DEVELOPMENT, EXTENSION AND VALIDATION OF (THEORY-BASED) SARS

DECEMBER 2018 I LUC VEREECKEN

Member of the Helmholtz Association

JÜLICH
Forschungszentrum

THE NEED FOR DATABASES AND SARS

Models for atmospheric chemistry are ever more complex

- Air quality : more, and more subtle chemistry, in interaction with biosphere

  O_3, NOx, Aerosols, clouds, oxidative capacity, ...

- Climate change : long-term predictions in varying conditions
- More compounds to describe : more VOC, oxygenates, 2nd gen. products

Kinetic chemical models in atmospheric research:

Models at multiple levels of complexity exist, and must exist

- Explicit mechanisms : e.g. Gecko-A
- Semi-explicit mechanisms : e.g. Master Chemical Mechanism, SAPRC
- Condensed mechanisms : e.g. RADM, RACM
- Parametrized models : e.g. GEOS-Chem, UKCA

Mechanism development

Need for comprehensive and reliable detailed mechanisms

- Serve as benchmark case
- Serve as basis for reduced mechanisms

However, more detail does not imply more accuracy or predictive capability

- Additional details may not be sufficiently characterized

  State of body of knowledge / Structure-activity relationships ?

- Advantages for models specifically tuned to problem case

Comprehensive mechanisms have their specific operational problems

Needs in mechanism development (SAR evaluation group)

- Improvement in methodologies for model generation and maintenance
- Collection and evaluation of experimental and theoretical databases
- Extension and evaluation of Structure-activity relationships (SARs)
- A community supporting these long-term goals

First step : SAR Evaluation group (led by W.P.L. Carter)

Summary of Project Participants (listed alphabetically within groups)

Name	Institution	Experience [a]
	Full Participants	
Bernard Aumont	LISA, Univ. of Paris	Mech-G, MCM, EC
Ian Barnes [b]	Univ. of Wuppertal	Lab, MCM, EC
Joseph Bozzelli	New Jersey Institute of Tech.	Lab, Theory, Therm
William Carter	Univ. of Calif. Riverside	SAR, Mech-A, Mech-G, Model
Mark Goldman	MIT (Student in Green group)	SAR, Mech-C, Mech-G
William Green	MIT	SAR, Mech-C, Mech-G, Theory
Sasha Madronich	NCAR	Mech-A, Calvert, Phot
Max McGillen	Univ. of Bristol	SAR, Lab, Phot
Wahid Mellouki	CNRS, Univ. Orleans	SAR, Lab, MCM, Calvert, IUPAC, EC
John Orlando	NCAR	Mech-A, Lab, Calvert
Benedicte Picquet-Varrault	LISA, Univ. of Paris	SAR, Lab, EC
Andrew Rickard	York University	Mech-A, Lab, MCM, EC
William Stockwell	University of Texas at El Paso	Mech-A, Calvert, Model
Luc Vereecken	Forschungszentrum Jülich	SAR, Mech-A, Theory, MCM
Tim Wallington	Ford	Lab, MCM, Calvert, IUPAC
	Observers	
Richard Derwent	RDScientific	Model, Calvert
Mike Jenkin	Atmos. Chem. Services	MCM, IUPAC
Ajith Kaduwela	California ARB	Model, Fund
Deborah Luecken	U.S. EPA	Model, Fund
Jozef Peeters	Univ. of Leuven (Emeritus)	Mech-A, Theory
Michael Pilling	Univ. of Leeds (Emeritus)	Mech-A, Mech-C, MCM, Calvert
Dudley Shallcross	University of Bristol	Mech-A, Lab

Next step: Formulation of the problem

Int. J. Chem. Kinet. 50, 435-469, 2018.

Perspective on Mechanism Development and Structure-Activity Relationships for Gas-Phase Atmospheric Chemistry

L. VEREECKEN,[1] B. AUMONT,[2] I. BARNES,[3,*] J.W. BOZZELLI,[4] M.J. GOLDMAN,[5] W.H. GREEN,[5] S. MADRONICH,[6] M.R. MCGILLEN,[7] A. MELLOUKI,[8] J.J. ORLANDO,[6] B. PICQUET-VARRAULT,[2] A.R. RICKARD,[9,10] W.R. STOCKWELL,[11] T.J. WALLINGTON,[12] W.P.L. CARTER[13]

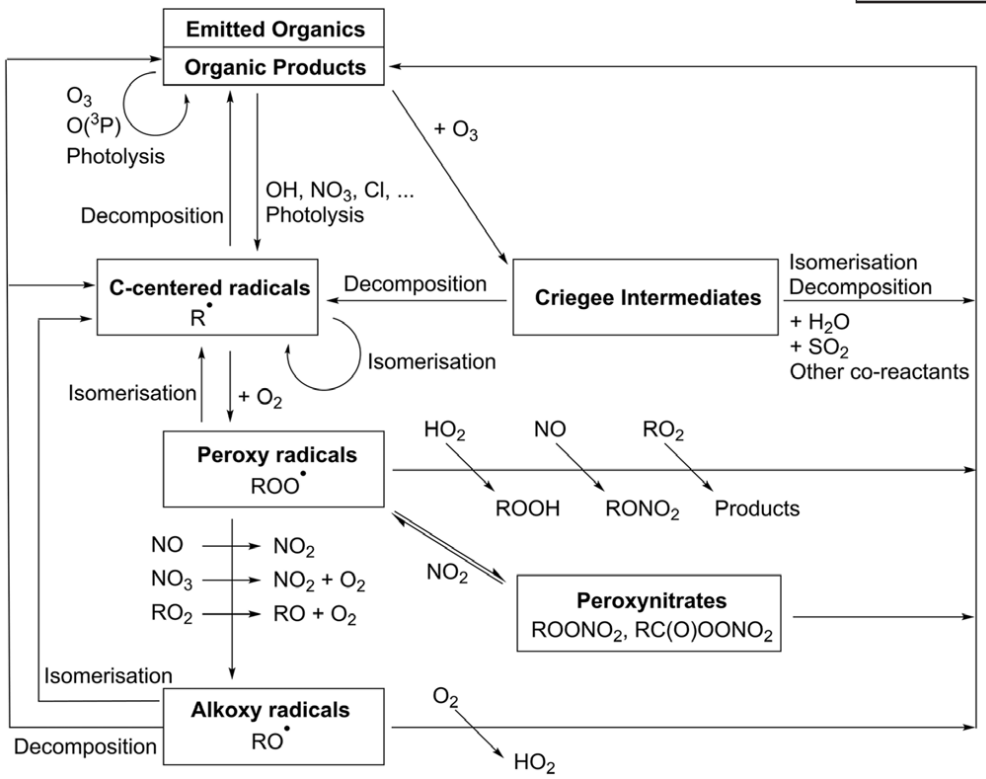
Chemistry:

- Which mechanisms
- Which intermediates

Knowns: (Databases)

- Which values
- What reliability

Hydrocarbon	$-CH_3$	$-CH_2-$	$-CH<$	$>C<$	$C=C$	$C=C-C=C$
	Aromatic	$C=C=C$	$C\equiv C$			
Oxygenate	$-CHO$	$-C(O)-$	$-OH$	$-O-$	$-OOH$	$-OO-$
	$-C(O)O-$	$-C(O)OH$	$-C(O)OOH$	$-OC(O)O-$		
Oxidized Nitrogen	$-ONO_2$	$-ONO$	$-OONO_2$	$-OONO$	$-NO_2$	$-NO$
	$C(O)OONO_2$					
Reduced Nitrogen	$-NH_2$	$-NH-$	$-N<$	$=NH$	$=N-$	
Radical centers	$-CH_2^{\bullet}$	$-CH(^{\bullet})-$	$>C(^{\bullet})-$	$=CH^{\bullet}$	$=C(^{\bullet})-$	$-C(O)^{\bullet}$
	$-O^{\bullet}$	$-OO^{\bullet}$	$-C(O)O^{\bullet}$	$-C(O)OO^{\bullet}$	$-CH:$	$-C(:)-$
	$-NH^{\bullet}$	$-N(^{\bullet})-$	$=N^{\bullet}$			
Criegee Intermediates	$CH_2O^{(+)}O^{(-)}$	$-CHO^{(+)}O^{(-)}$ (*syn*-isomer)		$-CHO^{(+)}O^{(-)}$ (*anti*-isomer)		$>CO^{(+)}O^{(-)}$



Unknowns: (SARs)

- How predictable
- How accurate

Understanding:

Generation & pruning

Reduction

Source: Vereecken et al., IJCK 50, 435, 2018

JÜLICH Forschungszentrum

NEW EFFORTS IN DATABASES

Databases on experimental data are being updated

- IUPAC / NASA-JPL / ... : critically evaluated data
- Databases : NIST Kinetics database, others...
- SAR evaluation group : extending database with k(VOC+OH/Cl/O_3/NO_3)

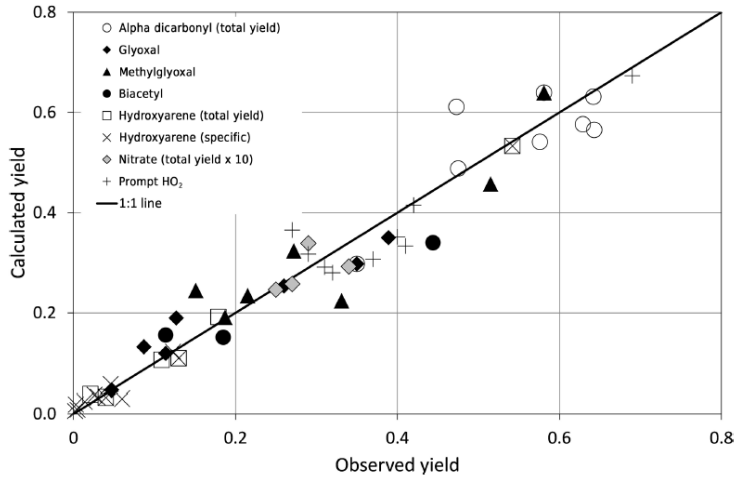
Databases with theoretical data more scarce, but will become more important

- NIST kinetics database and some others have partial theoretical data
- No systematic effort exists for theoretical data for atmospheric kinetics
- SAR evaluation group : TheoKinDB (flyer available, see Luc Vereecken)
  - New effort for construction database with kinetic data
  - Extensive meta-data for higher-level analysis (e.g. error estimates)

Usefulness of databases :

- Data directly useful in models
- Source data for SAR development / evaluation

NEW EFFORTS IN SARS



Experimental SARs

Initiation reactions : VOC + $OH/O_3/NO_3$/ ...

e.g. Jenkin et al., Atm. Chem. Phys. 18, 9297, 2018 : VOC + OH

Jenkin et al., Atm. Chem. Phys. 18, 9229, 2018 : Aromatics + OH

Deconvolute experimental data to identify chemical trends

Thermodynamical data : evaluation of group factors (SAR evaluation group)

SARS based on theoretical data : reliability increases with computational power

Gives direct site-specific information, very good at trend analysis (e.g. subst.)

Useful for experimentally challenging aspects (e.g. radicals, products)

See e.g. Vereecken et al., Phys. Chem. Chem. Phys. 19, 31599, 2017

SAR for CI unimolecular, H_2O, $(H_2O)_2$

SAR development challenges

Rarely validated and evaluated: reliability unknown when outside training set

Lack of validation data is limiting in this respect → ongoing effort

Need to support multi-functionalized species: key to many modern problems

- Combinatorial problem : impossible to do all combinations
- Practical problem: molecules hard for theory, experiment, and SAR

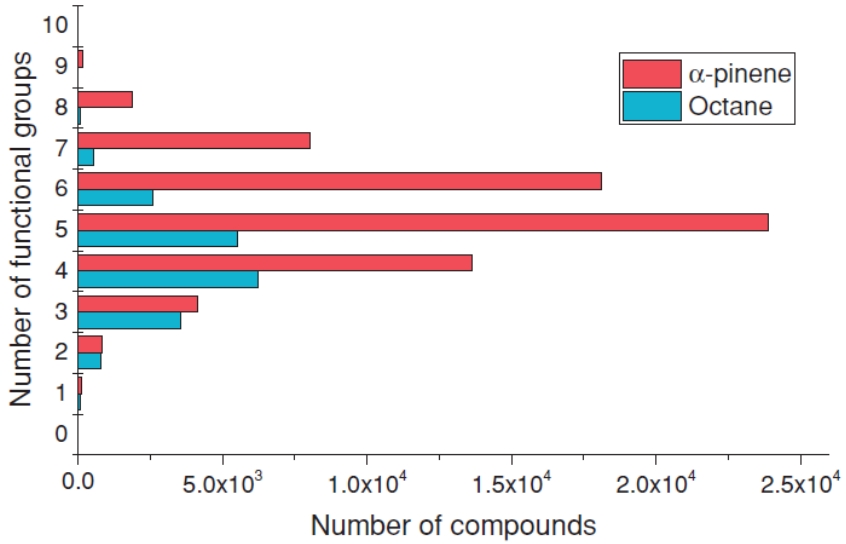


Fig: Gecko-A (octane & α-pinene) >10^5 compounds with >2 funct. groups

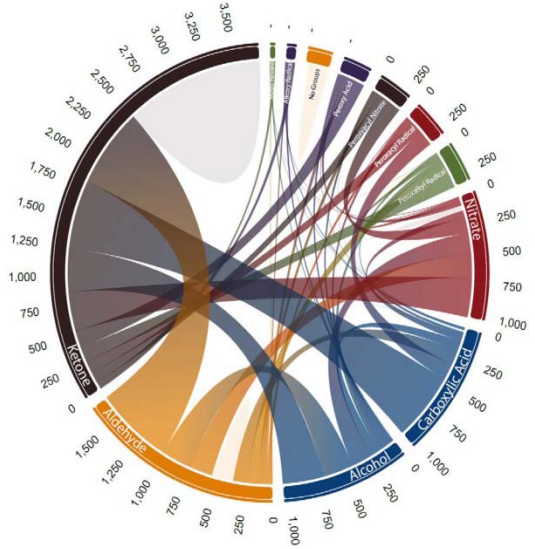


Fig: MCM multi-functionalization not all combinations equally likely

Source: Vereecken et al., IJCK 50, 435, 2018

JÜLICH Forschungszentrum

SAR EVALUATION

Criteria for SAR evaluation

a) Recovery of training set data (= goodness of fit)
b) Recovery of literature (experimental/high-level theory ; can be same as (a))
c) Predictions for substituents within scope of applicability but outside training set (interpolation)
d) Predictions for substituents outside scope of applicability (extrapolation)
e) Predictions for "uniques" - a lost cause?
f) Atmospheric model applications

Example SAR :

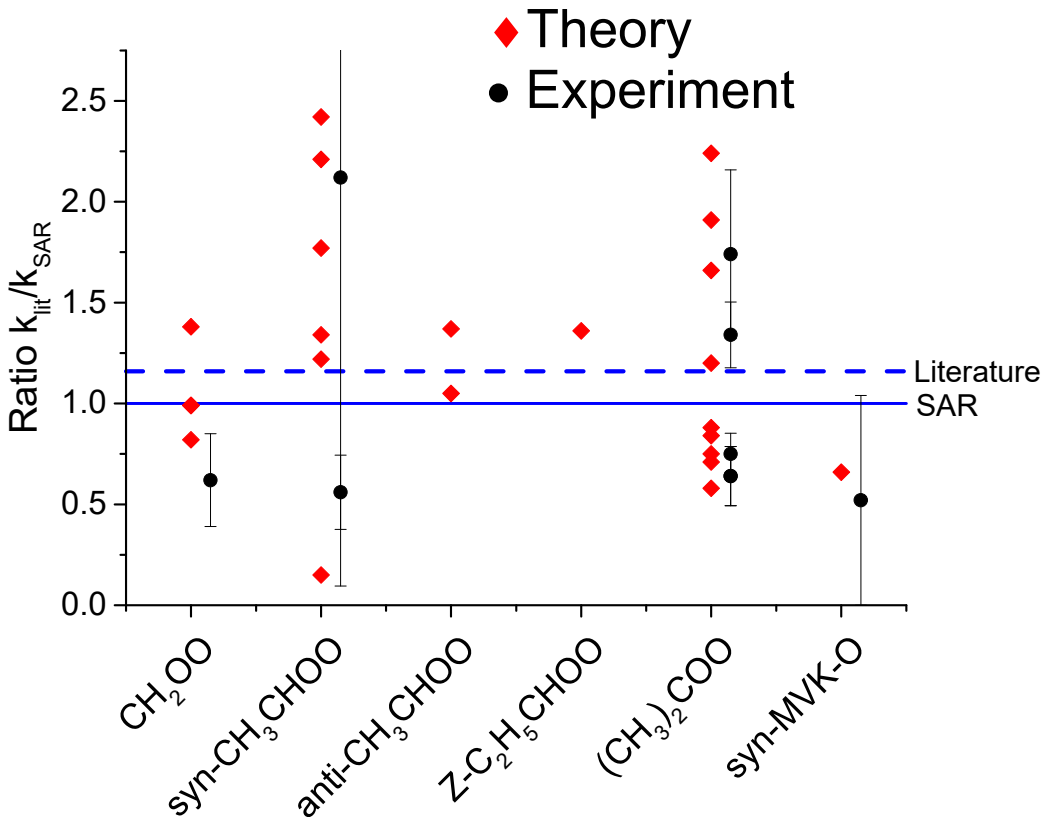
Unimolecular decomposition of carbonyl oxides (Criegee)

Vereecken et al., Phys. Chem. Chem. Phys. 19, 31599, 2017

Theory-based SAR, with targeted calculations for SAR components (*goodness of fit* is mostly meaningless)

Use of theoretical data for validation

Recovery of literature data



16 new studies since SAR publication

Comparison against all literature:

- Factor 0.86 underestimation
- Within literature uncertainty
- *A priori* SAR factor 10 uncertainty

Validity of metric :

- Only simple CI available
- Some literature data biased

Complex compounds:

- Few speciated data
  Experimental data
  is population-averaged

α/β-pinene: ×2.5

(Limonene: ×8)

Sabinene: ×1

Myrcene: ×3

Isoprene: ×2

Within SAR uncertainty, but with caveats

Predictions in scope of applicability but outside training set - interpolation

- Scope : H-atom, alkyl, β/γ-unsaturated, carbonyl, α-hydroxy, α-ether
- Most relevant substituent combination where already calculated.
- Use new data to validate SAR performance at chosen level of theory

  Extra calculations on CI of interest based on field data and model needs

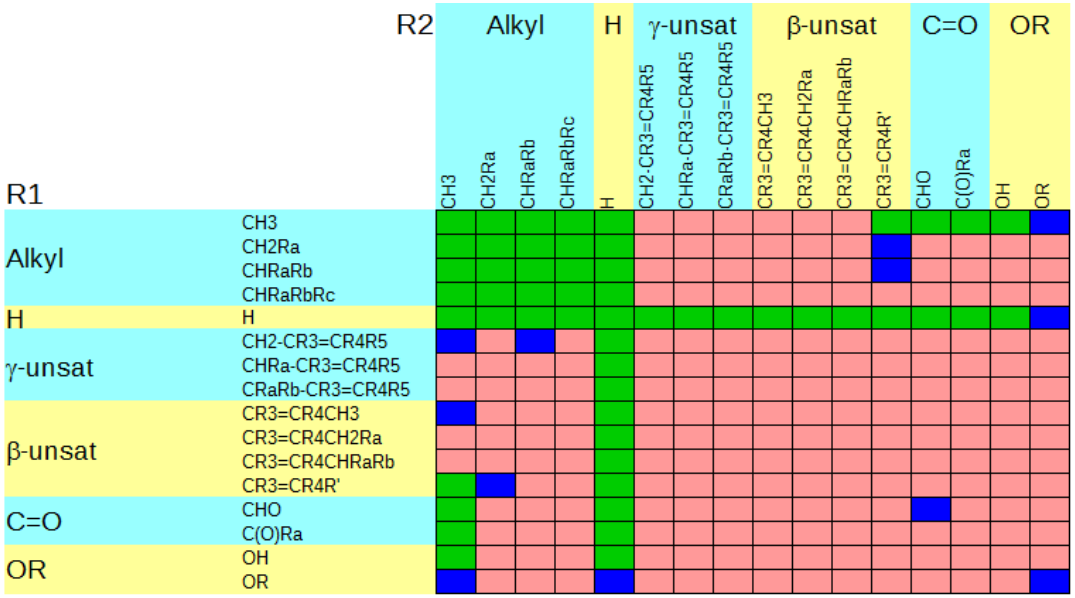


Fig: Matrix of all substituent combinations

Each CI compound:

Full conformer analysis

Several mechanisms

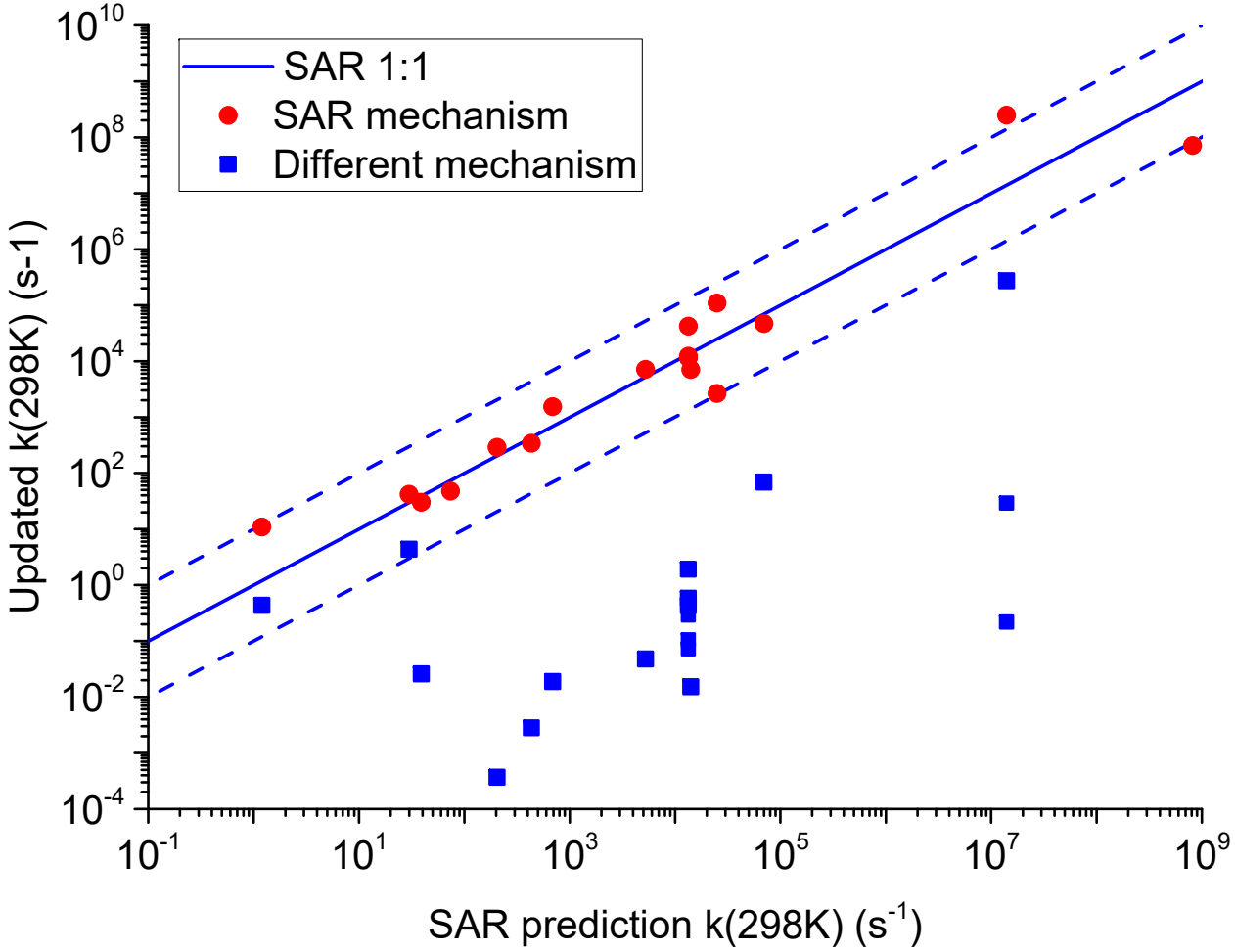
(19 actively studied)

Each reaction:

All reactant/TS conformers

Tunneling

This new data only tests quality of interpolation, not absolute performance



New k(298K) within x4 of SAR (advertised uncertainty x10)

Largest Δ : *syn*-*Z*-CH=CH-CH_3 (1.4×10^7 s^{-1} → 2.5×10^8 s^{-1})

Predicted mechanism: always correct

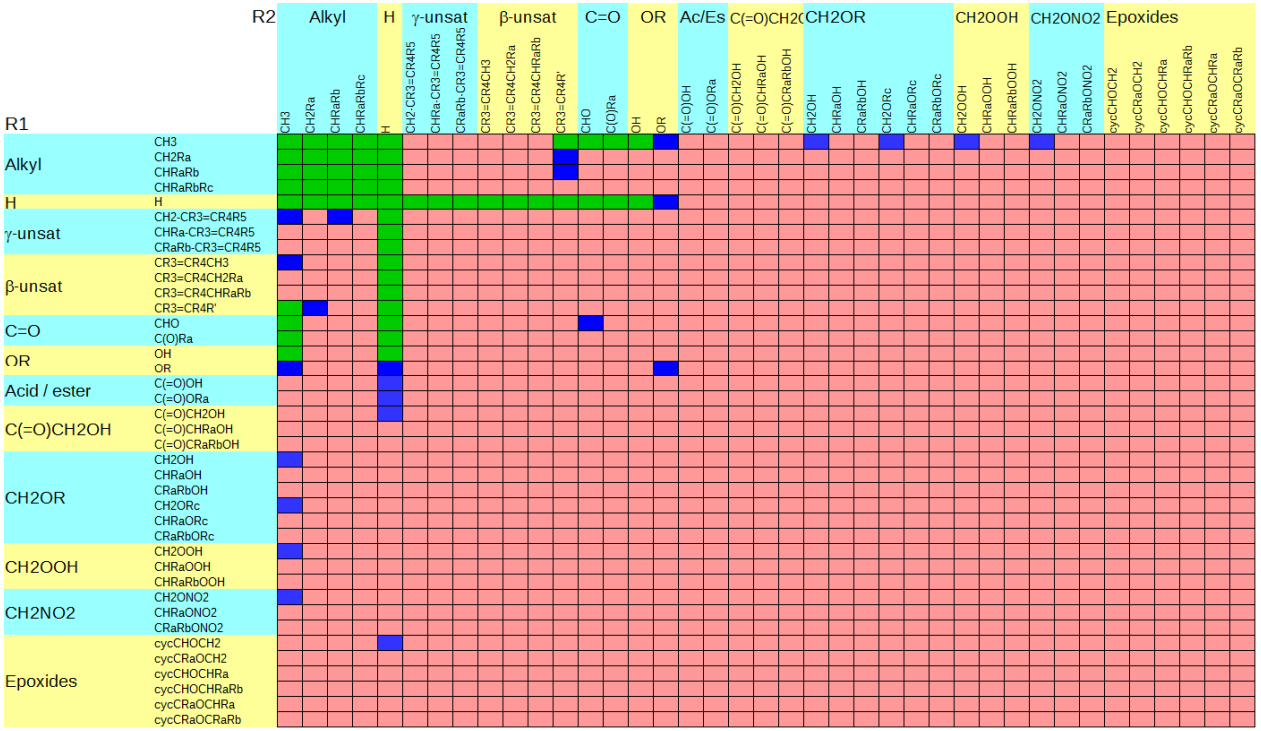
Predictions outside scope of applicability - extrapolation

Application of SAR to substituents not included in SAR → guesswork

Simplistic application: treat everything as "non-cyclic R, C=C, or C=O"

Extension needed for :

alcohols, ethers, acids, esters, hydroperoxides, epoxides, nitrates, ...



Matrix of all substituent combinations - combinatorial nightmare

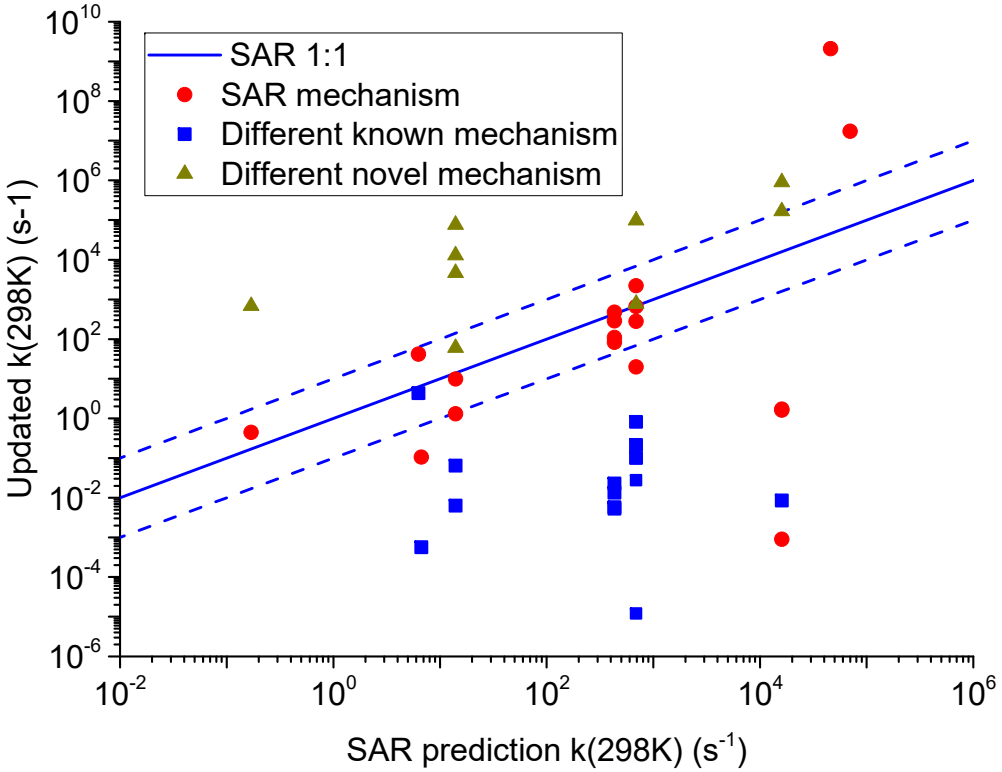
Extrapolation works fairly good in few cases (e.g. nitrates)

Orders of magnitude off in some cases (e.g. α-ethers)

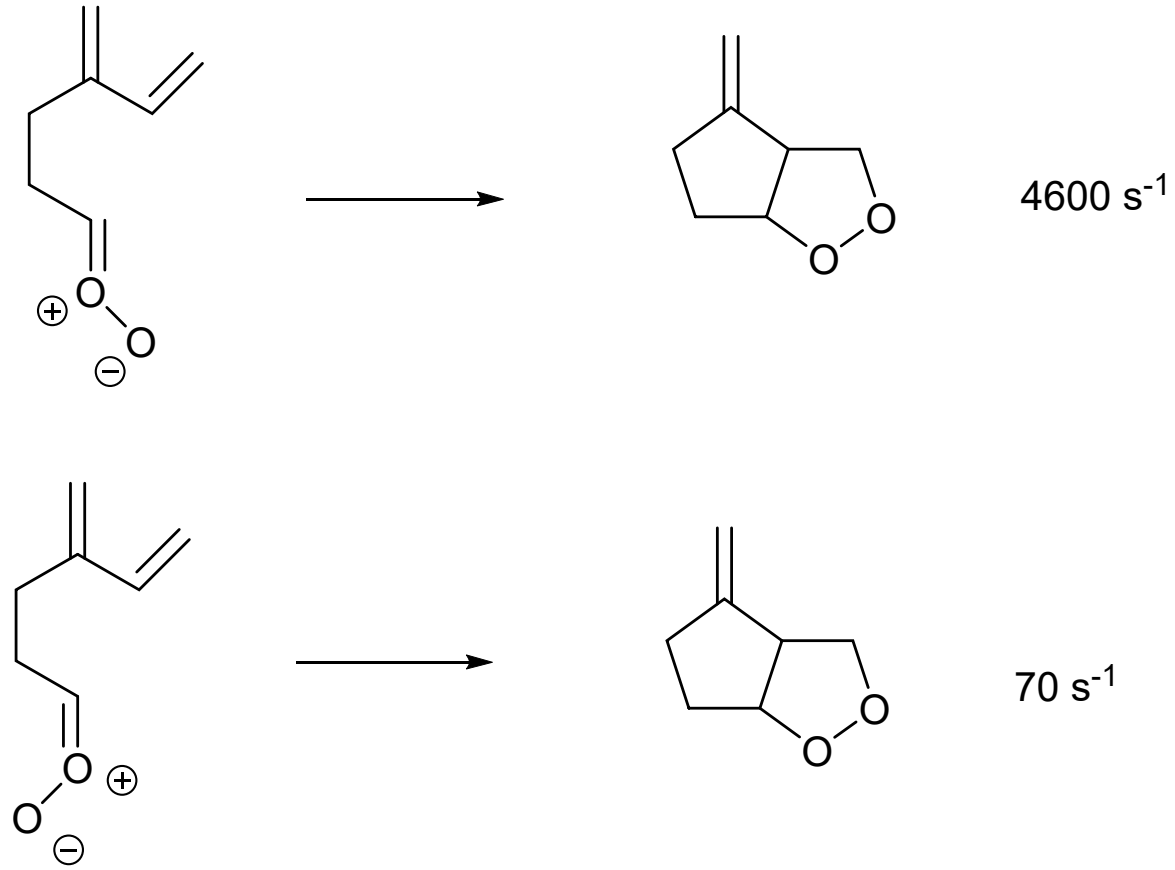
Totally new mechanism at play for some (-ROH, -ROOH, -RCOOH)

$\rightarrow$ extrapolation carries a high uncertainty

Extrapolation is not a "fair" metric, but an unavoidable use case



Example of different mechanism: bicyclic peroxide ring closure



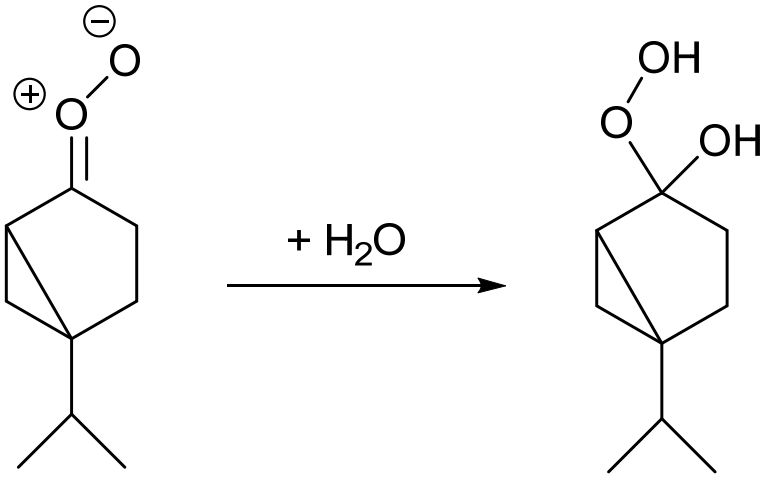
Z/E-Myrcene-Cl

Deng et al. 2018

Predictions for "uniques"

Some compounds have a one-of-a-kind structure → exception to any SAR

Possible example : Sabinene with rare 3-membered ring



E-sabinene-Cl	48 kJ/mol	Almatarneh et al. 2018
E-β-pinene-Cl	22 kJ/mol	Vereecken et al. 2017
E-(C_2H_5)(iC_3H_7)COO	15 kJ/mol	SAR

CONCLUSIONS

Data collection: SAR evaluation group

- Improved database of experimental data on VOC initiation
  - VOC + OH / NO_3 / Cl / O_3 : Data compilation and recommendations
- Work initiated on database for theoretical kinetic data (TheoKinDB)
  - Currently focusing on alkoxy chemistry to kick-start development
- Literature data collection on best available thermodynamic group values

SAR development

- Several new SARs have come available recently for key reaction classes
- SAR evaluation group: working on photolysis
  - Recommendation for estimate of cross sections and quantum yields
- Theoretical calculations are good supplements to experimental data
  - Well-suited for systematic studies
  - Provides access to experimentally difficult reactions

Evaluation of SAR and estimation methods : SAR evaluation group

- Progress on evaluation of VOC initiation reaction SARs,
  - Comparison of SARs / Comparison of SAR implementations / Validation
- Progress on evaluation of alkoxy radical SARs
  - Data collection / preliminary SAR implementations
- Progress on evaluation of CI chemistry SAR
  - Several metrics are examined
  - Extrapolation is probably unavoidable, but carries large uncertainty
  - Theory-based predictions are useful to validate the SARs

Model evaluation: SAR evaluation group

- Evaluation how photolysis is used in models / recommendations