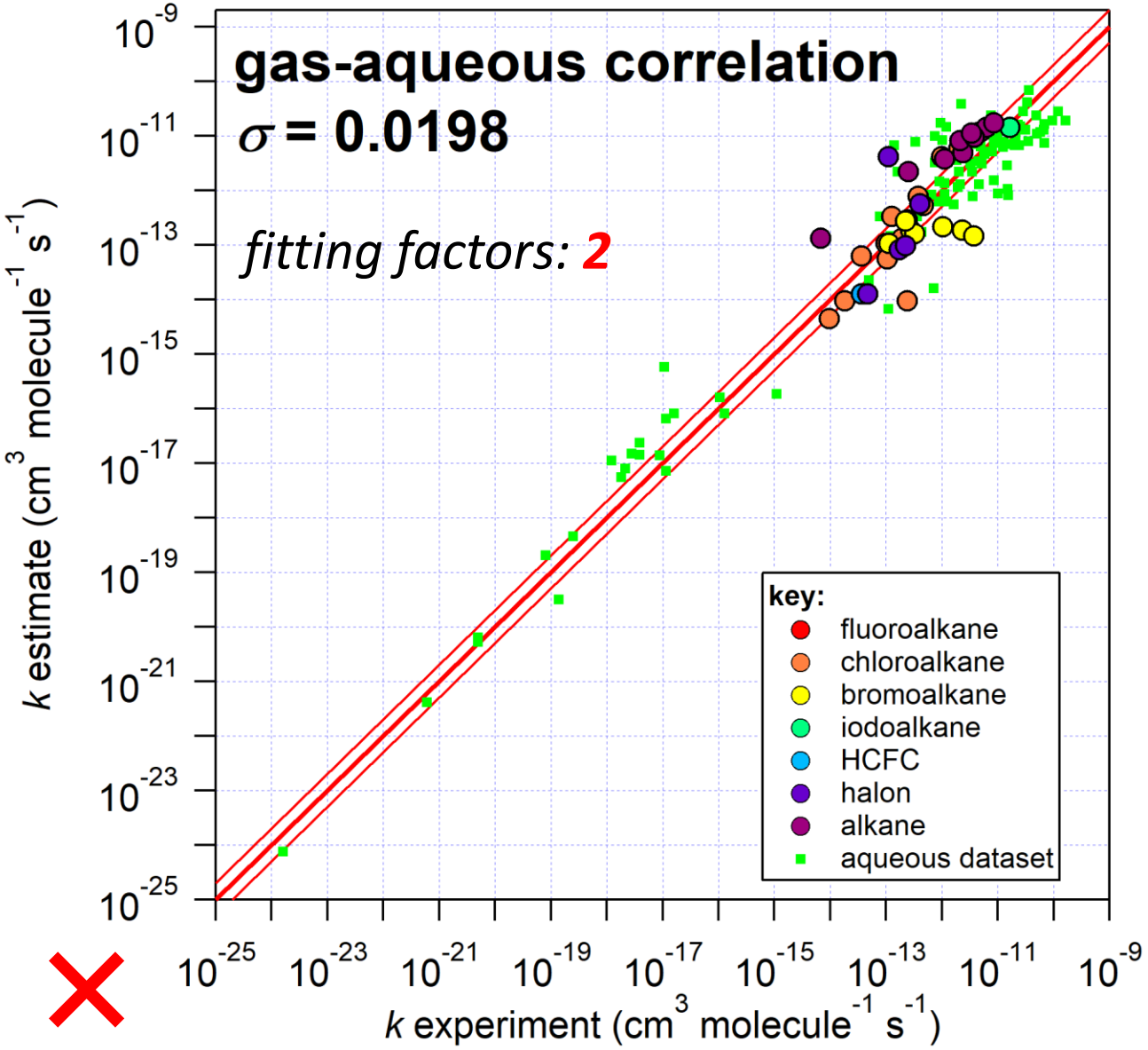
Linear free-energy relationship methods



- Correlations between aqueous-phase and gas-phase rate coefficients have been attempted previously
- **For this set of compounds, it appears (at least with the quality of the data available) that this relationship is too scattered to be useful**

- Klöpffer, W., Kaufmann G. & Frank, R. Phototransformation of Air Pollutants: Rapid Test for the Determination of *k*OH. *Zeitschrift für Naturforschung A* **40**, 686 (1985).
- Wallington, T. J., Dagaut, P. & Kurylo, M. J. Correlation between gas-phase and solution-phase reactivities of hydroxyl radicals towards saturated organic compounds. *J. Phys. Chem.* **92**, 5024–5028 (1988).
- **This work, focussing on a more complete alkanes and haloalkane dataset**

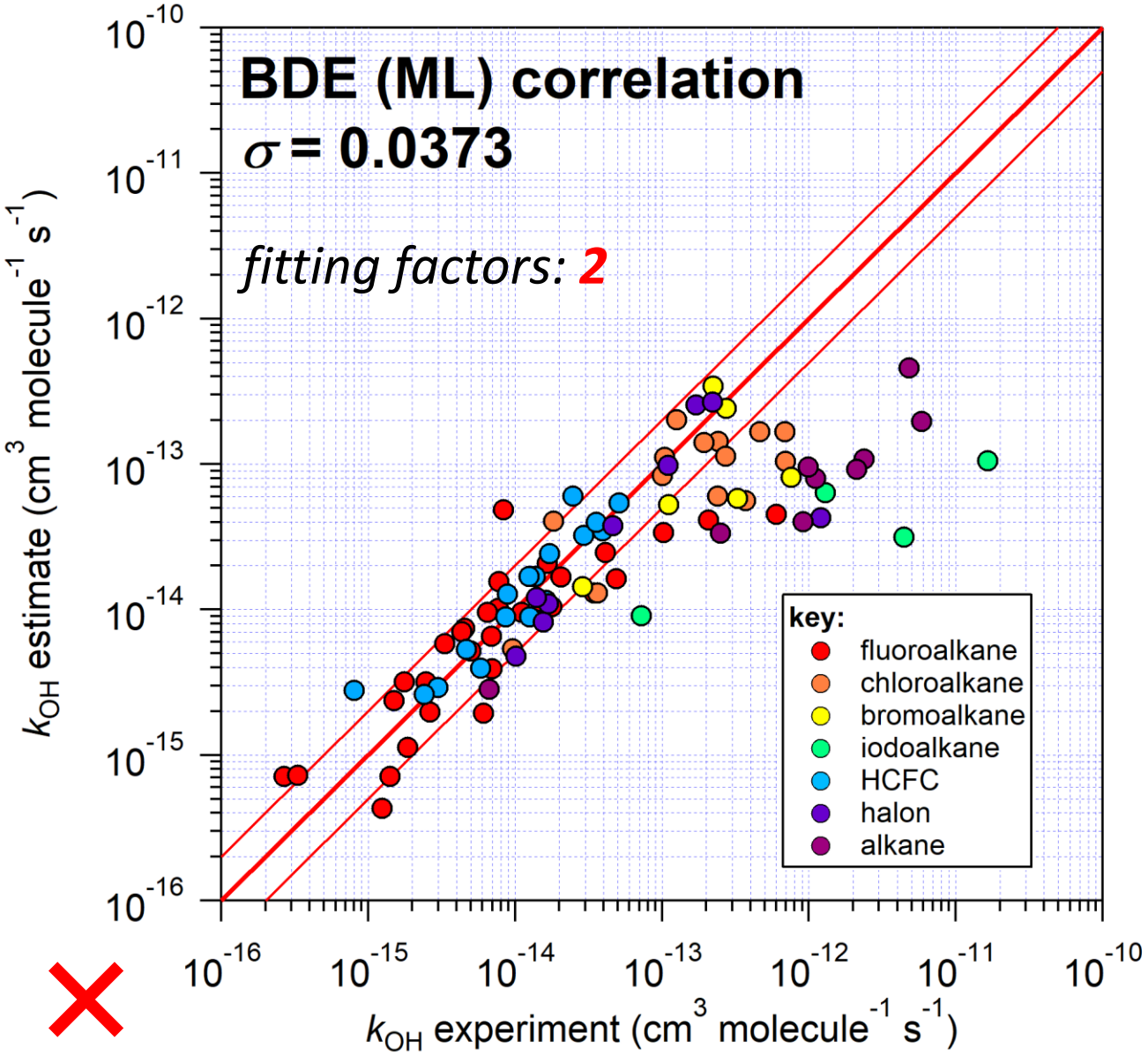
Linear free-energy relationship methods



- But, it is more compelling if we look at the broader picture
- Similarities between gas- and solution-phase have been observed for ≥100 years!
- **Alkanes and halocarbons are among the biggest outliers, which may result from difficulties getting them into aqueous solution**

- Daniels, F. & Johnston, E. H. The thermal decomposition of gaseous nitrogen pentoxide. A monomolecular reaction. *J. Am. Chem. Soc.* **43**, 53–71 (1921).
- Lueck, R. H. The thermal decomposition of nitrogen pentoxide in solution. *J. Am. Chem. Soc.* **44**, 757–769 (1922).

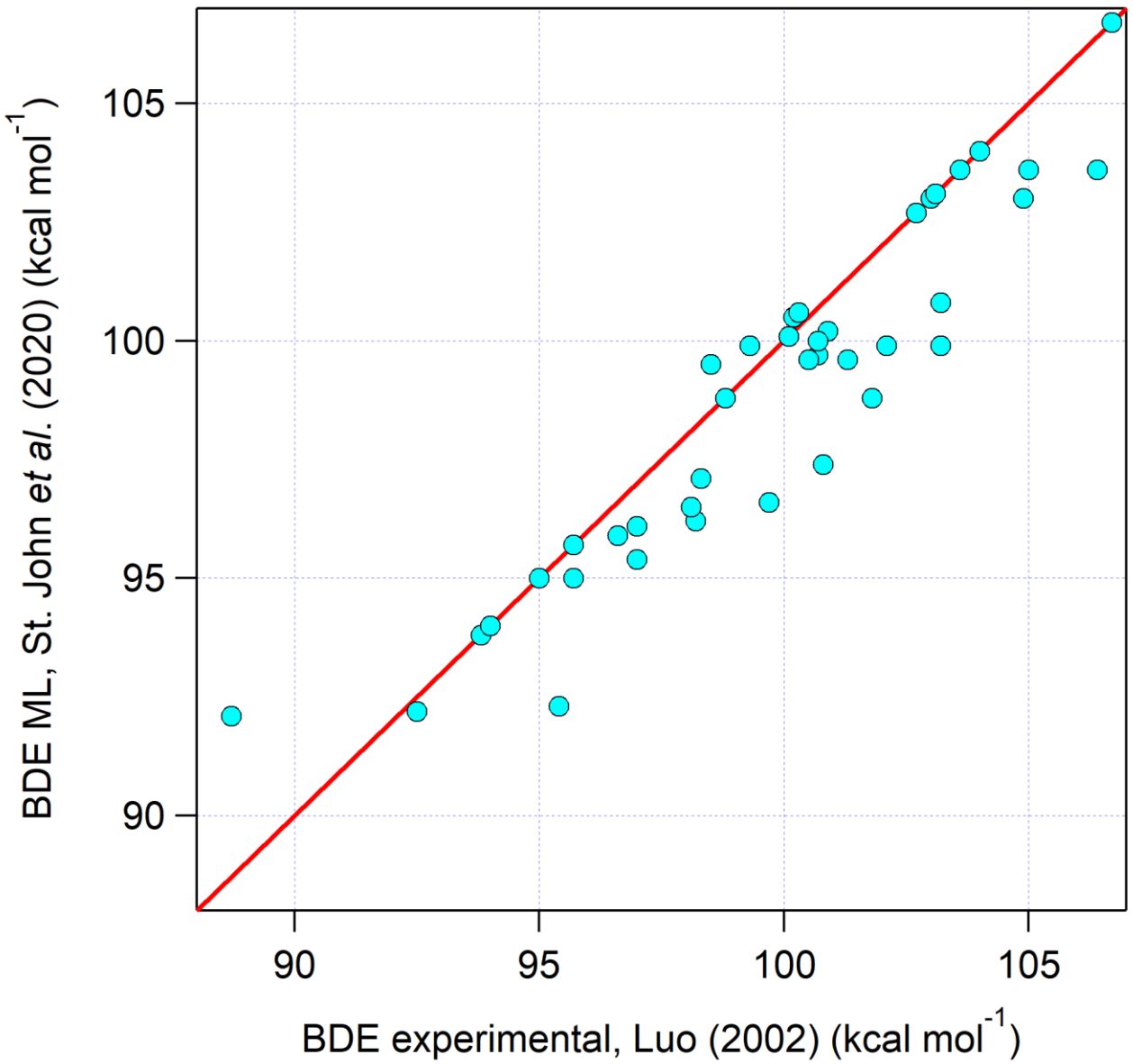
Linear free-energy relationship methods



- It is expected that BDE is one of the primary determinants in this hydrogen abstraction mechanism
- Accurate BDEs are few and far between, a machine-learning method is available
- **This relationship appears to break down for the more reactive molecules (E_a vs A?)**

- Hsu, K.-J. & DeMore, W. B. Temperature-Dependent Rate Constants and Substituent Effects for the Reactions of Hydroxyl Radicals with Three Partially Fluorinated Ethers. *J. Phys. Chem.* **99**, 11141–11146 (1995).
- St. John, P. C., Guan, Y., Kim, Y., Kim, S. & Paton, R. S. Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nat. Commun.* **11**, 2328 (2020).

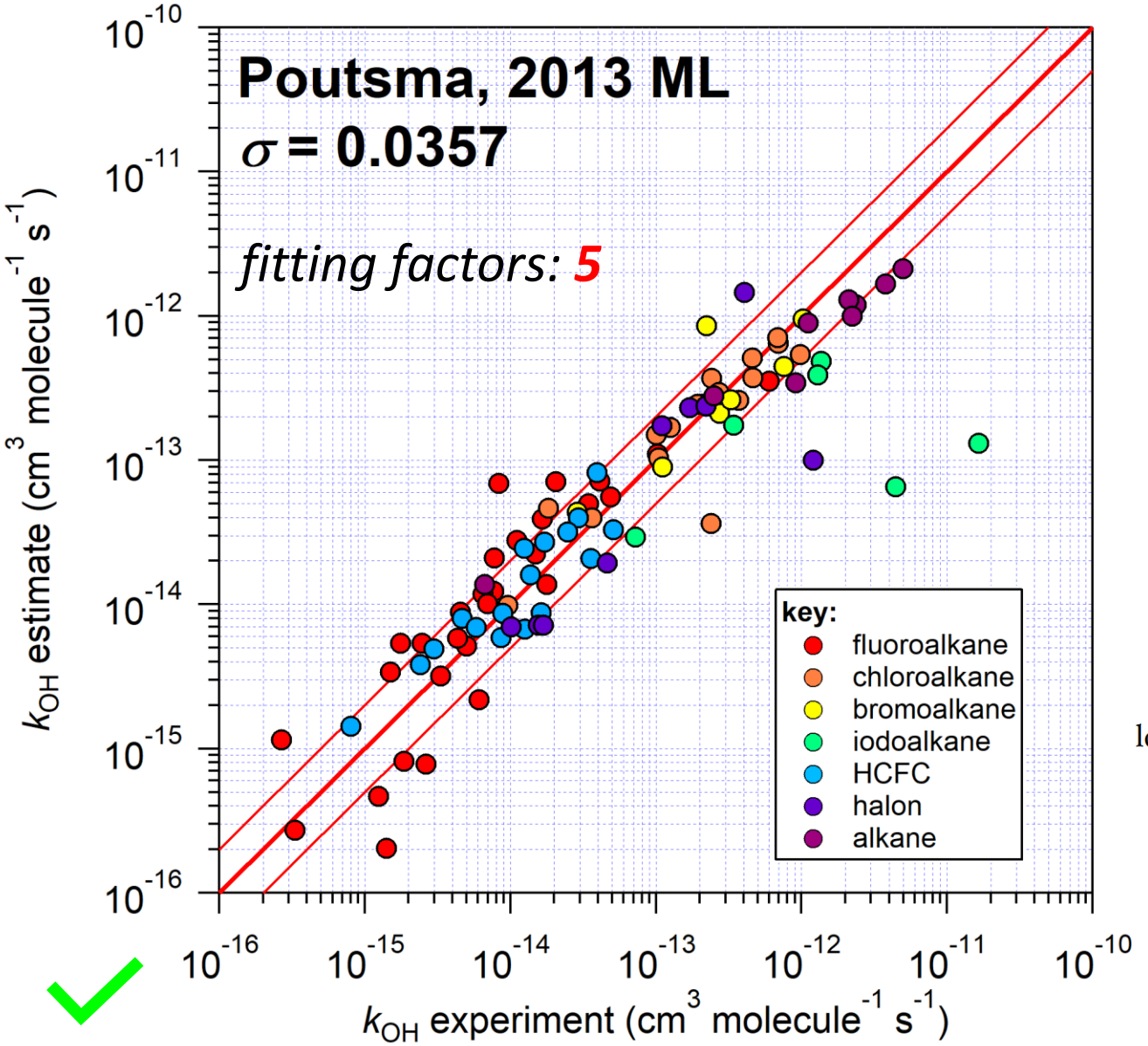
Linear free-energy relationship methods



- Is the machine-learning approach accurate for BDEs in our dataset?
- **It seems to perform quite well, of course there are also uncertainties in the experimental database…**

- Luo, Y.-R. *Handbook of Bond Dissociation Energies in Organic Compounds*. (CRC Press, 2002). doi:10.1201/9781420039863.
- St. John, P. C., Guan, Y., Kim, Y., Kim, S. & Paton, R. S. Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nat. Commun.* **11**, 2328 (2020).

Linear free-energy relationship methods



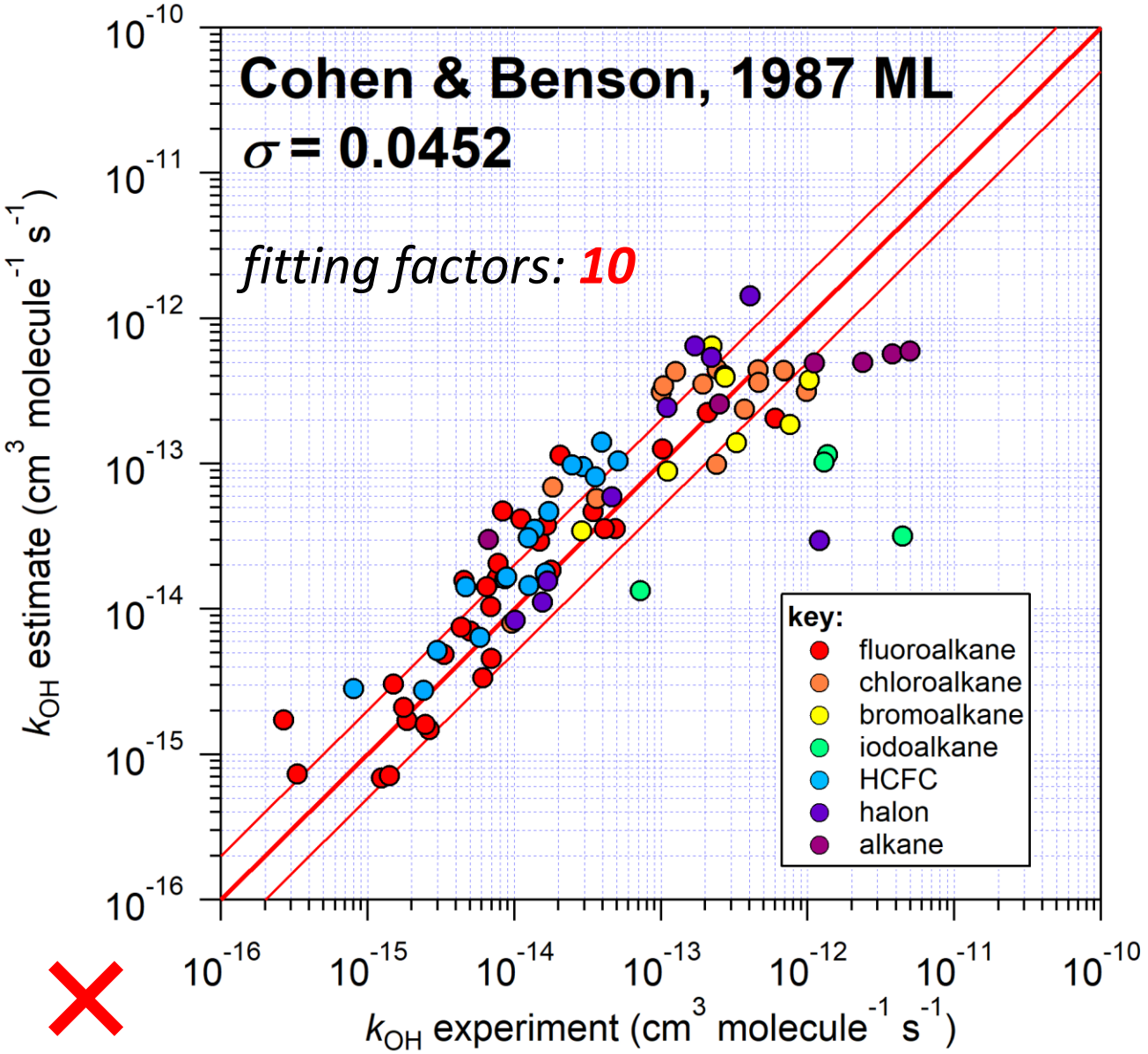
- This method requires a lot of independent variables: ($\Delta_r H$, F, R)
- This isn't very practical! However, it becomes more practical with ML
- **This extends the applicability of Poutsma's approach, which is otherwise quite reliable**

$$\log k(\mathrm{OH}) = -0.000630\Delta_r H^2 - 0.151\Delta_r H - 1.056\sum F - 1.053\sum R - 21.26$$

$$where\ \Delta_r H = \mathrm{BDE(R{-}H)}_{ML} - \mathrm{BDE(HO{-}H)}$$

- Poutsma, M. L. Evolution of Structure–Reactivity Correlations for the Hydrogen Abstraction Reaction by Hydroxyl Radical and Comparison with That by Chlorine Atom. *The Journal of Physical Chemistry A* **117**, 6433–6449 (2013).

Linear free-energy relationship methods



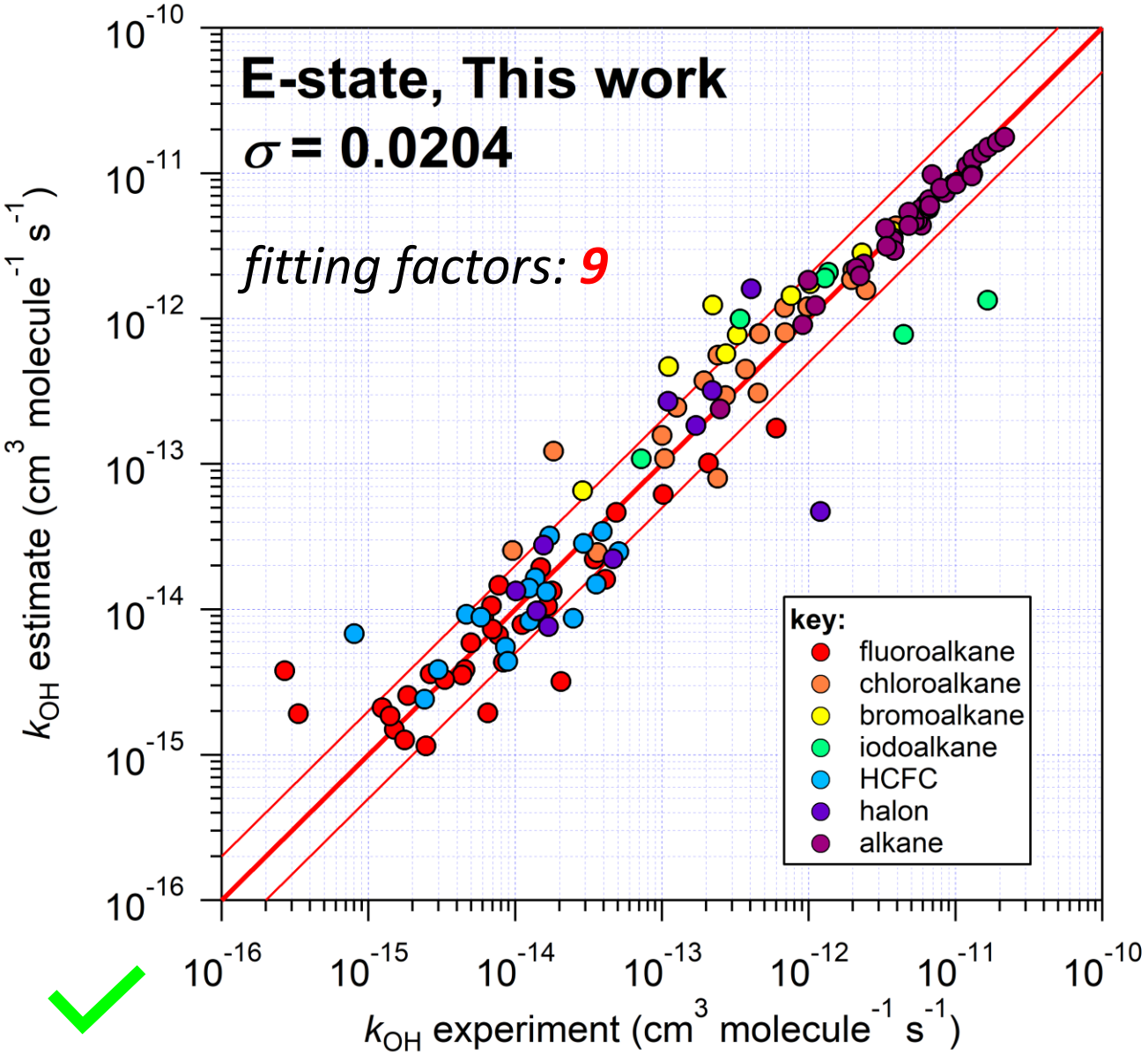
- This method was developed at a time when there was less kinetic information available
- **Considering this, it performs quite well, and it can probably be improved in the future**

$$k(T) = n_{\mathrm{H}} 10^{6.55} M^{-1} T^{1.5} \exp[-190(D_{\mathrm{RH}} - 95.3)/T]$$

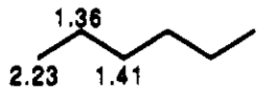
$$where\ D_{\mathrm{RH}} = D_{\mathrm{RH}_{ML}} = \mathrm{BDE(R\text{–}H)}_{ML}$$

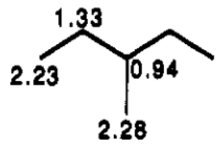
- Cohen, N. & Benson, S. W. Empirical correlations for rate coefficients for reactions of hydroxyl radicals with haloalkanes. *J. Phys. Chem.* **91**, 171–175 (1987).
- **This work, extending application with ML BDEs**

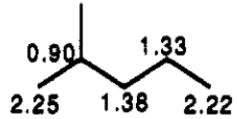
New approach: the electrotopological state



- A new method has been developed, which makes use of the Kier-Hall index
- This describes the electronic character and the connectivity of a hydrogen environment
- **This approach yields site-specific rate coefficients with high accuracy and a low number of fitting parameters**







Hall, L. H., Mohney, B. & Kier, L. B. The electrotopological state: structure information at the atomic level for molecular graphs. *J. Chem. Inf. Model.* **31**, 76–82 (1991).

Can we combine these results somehow?

- It seems plausible that in a highly uncertain situation, you should make use of as much information as possible
- The following statement may apply to our individual SAR models:

"Every attempt to represent nature in a set of equations, resolvable on a digital computer, inevitably introduces inaccuracy." (Hagedorn *et al*., 2005)

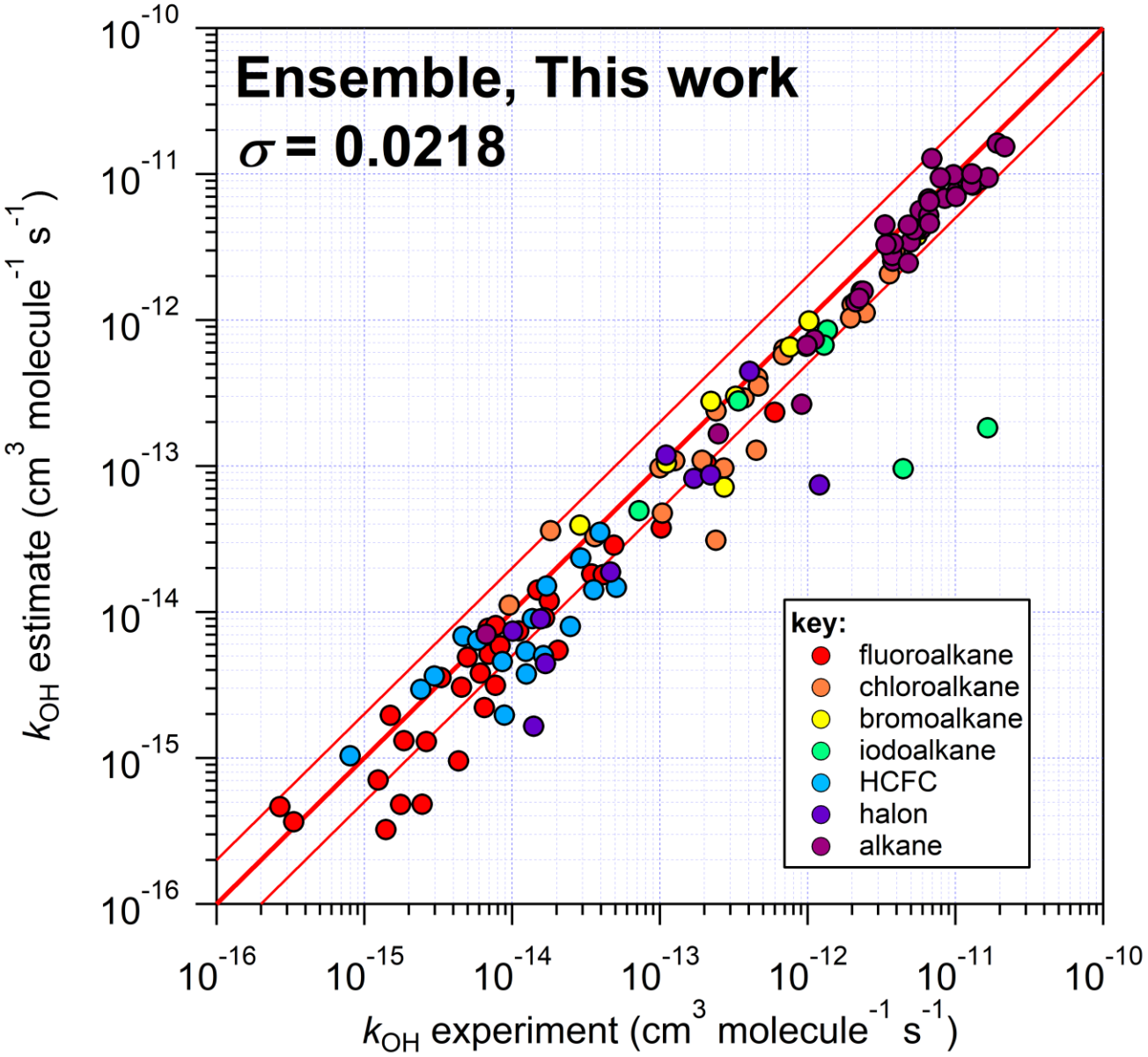
- Hagedorn *et al*. were discussing the macroscopic world of numerical weather prediction – not the microscopic world of chemical reactivity
- However, there is an important parallel: we all must extrapolate
- To test the ensemble, I took a simple "multiple measurement" approach:

$$k_{ensemble} = \sum_{i=1}^{N} w_i k_i \Big/ \sum_{i=1}^{N} w_i$$

$$where, w_i = 1/\sigma_i^2$$

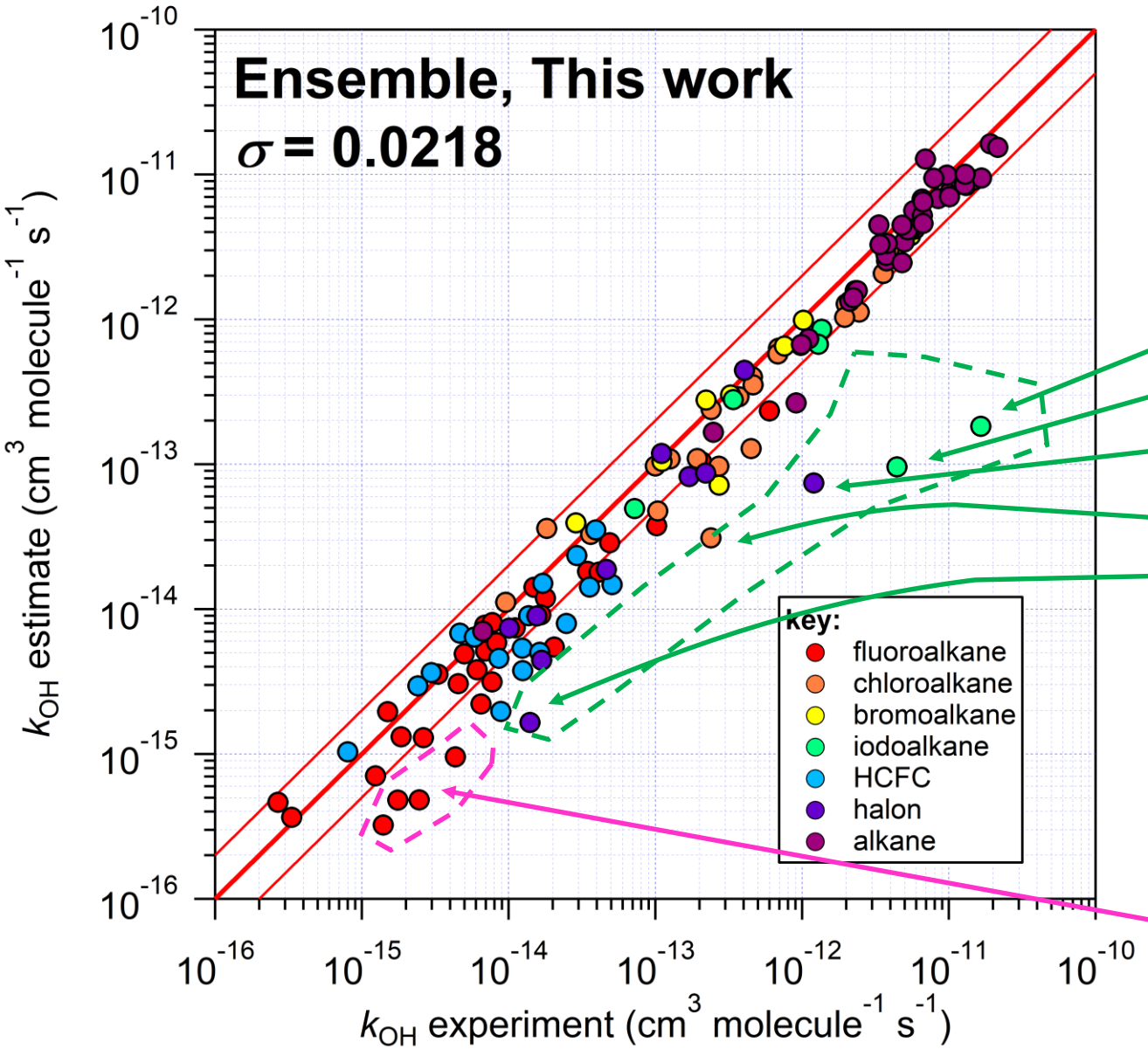
Hagedorn, R., Doblas-Reyes, F. J. & Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus A* **57**, 219–233 (2005).

Bof … ?

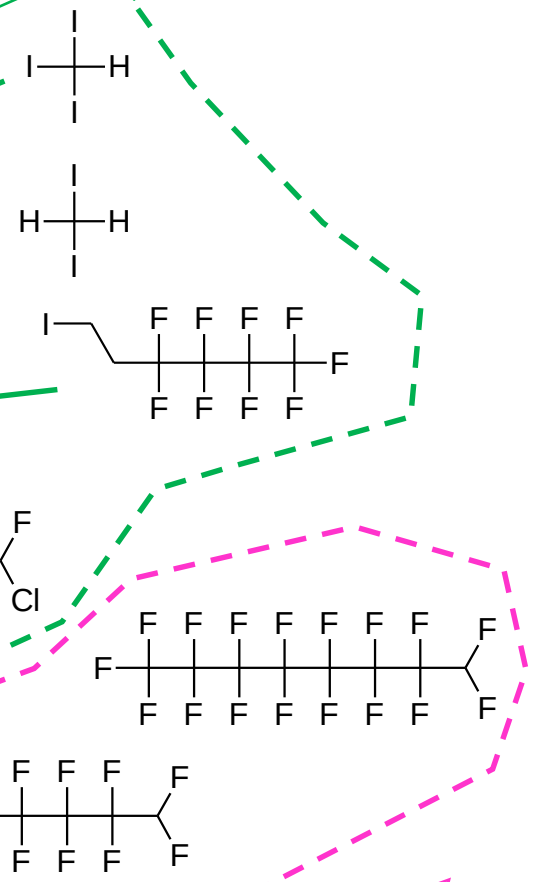


- The ensemble is less accurate at reproducing the experimental data than some individual approaches
- This was (initially) disappointing
- But, we shouldn't assume that our database is perfect
- **There is an interesting asymmetry to the errors in estimation space, with a tendency towards underprediction**

Is the ensemble helpful?



- High UV cross sections
- Low vapour pressure
- Single measurements



- Very low *k* values
- High sensitivity to impurities
- High(er?) GWPs

Conclusions and perspectives

- The Sandbox is a useful place to conduct this thought experiment
- There is a diversity of methods for making predictions, several of which are at least as good as the Atkinson G-A approach
- In addition, I present you with an entirely new site-specific method for rate coefficient estimation based on the electrotopological state
- The ensemble appears to be able to identify possibly erroneous data, **this should be checked with new experiments**
- There are several things that I didn't have time to discuss, but which are also very important:
- Site-specific reaction channels; influence of temperature; **consensus in estimation space** – next time

Thank you for your attention,
and thanks to my colleagues in the CRC SAR panel for helpful discussions