EurJOC
European Journal of Organic Chemistry

Research Article
doi.org/10.1002/ejoc.202400949

Chemistry Europe
European Chemical Societies Publishing

www.eurjoc.org

Case Studies of Dimensionality in Chemical Data

Alex Blokhuis*[a, b, c, d] and Robert Pollice*[a]

The ambition to relate intrinsic features of chemical data to the underlying chemical reaction networks (CRNs) is not new, but has experienced only modest success. This may partly be attributed to a lack of theoretical groundwork connecting idealized theory to actual experimental data with added complexity. In particular: i) many CRNs have species that cannot be directly observed experimentally; ii) the apparent number of underlying reactions is a function of the resolution of the data; iii) chemical phenomena can change the number of discernable independent processes of the data. In this work, we illustrate the application of the recently introduced concept of data dimension, which quantifies the linearly independent dimensions the system composition in a CRN can change. We perform case studies inspecting the dimensionality of chemical data characterizing CRNs, and outline how it can be used for mechanistic interpretation. In some instances, these extended considerations allow us to directly recover the CRN proposed in the literature without any fitting. This demonstrates that, with incomplete information, important clues about CRN structure can still be recovered. Additionally, our approach detects critical subtleties, preventing important candidate reactions from being discarded in mechanistic studies.

Introduction

Chemical reaction networks (CRNs) constitute the core of mechanistic models for catalytic reactions such as hydroformylation[1] and complicated atmospheric processes such as photochemical ozone production.[2] The ability to deduce complex CRNs directly from experimental data in a systematic manner influences our understanding and prediction capabilities. Furthermore, the extent to which we can build reaction networks à la carte, for instance in chemical reservoir computation,[3] DNA computing,[4] and enzymatic reaction networks,[5] is constrained by our capacity to capture and validate the CRNs we end up constructing.

Several cues in data are commonly used to deduce CRNs. Isosbestic points,[6,7] for instance, are used to characterize the absence of side reactions. Keeping track of the mass balance allows spotting hidden reactions.[8] Phase rules relate the number of phases to the number of components and degrees of freedom in a system.[9]

Significant advances have been made in automated computational exploration[10–12] of CRNs, and in both analytical techniques[13] and automation[14] to monitor an ever larger number of species in ever smaller concentrations simultaneously. In addition, hybrid computational and analytical approaches have emerged, for instance combining mass spectrometry with simulations based on density functional theory.[15,16] However, systematic approaches to agnostically deduce structural information about underlying CRNs from experimental data – without prior knowledge about the system – are currently limited to small CRNs and rely on simplified models.[17–19]

For systematic CRN elucidation to scale to large CRNs in a robust manner, a solid theoretical foundation is indispensable. CRN theory provides this foundation[20] but has largely been constructed under the assumption of a known CRN,[21] which tends to be the exception rather than the rule in the study of chemical reactions and their mechanisms. Recently, relying on the concept of the dimension of CRN data,[22] we formalized the impact of common experimental complexities such as irreversibility and so-called concealed, that is undetectable, species.[23] This provides the basis for analyzing CRN data to distinguish feasible from infeasible mechanistic models from the first experiment.

In this work, we test the utility of these approaches by applying both CRN theory and the analysis of the dimensionality of experimental data to four diverse case studies in organic chemistry. Our framework enables the extraction of hitherto overlooked structural information of mechanisms from minimal experimental data. In some instances, this enables direct CRN elucidation, in other instances further clues are needed for complete CRN deduction (similar to how molecular structures may be elucidated by a combination of techniques). Our findings illustrate the power and generality of systematic CRN

[a] Dr. A. Blokhuis, Dr. R. Pollice
Stratingh Institute for Chemistry, University of Groningen,
Nijenborgh 3, 9747 AG Groningen, The Netherlands
E-mail: alex.blokhuis@hotmail.com
r.pollice@rug.nl

[b] Dr. A. Blokhuis
Groningen Institute for Evolutionary Life Sciences,
University of Groningen,
Nijenborgh 4, 9747 AG Groningen, The Netherlands

[c] Dr. A. Blokhuis
University of Strasbourg & CNRS, UMR7140,
67083 Strasbourg, France

[d] Dr. A. Blokhuis
Instituto IMDEA Nanociencia,
Calle Faraday 9, 28049 Madrid, Spain

Supporting information for this article is available on the WWW under https://doi.org/10.1002/ejoc.202400949

deduction for mechanistic studies and lay the foundation to more general application.

Chemical Reaction Network Data

We will work with either series of concentrations directly or properties that can be linearly related to concentrations such as UV-VIS, IR, or NMR signals. These concentration series are either related in time, characterizing the time evolution of chemical reaction networks, or related through small perturbations applied to the underlying systems such as the gradual addition of chemical species.

When monitoring the concentrations of s species in a CRN, which are collected in a system composition vector $[\mathbf{X}] = ([X_1], [X_2], \ldots, [X_s])^T$, it can be directly related to the data dimension *d*. The composition can be expressed in terms of an initial condition composition vector $[\mathbf{X}]^0$, and *d* linearly independent[22] extents of reaction $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(d)}$,

$$[\mathbf{X}] = [\mathbf{X}]^0 + \sum_{k=1}^{d} \xi^{(k)} \mathbf{\Delta}_k, \tag{1}$$

where $\mathbf{\Delta}_k$ corresponds to signed stoichiometric coefficients of a reaction. Intuitively, *d* is the number of linearly independent directions in which the system composition can change. For a deeper discussion of data dimension, we refer to the Supporting Information and the Ref. [23].

Various spectroscopic techniques such as UV-VIS and IR provide signals that are linearly related to concentrations via the Lambert–Beer law.[24] Absorbance A_λ at wavelength λ is obtained by multiplying the concentration of each species with its molar extinction coefficient $\varepsilon_k(\lambda)$ at λ,

$$A_\lambda = L_p \sum_{k=1}^{s} [X_k]\ \varepsilon_k(\lambda) \tag{2}$$

$$= L_p [\mathbf{X}] \cdot \boldsymbol{\epsilon}(\lambda) \tag{3}$$

where L_p is a path length. Similar to Eq. (1), spectral changes can then be written in terms of (up to Ref. [25]) *d* changes due to independent reactions, which corresponds to the corresponding data dimension.

Each spectroscopic technique differs in its ability to differentiate between distinct species and characterize processes taking place at specific timescales. To enhance their differentiating capacity, they may be coupled with separation methods such as chromatographic techniques. In doing so, some species may only be indirectly visible, e.g., loosely bound complexes may disassemble during separation and only be observed as their constituents. Accordingly, what is visible to any analytical technique has direct consequences to the dimension of the data it provides.

Fundamental Law for Dimension

In Eq. (1), dimension *d* corresponds to the number of observable independent reactions. A system may have more observable reactions than the value of $d_\circ$, i.e., some combinations of reactions accomplish identical composition changes. To illustrate, in the CRN:

$$X_1 \underset{1}{\leftrightharpoons} X_2 \underset{2}{\leftrightharpoons} X_3 \underset{3}{\leftrightharpoons} X_1, \tag{4}$$

where reaction numbers are marked under the reaction arrows, reaction R_3 leads to the same change as both R_1 and R_2 in reverse. Performing all three reactions successively in the forward direction leaves the composition unchanged, and is called a cycle. A fundamental law relates the theoretical dimension $d_\circ$ of a CRN to reactions, cycles,[20,22,26] species, and linear constraints on their abundances:

$$d_\circ = s_\circ - \ell_\circ = r_\circ - c_\circ \tag{5}$$

$s_\circ$ number of species (X_1, …, X_s)
$\ell_\circ$ number of stoichiometrically conserved quantities
$r_\circ$ number of reactions (R_1, …, R_r)
$c_\circ$ number of cycles

This law captures the theoretical dimensionality CRN data can have, assuming i) reaction rates are independent, ii) all species are independently observed, and iii) sufficient resolution to quantify them.[23] In experimental practice, chemical data frequently does not match all these assumptions. For instance, irreversible reactions can have collinear rates, and some species may be hidden or indistinguishable, leading to an alternative formulation that is immediately applicable to experimental data:[23]

$$d = s - \ell = r - c \tag{6}$$

s number of observable species
ℓ number of conserved quantities in observable species
r number of observable independent reactions
c number of observable cycles

This extension introduces additional conserved quantities, which provide key mechanistic information.

Conserved Quantities

In general, a linear conserved quantity *L* for a CRN is a linear combination of concentrations $[X_k]$ with coefficients ℓ_k that remains constant with respect to time:

Research Article
doi.org/10.1002/ejoc.202400949

$$\sum_{k=1}^{n} \ell_k[\mathrm{X}_k] = L \tag{7}$$

$$\frac{dL}{dt} = 0 \tag{8}$$

Among the most elementary conserved quantities are mass balances, linear combinations of species amounts with integer coefficients $\ell_k^{(i)}$ conserved by reactions, for instance:

$$\mathrm{AB} \rightleftarrows \mathrm{A} + \mathrm{B} \tag{9}$$

$$L^{(1)} = [\mathrm{AB}] + [\mathrm{A}] \tag{10}$$

$$L^{(2)} = [\mathrm{AB}] + [\mathrm{B}] \tag{11}$$

Experimentally, concentrations and related quantities are measured with limited precision. Hence, there is some level of experimental error η. Conserved quantities are then only conserved up to the experimental error:

$$|\Delta L| \leq \eta \tag{12}$$

This implies that for details to manifest in the dimension d of the underlying data, they need to be observed with sufficient resolution, i.e., with low enough η. For larger values of η, additional quantities may appear as conserved that would not appear as conserved with higher precision.[28]

Isosbestic Points

The dimensionality of a CRN has immediate implications for the structure of the data that can be observed experimentally, especially for low numbers of dimensions. For a CRN with $d = 1$ that is monitored spectroscopically via UV-VIS or IR, the composition can be characterized in terms of a single reaction extent ζ and an initial condition A_λ^0. The absorbance $A_\lambda(\zeta)$ then becomes:

$$A_\lambda(\zeta) = A_\lambda^0 + \zeta \Delta A_\lambda \tag{13}$$

where ΔA_λ is the change in absorbance associated with the reaction extent ζ. From this expression, we can directly see the following properties, which are intrinsic to CRN data with $d = 1$ (cf. Figure 1):

- **Isosbestic point**: When two distinct spectra intersect at wavelength λ^*, all of them do. That is, $A_{\lambda^*}(\zeta_1) = A_{\lambda^*}(\zeta_2)$, $(\zeta_1 \neq \zeta_2)$, then $\Delta A_{\lambda^*} = 0$ and we will have an intersection at λ^* for all spectra (i.e., for all ζ).
- **Strict non-intersection**: When two distinct spectra do not intersect at wavelength λ^*, no distinct spectra do. As $\Delta A_{\lambda^*} \neq 0$ implies $A_{\lambda^*}(\zeta_1) \neq A_{\lambda^*}(\zeta_2)$ if $\zeta_1 \neq \zeta_2$.

While $d_\circ = 1$ guarantees the above properties, deducing $d = 1$ from spectra requires that enough species absorb at the wavelengths under consideration. We can use this to our advantage to discover more structure.

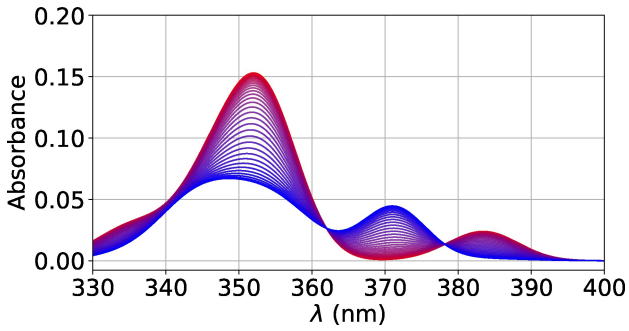


Figure 1. Artificially generated spectra over the course of a hypothetical reaction for a system with $d_0 = 1$, illustrating two isosbestic points, and strict non-intersection everywhere else. Red curve: reaction start, blue curve: reaction end.

From Eq. (13), we can directly see that an isosbestic point implies a conservation law with positive coefficients:

$$\frac{dA_\lambda}{d\zeta} = 0 \tag{14}$$

$$\sum_{k=1}^{s} \frac{\varepsilon_k(\lambda)}{\varepsilon^\circ}[\mathrm{X}_k] = L \tag{15}$$

where $\varepsilon^\circ = 1$ is a constant with identical units as $\varepsilon_k(\lambda)$ that is introduced to obtain dimensionless coefficients. Such conservation laws with $d = 1$ have several other visual signatures. For instance, in a reaction:

$$\mathrm{A} \rightleftarrows \mathrm{B} + \mathrm{C}, \tag{16}$$

changes in [A] exactly mirror changes in [B] and [C]:

$$[\mathrm{A}] + [\mathrm{B}] = L_2 \tag{17}$$

$$\Delta[\mathrm{A}] = -\Delta[\mathrm{B}] \tag{18}$$

When we observe the concentrations of A and B to counterbalance each other (cf. Eq. (18)), we can immediately infer a $d = 1$ conservation law. We could make the same inference from the observation of an isosbestic point.

In general, a $d = 1$ conservation law can be detected by plotting absorptions at two distinct wavelengths λ_1 and λ_2 against each other. All points will fall on a straight line:

$$\begin{pmatrix} A_{\lambda_1} \\ A_{\lambda_2} \end{pmatrix} = \begin{pmatrix} A^0_{\lambda_1} \\ A^0_{\lambda_2} \end{pmatrix} + \zeta \begin{pmatrix} \Delta A_{\lambda_1} \\ \Delta A_{\lambda_2} \end{pmatrix} \tag{19}$$

For $d = 2$, points no longer form a straight line, but, when plotting three wavelengths against each other, all data fall on a plane. This feature generalizes to arbitrary dimensions but is increasingly hard to visualize. A more general approach to

Research Article
doi.org/10.1002/ejoc.202400949

reveal these signatures is performing singular value decomposition (SVD) of mean-subtracted data.[23]

Arbitrary Dimensions

Our goal is to analyze experimental concentration series to unravel $d_\circ$. We define a matrix $\mathbb{M}$ with data monitoring the concentrations of s species at n time points $(t_1, t_2, .., t_n)$:[29]

$$\mathbb{M} = \begin{pmatrix} [X_1](t_1) & [X_2](t_1) & .. & [X_s](t_1) \\ [X_1](t_2) & [X_2](t_2) & .. & [X_s](t_2) \\ \vdots & \vdots & \ddots & \vdots \\ [X_1](t_n) & [X_2](t_n) & .. & [X_s](t_n) \end{pmatrix} \quad (20)$$

We substract the mean of each species to get:

$$\Delta\mathbb{M} = \begin{pmatrix} \Delta[X_1](t_1) & \Delta[X_2](t_1) & .. & \Delta[X_s](t_1) \\ \Delta[X_1](t_2) & \Delta[X_2](t_2) & .. & \Delta[X_s](t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta[X_1](t_n) & \Delta[X_2](t_n) & .. & \Delta[X_s](t_n) \end{pmatrix}$$

where:

$$\Delta[X_k](t_i) = [X_k](t_i) - \frac{1}{n}\sum_{j=1}^{n}[X_k](t_j) \quad (21)$$

By performing SVD of the s-by-n matrix $\Delta\mathbb{M}$, we can decompose the time trajectories of each species into a linear combination of independent contributions of decreasing relative weight. We can use this to determine the discernible dimensions of CRN data by finding the number of independent contributions that separates noise from genuine time trajectories. Then, we let the discernable dimension d be the number of contributions above a threshold value that directly results from the experimental precision. Consequently, in our case studies, d will be qualitatively evident from visual inspection of the results of SVD. A detailed derivation is provided in the Supporting Information.

The discernable dimension d is an intrinsic property of the data being analyzed. For the trivial case of data that only consists of noise, the discernable dimension $d = 0$. As soon as there is 1 independent reaction that can be discerned and is significantly different from noise, $d = 1$. The upper bound of the discernable dimension d is dictated by the theoretical dimensionality $d_\circ$ of underlying CRN.

Case Studies

In this section, we will go through several case studies of increasing mechanistic complexity and investigate what mechanistic information can be deduced based on analyzing the dimensionality in experimental data. Whereas in the simplest case study, all the mechanistic information obtained can be readily obtained via alternative kinetic data analysis methods, for the more complex case studies our methodology allows extracting information that is otherwise inaccessible. All the data analyzed here was originally obtained through experiments and is taken from the literature. We will refer to specific species as X_n where n is the number indicated below the structures in the corresponding figures.

Case Study 1: Molecular Walkers

First, we will inspect data extracted from Ref. [30] regarding molecular walkers. We selected this as our first case study because the underlying mechanism is simple and the analysis of data dimension alone allows to deduce the CRN proposed in the literature, but without any data fitting. Figure 2a shows three isomeric structures of a molecular walker, a molecule harboring a moiety that can be exchanged among binding sites.

Figure 2b shows kinetic data monitoring walkers X_1 to X_3 in time using ^{1}H-NMR.[30] Let us inspect properties of the data that allow us to propose a CRN and eliminate alternative hypotheses without fitting specific CRNs, independent of molecular structures. In the absence of chemical reasons for irreversibility (e.g., photochemistry, radiochemistry, strongly favored reactions), for

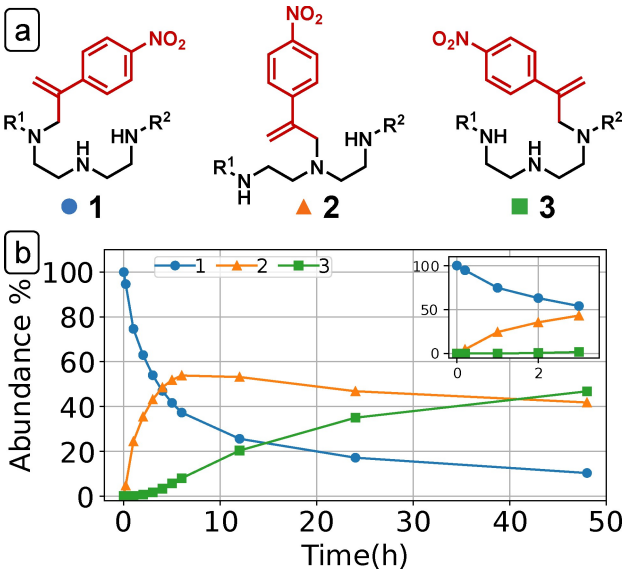


Figure 2. Case Study 1: (a) Molecular walker structures. (b) Relative abundance of species X_1, X_2, and X_3 against time, taken from the Ref. [30]. Inset: close-up of early reaction times. Initially, changes in X_3 are not discernable. Transiently, X_1 and X_2 counterbalance and, thus, follow a $d = 1$ conservation law.

ease of instruction, candidate reactions are introduced as reversible reactions.

As we monitor three species, $s = 3$. Using SVD on the full data, we find $d = 2$ through 2 smooth trajectories, and 1 noisy one, which coincides with mass conservation:

$$L = [X_1]+[X_2]+[X_3], \tag{22}$$

i.e., a stoichiometric conservation law, and $\ell = 1$.

Confined to short times, only the consumption of species X_1 and the production of species X_2 can be resolved. Consumption and production counterbalance, i.e., intermittently, the following quantities appear to be conserved:

$$L_1^* = [X_1] + [X_2] \tag{23}$$

$$L_2^* = [X_3] = 0 \tag{24}$$

Deviations from L_k^* are initially too small to be distinguishable from noise. This also follows from a $d = 1$ conservation law in the initial reaction period (cf. Figure 2b). Spectroscopically, it could give rise to a genuine transient isosbestic point, a feature also commonly used to study the isomeric states for photochemical motors[31] and switches.[32]

An interpretation consistent with these observations is that X_1 reacts to form X_2, for instance according to the following equation:

$$X_1 \leftrightarrows X_2 \tag{25}$$

This is the simplest possibility, but not the only one. Any reaction that accomplishes the same stoichiometric transformation is, *a priori*, consistent with the same conservation laws. More complex candidate reactions that are still consistent with these observations would be:

$$X_1 + X_2 \rightleftarrows 2\ X_2 \tag{26}$$

$$2\ X_1 \rightleftarrows X_2 + X_1 \tag{27}$$

Distinguishing these cases requires more information. For instance, reaction (26) requires initial X_2 and, thus, should not occur here. Additionally, the presence of hidden species, denoted as ■, cannot be excluded. Eq. (25) may derive from $X_1 + ■ \leftrightarrows X_2 + ■$. Yet, CRN descriptions with missing species are still insightful and common.

At longer reaction times, species X_2 accumulates, and species X_3 starts to form in detectable quantities. L_1^*, L_2^* are no longer conserved, but their sum still is:

$$L = L_1^* + L_2^* = [X_1]+[X_2]+[X_3] \tag{28}$$

The simplest candidate network consistent with these conservation laws is:

$$X_1 \leftrightarrows X_2 \leftrightarrows X_3, \tag{29}$$

which coincides with the CRN proposed in Ref. [30].

To highlight how this is useful, let us consider hypotheses that we can eliminate with this information, using only the data in Figure 2b and its analysis in Figure 3. This is an important question to answer for characterizing whether a species is a motor or a switch.[33] As there are three isomers to be considered, there are $3 \cdot 2 \cdot 0.5 = 3$ distinct pairs of isomers, 3 candidate isomerizations, and, thus, $2^3 = 8$ theoretical candidate CRNs (see also the Supp. Matt. for candidate CRN combinatorics).

Let us consider the following candidate CRNs:

$$X_2 \leftrightarrows X_1 \leftrightarrows X_3 \tag{30}$$

$$X_2 \leftrightarrows X_1 \leftrightarrows X_3 \leftrightarrows X_2 \tag{31}$$

At short times, both networks behave like:

$$X_2 \leftarrow X_1 \rightarrow X_3 \tag{32}$$

We would measure $s = 3$ species at short times. This network exhibits co-production,[23] and would give rise to a non-integer conservation law, with the coefficients o_1 and o_2 being real numbers:

$$L^{(2)} = o_2[X_2] - o_1[X_3] \tag{33}$$

We would expect $\ell = 2$ conserved quantities in total. As these are not observed, these networks can be excluded. Additionally, to explain the data over the entire measurement time, CRNs consisting of only one reaction can be excluded as this would lead to $d = 1$ with two conservation laws.

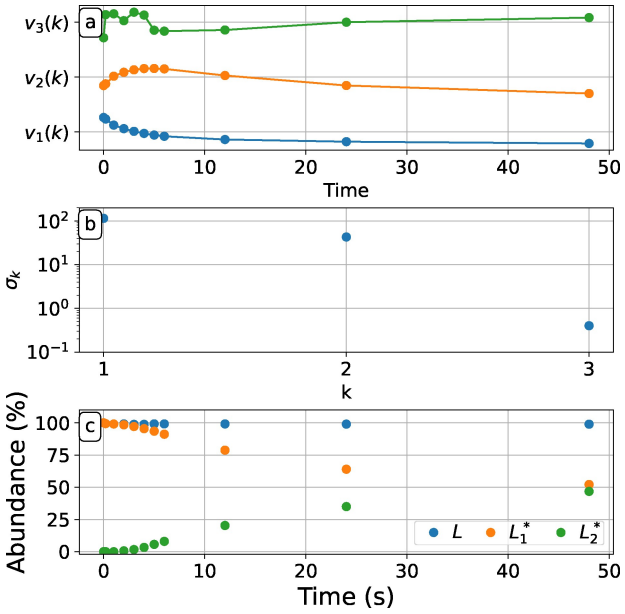


Figure 3. Case Study 1: (a) Singular vectors with 2 non-noisy singular vectors, (b) singular values, (c) conservation law L, and transient conservation laws L_1^*, L_2^*.

Research Article
doi.org/10.1002/ejoc.202400949

We may not, *a priori*, know that an isomerization takes place, and so reactions with more than one reactant or product are in general relevant. For instance, the single reaction:

$$X_1 \leftrightarrows X_2 + X_3 \tag{34}$$

satisfies two conservation laws:

$$L^{(1)} = [X_1]+[X_2] \tag{35}$$

$$L^{(2)} = [X_1]+[X_3] \tag{36}$$

Here, we can rule this CRN out. At short times, the data matches $L^{(1)}$ but not $L^{(2)}$. At long times, the data in Figure 3b is too high-dimensional ($d = 2$) as Eq. (34) would provide data with $d = 1$. Alternatively, the CRN:

$$X_1 \leftrightarrows X_2 + X_3,\ X_2 \leftrightarrows X_3 \tag{37}$$

would transiently provide data with $d = 1$, and $d = 2$ would be expected for the full data, so far consistent with the observed data. At long reaction times, however, the unique conserved quantity would be:

$$L^{(3)} = 2\,[X_1]+[X_2]+[X_3] \tag{38}$$

For the observed data there is a unique conservation law $L = [X_1]+[X_2]+[X_3]$ that holds at all times. Since $L^{(3)}$ is different, we can also eliminate this candidate CRN.

Thus far, we used concentrations, not molecular structure, to infer the underlying CRN, highlighting how much mechanistic information is contained in this data. In practice, we often know more. For instance, the composition of our species may be characterized spectroscopically. We could use the information that all species involved are isomers to dismiss some candidate CRNs like Eqs. (34) and (37). Generally, molecular structure and conserved quantities give complementary insight (cf. Case Study 4).

Case Study 2: Ruthenium Hydrides

Our next case study involves conservation laws similar to those in Case Study 1. However, CRN deduction is considerably more non-trivial, which is why we selected it as our second case study. Additionally, it demonstrates that the insight obtained from dimensional analysis tends to increase with mechanistic complexity. We consider 4 N-heterocyclic carbene (NHC)-ligated ruthenium hydride complexes (Figure 4a), 3 of which are isomeric, which interconvert photochemically[35] at low temperature (cf. Figure 4b).

Due to the low temperature, thermal reactions are slow and, hence, at the timescale of the experiment, some isomerizations may not be appreciably observed. As species interconvert in a stoichiometric one-to-one relationship, we expect the following mass balance to hold at all times:

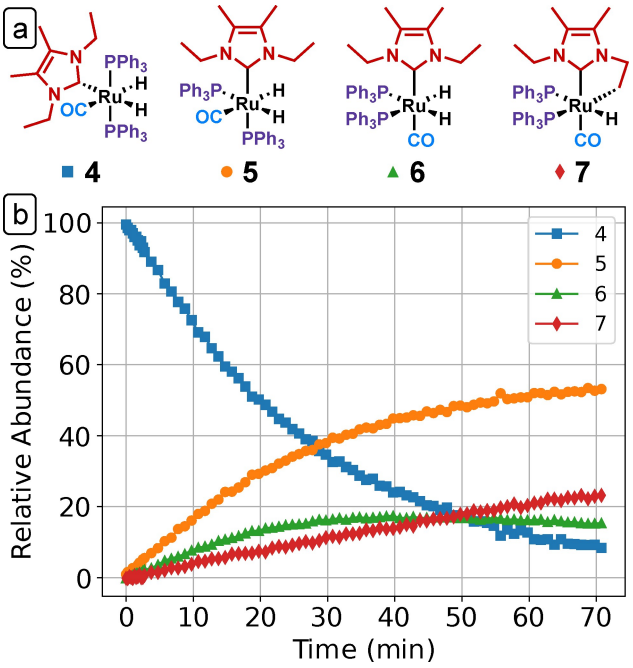


Figure 4. Case Study 2: (a) Structures of NHC-ligated ruthenium hydride complexes. (b) Abundance of 4 NHC-ligated ruthenium hydride complexes under illumation at $T = 223$ *K*, monitored using ^{1}H-NMR. Kinetic data were taken from the Ref. [35].

$$L_1 = [X_4]+[X_5]+[X_6]+[X_7] \tag{39}$$

Formation of X_7 releases H_2, which is a hidden species. Looking at the first 20 minutes, we find that our $s = 4$ observable species yield data with $d = 1$ dimensionality (cf. Figure 5a–b). This requires that all changes are collinear, which is consistent with three irreversible reactions of X_4 forming X_5, X_6, and X_7, respectively. These reactions are not linearly independent and can be merged into one effective reaction,[23] which we refer to as R_4 and indicate its number under the reaction arrow. The corresponding fractional stoichiometric coefficients p_1, p_2, and p_3 are obtained directly from SVD:

$$\begin{aligned} &X_4 \underset{4}{\rightarrow} p_1\,X_5 + p_2\,X_6 + p_3\,X_7, \\ &p_1 \approx 0.575,\ p_2 \approx 0.275, vp_3 \approx 0.150 \end{aligned} \tag{40}$$

From this, we get two additional conservation laws, L_2^* and L_3^*, due to co-production:

$$L_2^* = p_2[X_5] - p_1[X_6] \tag{41}$$

$$L_3^* = (1 - p_1 - p_2)\,[X_5] - p_1[X_7] \tag{42}$$

In Case Study 1, all co-production conservation laws were transient. Here, we find $d = 2$ observable dimensions for the full data (Figure 5c–d). From the singular vectors, we identify the conserved quantity:

Research Article
doi.org/10.1002/ejoc.202400949

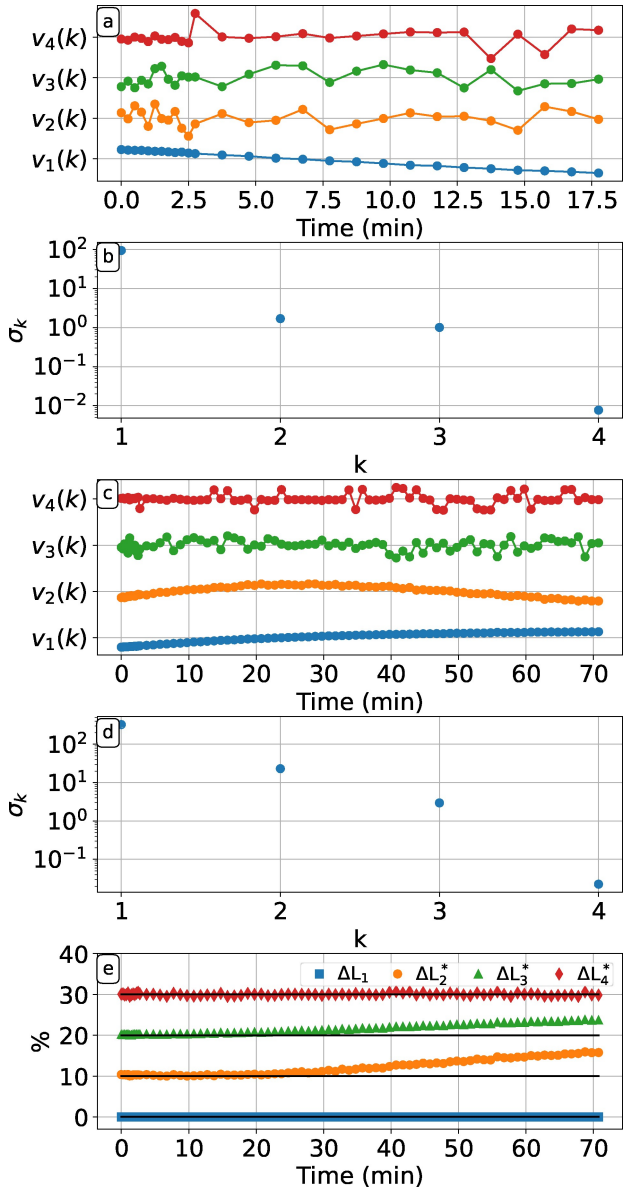


Figure 5. Case Study 2: (a–b) SVD of the data in Figure 4b. For the first 20 minutes, we find dimensionality $d = 1$ and $\ell = 3$ conservation laws. (c–d) SVD of the full data, revealing $d = 2$ and $\ell = 2$. (e) Observed variation for (transiently) conserved quantities L_1, L_2^*, L_3^*, and L_4^*. Variations are depicted with an ordinate offset of 10%, 20%, and 30% for ΔL_2^*, ΔL_3^*, and ΔL_4^*, respectively. Both L_2^* and L_3^* are only initially conserved.

$$L_4^* = p_1[X_4] + [X_5] = p_1 L_1 + L_2^* + L_3^*, \tag{43}$$

which is a linear combination of L_1, L_2^*, and L_3^* (Figure 5e). From L_4^*, we deduce that, at the timescale of the experiment, X_5 does not react further. Accordingly, there is no other reaction in the CRN that forms X_4 or X_5.[36] As data dimension d changes from 1 to 2 at longer timescales, the same cannot be true for X_5 and X_6. The following two candidate reactions R_2 and R_3 remain, where the reaction number is indicated under the corresponding reaction arrow:

$$X_6 \xrightarrow[5]{} X_7 \tag{44}$$

$$X_7 \xrightarrow[6]{} X_6 \tag{45}$$

As the abundance of species X_6 eventually changes from increasing to decreasing, X_6 must react further, for which the only candidate reaction is R_5. X_7 keeps increasing and, hence, R_6 is slow. Thus, we propose the following minimal CRN, with rate constants noted over the reaction arrows:

$$X_4 \xrightarrow{k_{45}} X_5,\ X_4 \xrightarrow{k_{46}} X_6,\ X_4 \xrightarrow{k_{47}} X_7,\ X_6 \xrightarrow{k_{67}} X_7 \tag{46}$$

When R_3 is significant, the last reaction has a reverse:

$$X_4 \xrightarrow{k_{45}} X_5,\ X_4 \xrightarrow{k_{46}} X_6,\ X_4 \xrightarrow{k_{47}} X_7,\ X_6 \xrightarrow{k_{67}} X_7,\ X_7 \xrightarrow{k_{76}} X_6 \tag{47}$$

Letting k_4 be the total rate constant for first-order consumption of $[X_4]$, we can then directly estimate rate constants $k_{45} = p_1 \cdot k_4$, $k_{46} = p_2 \cdot k_4$ and $k_{47} = p_3 \cdot k_4$. We denote with an asterisk the abundances at the time where X_6 reaches its maximum, i.e., $[X_6]^*$. We can then find:

$$0 = k_{47}[X_4]^* - k_{67}[X_6]^* \tag{48}$$

Overall, from inspecting the data and its conservation laws, we could infer a candidate model and provide a quick estimate of its parameters. *A priori*, there are 12 possible candidates for irreversible photochemical isomerizations, and, hence, $2^{12} = 4096$ possible CRNs. Analysis of data dimensions for kinetic data under one set of reaction conditions reduced this to only 2 related candidate CRNs. This is in marked contrast to most current kinetic analysis methods that either need kinetic data under multiple conditions or are based on assuming specific kinetic models (*vide infra*).

Case Study 3: Buchwald-Hartwig Coupling

Next, we inspect data for a Buchwald-Hartwig coupling from the Ref. [38], depicted in Figure 6 with concentration changes for species X_8 to X_{11}. We selected this data because it showcases how dimensional analysis immediately uncovers data dimensionality that is unexpectedly small for a mechanism that would normally be expected to be best modeled as a multi-step catalytic process, thus providing guidance to kinetic data interpretation. One of the *a priori* expected products does not form and we can choose to ignore species X_{12}. When its concentration is introduced as a variable, it will satisfy a trivial conservation law:

$$L_4^* = [X_{12}] = 0 \tag{49}$$

We have $s = 4$ observed species, derived from peak areas after chromatographic separation. To a good approximation, we have four collinear curves, i.e., observed dimensions $d = 1$ and observed conserved quantities $\ell = 3$ (Figure 5a–b). Following

Research Article
doi.org/10.1002/ejoc.202400949

Chemistry Europe
European Chemical Societies Publishing

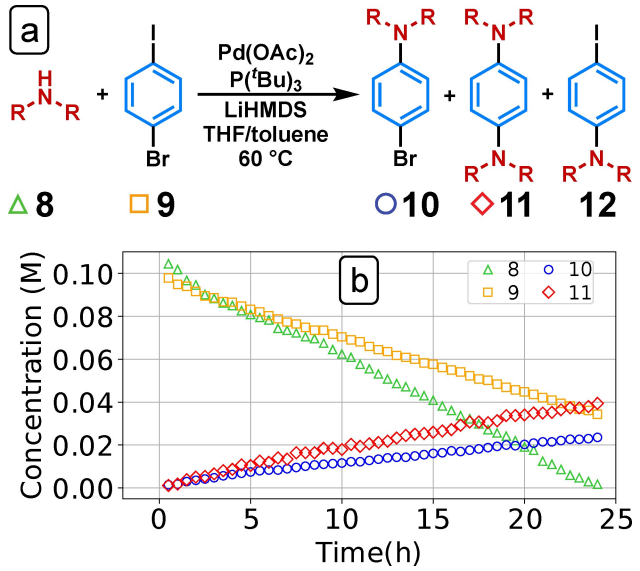


Figure 6. Case Study 3: (a) Buchwald-Hartwig amination performed in Ref. [38] (R = p-MeO-Ph). Compound **12**, was expected to form, but not observed. (b) HPLC reaction monitoring data for this reaction.

the molecular structure and mechanistic considerations from the original reference,[38] one would assume that doubly functionalized X_{11} would naturally form from monofunctionalized X_{10}. The direct reaction of X_9 with two molecules of X_8 to form X_{11} was not considered a plausible candidate reaction in the original work.[38] Accordingly, one would expect 2 independent reactions leading to $d = 2$ data. We would write:

$$X_8 + X_9 \rightarrow X_{10},\ X_8 + X_{10} \rightarrow X_{11} \tag{50}$$

This CRN has $\ell_\circ = 2$ stoichiometric conservation laws:

$$L_1 = [X_8] + [X_{10}] + 2\,[X_{11}] \tag{51}$$

$$L_2 = [X_9] + [X_{10}] + [X_{11}] \tag{52}$$

While one might anticipate independent successive reactions, X_{10} and X_{11} appear to form in a collinear fashion that appears to be zeroth order in reactants, with an emergent co-production conservation law:

$$L_3^* = [X_{10}] - q\cdot[X_{11}] = 0\ (q \approx 0.6) \tag{53}$$

Unlike L_1 and L_2, a conserved quantity caused by co-production can have negative and non-integer coefficients. Figure 7d confirms that these quantities are all conserved throughout the experiment. A single effective reaction consistent with these conserved quantities is (Figure 5c):

$$(2 - q)\,X_8 + X_9 \rightarrow \frac{q}{1+q}\,X_{10} + \frac{1}{1+q}\,X_{11} \tag{54}$$

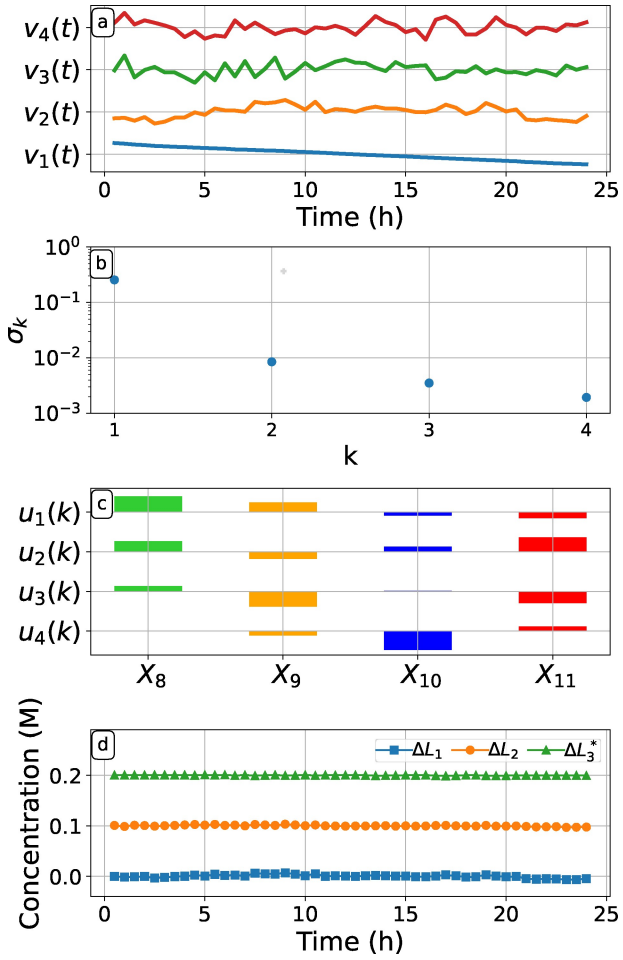


Figure 7. Case Study 3: SVD of the data in Figure 6b. (a) Trajectories $v_i(t)$, only $v_1(t)$ shows meaningful variation. (b) Singular values σ_k. The second singular value is comparable to noise. (c) Composition vectors $u_j(k)$. Since dimensionality $d = 1$, singular vector $u_1(k)$ coincides with the reaction stoichiometry. (d) Variation in conserved quantities L_1, L_2, and L_3^*. Variations are depicted with an ordinate offset of 0.1 M and 0.2 M for ΔL_2 and ΔL_3^*, respectively. All quantities are conserved throughout the experiment.

This reaction involves several species, including catalyst, and, to a good approximation, follows a rate law that appears to be zeroth order in both X_8 and X_9.

A more in-depth look via SVD (cf. Figure 7a–b) reveals a single non-noisy dimension, which accounts for 95% of the variation in the data. The low dimension suggests a mechanism that should simultaneously enable such a non-trivial outcome, yet be'simple' enough to yield experimental data with $d = 1$ (by default, additional reactions increase d). It is reasonable to expect that this mechanism involves species that are not monitored. Hence, the inferences that can be made based on the present data alone are limited.

To illustrate another mechanism by which the data dimension $d = 1$ and the conservation law in Eq. (53) may emerge, let us consider one alternative scenario. A dependence between eluted species may also be introduced when they were previously part of the same species, e.g., a complex or pseudophase. For instance, when the catalyzed reaction produces one net product complex:

$$11\ X_8 + 8\ X_9 \rightarrow (X_{10})_5(X_{11})_3, \tag{55}$$

then such a complex might disassemble during chromatography and give two distinct peaks. The effective variables $[X_{10}]^*$ and $[X_{11}]^*$, which are deduced from peak heights, then originate from the same species:

$$[(X_{10})_5(X_{11})_3] = \frac{1}{5}\,[X_{10}]^* = \frac{1}{3}\,[X_{11}]^* \tag{56}$$

In such a scenario, the conservation law L_3^* is not caused by co-production, but rather by the composition of a complex or pseudophase. Notably, based on further experiments and subsequent work, this reaction is currently thought to proceed through a ring-walking mechanism.[39]

Case Study 4: Thiophenol Chlorination

We now consider kinetic data for a thiol chlorination[40] where all but one species is monitored through ^{1}H-NMR (Figure 8a). Nevertheless, this one hidden species is critical, which is one of the reasons we selected this case study. Additionally, it also shows how dimensional analysis allows quantifying the concentration of a hidden species through an emergent conservation law. The following CRN was proposed in Ref. [40], with reaction numbers indicated below the reaction arrows:

$$\mathrm{ArSH} + \mathrm{Cl_2} \underset{7}{\rightleftharpoons} \mathrm{HCl} + \mathrm{ArSCl} \tag{57}$$

$$\mathrm{NCS} + \mathrm{HCl} \underset{8}{\rightleftharpoons} \mathrm{NHS} + \mathrm{Cl_2} \tag{58}$$

$$\mathrm{NCS} + \mathrm{ArSH} \underset{9}{\rightleftharpoons} \mathrm{NHS} + \mathrm{ArSCl} \tag{59}$$

$$\mathrm{ArSCl} + \mathrm{ArSH} \underset{10}{\rightleftharpoons} \mathrm{HCl} + \mathrm{Ar_2S_2} \tag{60}$$

$$\mathrm{Ar_2S_2} + \mathrm{Cl_2} \underset{11}{\rightleftharpoons} 2\ \mathrm{ArSCl} \tag{61}$$

From stoichiometry, $\ell = 4$ conservation laws arise, with observable species $s = 7$, observable reactions $r = 5$, and observable cycles $c = 2$. As any linear combination of conservation laws is also conserved, we can choose an orthogonal basis. A convenient choice with positive coefficients and a minimal number of species in each equation is:

$$L_1 = [\mathrm{NHS}] + [\mathrm{NCS}]$$

$$L_2 = [\mathrm{ArSH}] + [\mathrm{HCl}] + [\mathrm{NHS}]$$

$$L_3 = [\mathrm{ArSCl}] + 2\,[\mathrm{Ar_2S_2}] + [\mathrm{ArSH}]$$

$$L_4 = [\mathrm{ArSCl}] + [\mathrm{HCl}] + [\mathrm{NCS}] + 2\,[\mathrm{Cl_2}]$$

Without emergent conservation laws, i.e., when all species are monitored, we could extract L_1 to L_4 from the data, and consider candidate reactions that leave L_1 to L_4 unchanged. R_7 to R_{11} involve the fewest reactants and products that do so. This can be readily shown based on linear algebra, which is demonstrated in the Supporting Information.

This CRN was constructed through extensive experiments with various techniques. Here, we only consider the data of Figure 8b, where species were monitored via ^{1}H-NMR and Cl_2 cannot be observed. As it is typical for individual experiments that not all species can be monitored, let us consider some consequences of this hidden species.

Without Cl_2, we only see $s = 6$ species. This may either remove a conservation law or a dimension. For a closed system, the first missing species will always break a conservation law.[23] Consequently, L_4 is no longer conserved but L_1 to L_3 remain. When we rewrite the CRN above only in terms of visible species, the resulting CRN respects the remaining conservation laws. Reaction numbers with asterisk mark modified equations with hidden species ■:

$$\mathrm{ArSH} + \blacksquare \underset{7^*}{\rightleftharpoons} \mathrm{HCl} + \mathrm{ArSCl} \tag{62}$$

$$\mathrm{NCS} + \mathrm{HCl} \underset{8^*}{\rightleftharpoons} \mathrm{NHS} + \blacksquare \tag{63}$$

$$\mathrm{NCS} + \mathrm{ArSH} \underset{9}{\rightleftharpoons} \mathrm{NHS} + \mathrm{ArSCl} \tag{64}$$

$$\mathrm{ArSCl} + \mathrm{ArSH} \underset{10}{\rightleftharpoons} \mathrm{HCl} + \mathrm{Ar_2S_2} \tag{65}$$

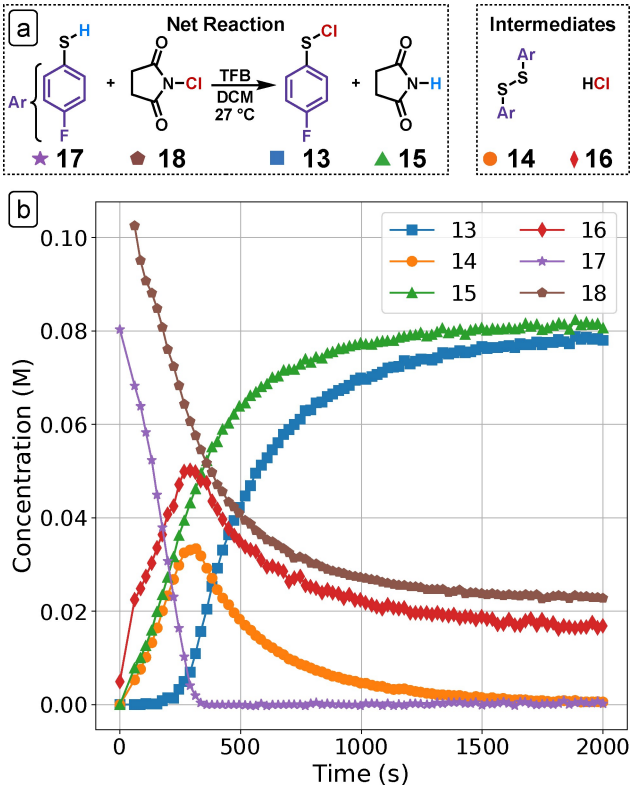


Figure 8. Case Study 4: (a) Thiophenol chlorination studied in Ref. [40]. Cl_2 is also produced, but not captured by ^{1}H-NMR. (b) Kinetic data for this reaction.[40] **18** is added in the first 30 seconds under agitation.

Research Article
doi.org/10.1002/ejoc.202400949

$$Ar_2S_2 + \blacksquare \underset{11}{\rightleftharpoons} 2\ ArSCl \tag{66}$$

In the corresponding concentration data, we can readily see a mass defect in reactions R_7^*, R_8^*, and R_{11}^*, which can be resolved through the inclusion of Cl_2. Upon inclusion of Cl_2, the CRN acquires a genuine conservation law. Hence, we can write $[Cl_2]$ as a linear combination of observable species. Because $[Cl_2]$ is not independent, it follows that we can quantify this hidden variable. This resolution is not obvious when we are agnostic about elemental composition:

$$X_{13} + \blacksquare \underset{4^*}{\rightleftharpoons} X_{15} + X_{18} \tag{67}$$

$$X_{14} + X_{15} \underset{5^*}{\rightleftharpoons} X_{17} + \blacksquare \tag{68}$$

$$X_{14} + X_{13} \underset{6}{\rightleftharpoons} X_{17} + X_{18} \tag{69}$$

$$X_{18} + X_{13} \underset{7}{\rightleftharpoons} X_{15} + X_{16} \tag{70}$$

$$X_{16} + \blacksquare \underset{8^*}{\rightleftharpoons} 2\ X_{18} \tag{71}$$

In such a case, we can still recover all reactions as candidates, but R_7*, R_8*, and $R_{11}*$ will be incomplete. When we find the dimension of the CRN and the number of species by another technique, we can directly deduce that one species and one conservation law are missing.

However, experimental conditions may lead to data with less dimensions. Figure 8b shows ^{1}H-NMR data for 6 species, with 3 monitored species present at the outset. Two additional species start being produced at $t = 0$, whereas the final product X_{13} is first detected only after $t = 200$ s, coinciding with the exhaustion of species X_{18}. Using SVD, we find $d = 2$ significant dimensions and 4 dominated by noise, with a possible exception in the first few datapoints (Figure 9). We can find four independent positive integer combinations of species that are conserved:

$$L_1^* = [NHS] + [NCS]$$

$$L_2^* = [ArSH] + [HCl] + [NHS]$$

$$L_3^* = [ArSCl] + 2\,[Ar_2S_2] + [ArSH]$$

$$L_4^* = [ArSCl] + [HCl] + [NCS]$$

Compared to our initial analysis, we have one more conservation law than expected. In addition to the expected L_1 to L_3, we obtain L_4^*, which is similar to L_4 except for the $[Cl_2]$ term (Figure 9). This law is emergent and can only hold approximately. The distinction between approximate and strict constraints is important because candidate reactions should be consistent with L_1 to L_3, but not necessarily L_4^*. Indeed, reactions R_7 and R_8 conserve L_4, but not L_4^*.

To leverage such emergent conservation laws, we need to interpret them. One explanation consistent with our previous considerations is that $[Cl_2]$ is approximately constant due to its slow production and rapid consumption. In the full CRN with 7 species, $L_5^* = [Cl_2]$ holds approximately, within experimental error η. Consequently, we have:

$$L_4 = [ArSCl] + [HCl] + [NCS] + L_5^*$$

$$L_4^* = L_4 - L_5^*$$

Thus, at the experimental resolution, L_4^* appears as conserved. Emergent conservation laws demonstrate that care must be taken in infering reactions from observed conserved quantities. At the same time, emergent conservation laws are informative. Here, the emergent conservation law L_4^* allows us to infer the concentration of a hidden species.

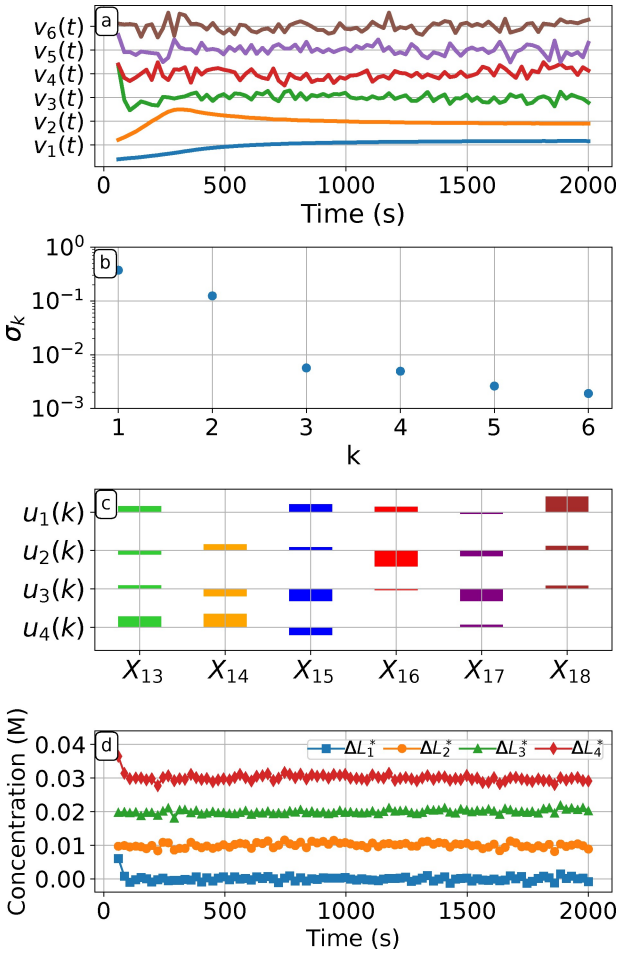


Figure 9. Case Study 4: (a–c) SVD for data shown in Figure 8b.[40] (a) Trajectories $v_i(t)$. (b) Singular values σ_k. (c) Compositions $u_j(k)$. (d) Observed variation for conserved quantities L_1^*, L_2^*, L_3^*, and L_4^*. Variations are depicted with an ordinate offset of 0.01 M, 0.02 M, and 0.03 M for ΔL_2^*, ΔL_3^*, and ΔL_4^*, respectively.

Discussion

The four case studies discussed in this work illustrate that the dimensionality of reaction monitoring data provides structural

information about the underlying CRNs. Data dimensions can be extracted without prior knowledge about the corresponding mechanisms, unraveling clues about both the possible and the impossible structural features of candidate CRNs. One distinguishing feature of this approach is that this analysis can be performed based only on data recorded in a single reaction monitoring experiment, supporting data- and hypothesis-driven mechanistic derivation from the beginning of such a research endeavor.

This is in marked contrast to most both classical and advanced analysis methods of kinetic data, which typically require at least two independent kinetic experiments under perturbed reaction conditions. For instance, the determination of reaction orders requires the measurement of reaction rates or kinetic profiles with modified concentrations of the substrates of interest. Such approaches include the classical method of initial rates and more modern approaches such as reaction progress kinetic analysis (RPKA)[41,42] and variable time normalization analysis (VTNA),[43–45] the latter of which enables visual data analysis.[46]

Other important methods for CRN elucidation rely on using pairs of analogous substrates in varying ratios such as the investigation of nonlinear effects (NLEs)[47–49] and the delayed addition of an alternative substrate under continuous reaction monitoring, delayed reactant labeling.[50] Nevertheless, there are a few advanced methods that allow for kinetic analysis for deriving specific features of CRNs based on kinetic data from individual experiments such as assessing the concentration ratios of reaction intermediates to obtain information about which one is formed first and infer the corresponding local CRN structure.[51] Yet, the only approach that is as generally applicable as the study of kinetic data dimension is kinetic modeling, that is fitting a system of ordinary differential equations to reproduce kinetic data, with the key difference that kinetic modeling relies on a proposed kinetic model. The more complex the underlying CRN, the more potentially viable kinetic models need to be assessed, making this approach not scalable without other mechanistic information.

The only prerequisite for the analysis of dimensionality of kinetic data to be fruitful is to have access to more than one independently measured concentration-time profile for a given chemical process. This is because this method analyzes the linearly independent contributions present in kinetic profiles. Any kinetic data consisting only of a kinetic profile for a single species, with statistically significant concentration changes over time, will result in a data dimension $d = 1$. In the trivial case that this data only consists of noise, the data dimension would be $d = 0$. As soon as a second independently measured set of concentrations is available, the analysis of data dimensions will provide valuable insight to deconvolute the underlying CRN.

While any of the methods for mechanistic elucidation based on kinetic analysis mentioned above provides information about the underlying CRN, analyzing data dimensionality is not meant to replace any of them but should rather be seen as complement that leverages information that has *hitherto* been neglected. Its general applicability to any kinetic time course data and its agnostic character with respect to prior mechanistic information make it an attractive tool for studying complex reaction mechanisms in organic chemistry.

Conclusions

Chemistry has a few well-established dimensional observables (e.g., isosbestic points) that are well-established for mechanistic inferences. Here, we have discussed extensions thereof for interpreting chemical observations. Such extensions require a physical theory that connects dimensionality to both observables and chemical phenomena, and the experimental context they are observed in. Additionally, experimental resolution needs to be accounted for. By formalizing these aspects, data dimensionalities become observations themselves. For instance, the number of conservation laws, their nature, and how species participate therein reveals abundant mechanistic information.

This information can be used to filter hypotheses and guide interpretations. In the case studies outlined above, we have shown how it can be leveraged in the search for CRNs. In some cases, this allowed us to directly propose an underlying CRN, without fitting. In other cases, it allowed us to characterize nontrivial chemical behavior and formulate hypotheses explaining such behavior, some of which might not have been considered otherwise. Overall, isosbestic points remain an indispensible source of mechanistic information for interpreting chemical experiments. By generalization, we surmise that deeper dimensional considerations, such as the ones leveraged here, will equip chemists to extract powerful new insights from experiments and lead to ever more accurate reaction mechanisms.

Acknowledgements

A. B. acknowledges support from the EU (Marie-Sklodowska-Curie grant 101155395).

Conflict of Interests

There are no conflict of interest to declare.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

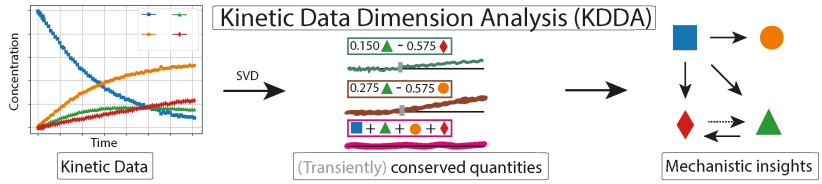
Keywords: chemical reaction networks · chemical kinetics · reaction mechanisms · data dimension · singular value decomposition

[1] R. Franke, D. Selent, A. Börner, *Chem. Rev.* **2012**, *112*, 5675.
[2] P. O. Sturm, A. S. Wexler, *Geosci. Model Dev.* **2022**, *15*, 3417.
[3] M. G. Baltussen, T. J. de Jong, Q. Duez, W. E. Robinson, W. T. Huck, *Nature* **2024**, *631*, 549.
[4] L. Qian, E. Winfree, J. Bruck, *Nature* **2011**, *475*, 368.

[5] S. N. Semenov, A. S. Y. Wong, R. M. van der Made, S. G. J. Postma, J. Groen, H. W. H. van Roekel, T. F. A. de Greef, W. T. S. Huck, *Nat. Chem.* **2015**, *7*, 160.
[6] G. Scheibe, *Angew. Chem.* **1937**, *50*, 212.
[7] M. D. Cohen, E. Fischer, *J. Chem. Soc.* **1962**, pages 3044–3052.
[8] Rigorously, this is based on a few assumptions about co-production.
[9] J. W. Gibbs, *Am. J. Sci.* **1878**, *3*, 441.
[10] J. P. Unsleber, M. Reiher, *Annu. Rev. Phys. Chem.* **2020**, *71*, 121.
[11] A. Nakao, Y. Harabuchi, S. Maeda, K. Tsuda, *J. Chem. Theory Comput.* **2023**, *19*, 713, pMID: 36689311.
[12] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2011**, *7*, 2335, pMID: 26606607.
[13] U. Bentrup, *Chem. Soc. Rev.* **2010**, *39*, 4718.
[14] Y. Shi, P. L. Prieto, T. Zepel, S. Grunert, J. E. Hein, *Acc. Chem. Res.* **2021**, *54*, 546.
[15] E. W. C. Spotte-Smith, S. M. Blau, D. Barter, N. J. Leon, N. T. Hahn, N. S. Redkar, K. R. Zavadil, C. Liao, K. A. Persson, *J. Am. Chem. Soc.* **2023**, *145*, 12181.
[16] X. Xie, E. W. Clark Spotte-Smith, M. Wen, H. D. Patel, S. M. Blau, K. A. Persson, *J. Am. Chem. Soc.* **2021**, *143*, 13245.
[17] A. D. Vogt, E. Di Cera, *Biochemistry* **2012**, *51*, 5894.
[18] Y. Hirono, T. Okada, H. Miyazaki, Y. Hidaka, *Phys. Rev. Res.* **2021**, *3*, 043123.
[19] J. Burés, I. Larrosa, *Nature* **2023**, *613*, 689.
[20] R. Aris, *Arch. Ration. Mech. Anal.* **1965**, *19*, 81.
[21] M. Feinberg, *Foundations of Chemical Reaction Network Theory*, Applied Mathematical Sciences, Springer Nature, Switzerland, 1 edition **2019**.
[22] R. Aris, R. H. S. Mah, *Ind. Eng. Chem. Fundam.* **1963**, *2*, 90.
[23] A. Blokhuis, M. van Kuppeveld, D. van de Weem, R. Pollice, *arXiv* **2023**, https://doi.org/10.48550/arXiv.2306.09553.
[24] Beer, *Ann. Phys.* **1852**, *162*, 78.
[25] If not all species visibly absorb or if some species cannot be distinguished, spectral data may have a lower dimension.
[26] M. Polettini, M. Esposito, *J. Chem. Phys.* **2014**, *141*, 024117.
[27] This law follows from the rank-nullity theorem applied to the underlying reactions (i. e., stoichiometric matrix).
[28] In an operational sense, they'are' conserved, *sensu* consistency with the conservation criterion in Eq. (12).
[29] We can perform this same exercise with data other than kinetic data. For instance, instead of time, we may have a specific reagent added during a titration.
[30] A. G. Campaña, D. A. Leigh, U. Lewandowska, *J. Am. Chem. Soc.* **2013**, *135*, 8639.
[31] Y. Shan, J. Sheng, Q. Zhang, M. C. A. Stuart, D.-H. Qu, B. L. Feringa, *Aggregate* **2024**, *5*, e584.
[32] A. Galanti, J. Santoro, R. Mannancherry, Q. Duez, V. Diez-Cabanes, M. Valášek, J. De Winter, J. Cornil, P. Gerbaux, M. Mayor, P. Samorì, *J. Am. Chem. Soc.* **2019**, *141*, 9273.
[33] G. B. Boursalian, E. R. Nijboer, R. Dorel, L. Pfeifer, O. Markovitch, A. Blokhuis, B. L. Feringa, *J. Am. Chem. Soc.* **2020**, *142*, 16868.
[34] For completeness, the 8 candidates also include the trivial CRN without any reactions for when none of the observable species concentrations change.
[35] K. A. M. Ampt, S. Burling, S. M. A. Donald, S. Douglas, S. B. Duckett, S. A. Macgregor, R. N. Perutz, M. K. Whittlesey, *J. Am. Chem. Soc.* **2006**, *128*, 7452.
[36] That is, there is no other candidate reaction where the candidates are irreversible unimolecular isomerization reactions.
[37] Since we assume photochemical transitions, such a reverse is likely not its microscopic reverse. We use a separate reaction to underscore this point.
[38] T. C. Malig, L. P. E. Yunker, S. Steiner, J. E. Hein, *ACS Catal.* **2020**, *10*, 13236.
[39] M. C. Deem, J. S. Derasp, T. C. Malig, K. Legard, C. P. Berlinguette, J. E. Hein, *Nat. Commun.* **2022**, *13*, 2869.
[40] A. García-Domínguez, N. M. West, R. T. Hembre, G. C. Lloyd-Jones, *ACS Catal.* **2023**, *13*, 9487.
[41] D. G. Blackmond, *Angew. Chem. Int. Ed.* **2005**, *44*, 4302.
[42] D. G. Blackmond, *J. Am. Chem. Soc.* **2015**, *137*, 10852.
[43] J. Burés, *Angew. Chem. Int. Ed.* **2016**, *55*, 2028.
[44] J. Burés, *Angew. Chem. Int. Ed.* **2016**, *128*, 16084.
[45] A. Martínez-Carrión, M. G. Howlett, C. Alamillo-Ferrer, A. D. Clayton, R. A. Bourne, A. Codina, A. Vidal-Ferran, R. W. Adams, J. Burés, *Angew. Chem. Int. Ed.* **2019**, *58*, 10189.
[46] C. D.-T. Nielsen, J. Burés, *Chem. Sci.* **2019**, *10*, 348.
[47] C. Girard, H. B. Kagan, *Angew. Chem. Int. Ed.* **1998**, *37*, 2922.
[48] D. G. Blackmond, *Acc. Chem. Res.* **2000**, *33*, 402.
[49] R. Pollice, M. Schnürch, *Chem. Eur. J.* **2016**, *22*, 5637.
[50] L. Jasikova, M. Anania, S. Hybelbauerova, J. Roithova, *J. Am. Chem. Soc.* **2015**, *137*, 13647.
[51] R. Pollice, N. Dastbaravardeh, N. Marquise, M. D. Mihovilovic, M. Schnürch, *ACS Catal.* **2015**, *5*, 587.

Manuscript received: August 16, 2024
Revised manuscript received: November 27, 2024
Accepted manuscript online: December 24, 2024
Version of record online: ■■, ■■

RESEARCH ARTICLE



Dr. A. Blokhuis, Dr. R. Pollice**

1 – 13

Case Studies of Dimensionality in Chemical Data

In this work, we use kinetic data dimension analysis, which is grounded in chemical reaction network theory, to analyze reaction monitoring data and extract the number of linearly independent reactions. We demonstrate the power of this approach for mechanism elucidation on 4 case studies with experimental data from the literature by deriving new insights.