Catalysis Science & Technology

Unveiling the relation between multiple chemical products and process conditions for Trichloroethylene and Perchloroethylene Production via Catalysis Network Analysis

Journal:	*Catalysis Science & Technology*
Manuscript ID	CY-ART-05-2024-000573.R2
Article Type:	Paper
Date Submitted by the Author:	12-Jul-2024
Complete List of Authors:	Takahashi, Lauren; Hokkaido University, Yamada, Taku; AGC Inc Okamoto, Hidekazu; AGC Inc Takahashi, Keisuke; Hokkaido University, Department of Chemistry

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Unveiling the relation between multiple chemical products and process conditions for Trichloroethylene and Perchloroethylene Production via Catalysis Network Analysis

Lauren Takahashi[a*], Taku Yamada[b], Hidekazu Okamoto[b], Keisuke Takahashi[a*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Trichloroethylene (TRI) and perchloroethylene (PER) are widely-produced in the chemical industry and used as solvents, varnishes, degreasers, and dry cleaning chemicals that involve complex process conditions. Data science and network analysis are used in order to unveil relationships between reactants, process conditions, and selectivities of select products with the aim to improve production efficiency. Data visualization and machine learning reveal the sets of conditions that have positive and inverse relations with TRI and PER selectivities, while transforming the data into networks reveals which sets of experimental conditions correlate with desired outcomes. Thus, it becomes possible to tailor experimental conditions in order to increase desired selectivities while avoiding production of undesirable selectivities.

1 Introduction

Oxyhydrochlorination of ethylene (C_2H_4) and 1,2-dichloroethane (1,2-DCE) is an industrially important process used to produce trichloroethylene and perchloroethylene[1,2]. In particular, trichloroethylene (C_2HCl_3) and perchloroethylene ($Cl_2C{=}CCl_2$) are chemicals that are in high demand within the chemical industry where trichloroethylene is widely used as a degreasing solvent in major industrial areas and commonly used for applications in printing ink, varnish, adhesives, and paint stripper while perchloroethylene is commonly used for dry cleaning and as a brake cleaner.[3–9]

Trichloroethylene (TRI) is produced from the following equation:

$$C_2H_4 + 3\ HCl + 3/2\ O_2 \rightarrow C_2HCl_3 \text{ and } 4H_2O$$

Perchloroethylene (PER) is produced from the following equation:

$$C_2H_4 + 4HCl + 2O_2 \rightarrow C_2Cl_4 + 4H_2O$$

However, despite how simple the equations appear, these reactions involve complex process conditions. Furthermore, the reactions result in numerous chemical byproducts that are undesireable. In particular, multiple reactions occur simultaneously during chlorination and dehydrochlorination reactions as well as during combustion reactions. Many intermediates such as chloroalkanes and chloroalkenes are also produced during these reactions, which impact the production of target products. Previous reports have investigated this through thermodynamic equilibrium theory, but it still remains quite difficult to propose specific process conditions[10,11]. Despite the extensive implementation of various characterization techniques, the complexities of oxyhydrochlorination of ethylene and 1,2-dichloroethane remains a mystery.

Developments in green chemistry have garnered much attention in recent years, particularly due to rising concerns regarding the negative impact the chemical industry has had on the environment. There is a growing need to make chemical production more efficient, to reduce toxins that are produced as reaction byproducts, and to improve the atom economy of synthesis methods, for example.[12–15] However, it is very difficult to do so without understanding how factors such as experimental conditions and the impact of intermediate reactions have on the production of target products. The number of factors that are involved in these reactions vary greatly, and likely involve a level of multidimensionality that is beyond what individual researchers are capable of. Fortunately, recent developments have made it possible for informatics to be considered a viable approach towards

[a] *Department of Chemistry, Hokkaido University, North 10, West 8, Sapporo 060-0810, Japan*
[b] *AGC Inc. Chemicals Company Research & Development General Division, 1-1 Suehiro-cho Tsurumi-ku, Yokohama,Kanagawa 230-0045, Japan*
* *email:keisuke.takahashi@sci.hokudai.ac.jp*
* *email:lauren.takahashi@sci.hokudai.ac.jp*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

trichloroethylene	CCl_4	C_2HCl_5
C_2Cl_6	CO	CO_2
perchloroethylene	C_2H_3Cl	$C_2H_2Cl_2$
1,1,2-$C_2H_3Cl_3$	1,2-$C_2H_4Cl_2$	$C_2H_2Cl_4$

Table 1 Extracted data pertaining to the selectivity of various products. Data extracted from literature[1,2].

Reactant	Cross-section area	Catalyst amount
Volume	Length	HCl/reactant Ratio
N_2/reactant ratio	O_2/reactant ratio	reaction temperature
pressure	pressure-reactant	pressure-HCl
pressure-N_2	pressure-O_2	total gas flow
linear speed	residence time	GHSV

Table 2 Extracted data pertaining to experimental conditions. Data extracted from literature[1,2].

analyzing the relationships between chemicals, experimental conditions, and their resulting products and byproducts.

Catalysts informatics is a data driven field of research centered on catalysts where various data science techniques are used to extract knowledge from multidimensional variables[16–19]. Machine learning has gained much attention and is widely used to process such complex data; however, due to machine learning's black-box nature, it is challenging to understand how such multidimensional data is processed by the machine[20,21]. In such circumstances, catalysis network methods are developed where the relations in the data is visualized as a network[22]. Network analysis is used in order to uncover any relations between experimental conditions and products in order to help improve process conditions, reduce undesireable byproducts, and increase production of target products. Here, machine learning analysis and network analysis are independently performed to unveil the relations between chemical plant and experimental conditions and resulting selectivities of select products of trichloroethylene and perchloroethylene during oxyhydrochlorination. In particular, oxyhydrochlorination of ethylene and 1,2-dichloroethane is chosen as the prototype reaction and data is collected from literature sources where catalysts informatics is performed in order to unveil the complexities of relations between process conditions and each chemical product.

2 Data Collection and Methods

Data used for data analysis is manually extracted from literature[1,2]. In particular, 51 types of data relating to single compositions C_2H_4 or 1,2-DCE are extracted from the literature data and consists of 32 data points pertaining to reactants, catalysts, gas/reactant ratios, experimental conditions, and selectivities of various products of the reaction. Extracted selectivities are listed in Table 1. The extracted experimental conditions are also listed in Table 2.

Reactants consist of ethylene (C_2H_4) or 1,2-dichloroethane (1,2-DCE). Cross-section area refers to the area of the reactor. Amount refers to the weight of catalyst found in the reactor while Volume refers to the volume of catalysts in the reactor. Length is height of catalysts in the reactor. HCl/reactant, N_2/reactant, and O_2/reactant are the ratios of each gas against total reactant. Reaction temperature is the temperature taken in the reaction. Pressure is total pressure in the reactor while pressure-reactant, pressure-HCl, pressure-N_2, and pressure-O_2 are the pressure of the reactant and each gas. Total flow is the total gas flow in the reactor while linear speed is the speed of gas in the reactor. Residence time refers to the time required for gas to pass through the catalyst in the reactor. GHSV represents the space velocity of gases in the reactor.

Supervised machine learning and catalysis network methods are used in this study. For supervised machine learning, random forest regression as implemented in scikit-learn is used for investigating features importance[23]. The number of trees is set to 100. Two separate random forest regressors are made where one sets trichloroethylene selectivity as the target variable and the other sets perchloroethylene selectivity the target variable.

In the case of network analysis, the data is preprocessed and transformed into a network for further analysis using Gephi.[22,24] Initial node placement is randomized upon importing the data into Gephi. To determine the network structure, the ForceAtlas2 algorithm is used.[25] When using the ForceAtlas 2 algorithm, the following parameters were defined: tolerance (speed) = 0.6, approximation = 1.2, scaling = 60, gravity = 1.0, Prevent Overlap = true, and edge weight influence = 1.

3 Results and Discussion

3.1 Data Analysis

Data science is used in order to analyse the relationships between reactants, process conditions, and selectivities of trichloroethylene, perchloroethylene, and other byproducts. In particular, data visualization is first utilized, followed by machine learning in order to uncover any relations or other information buried within the data. Pearson correlation coefficient is calculated in order to evaluate if there are any proportional and inverse proportional relations in the data. A Pearson correlation coefficient map of the collected data is presented in Figure 1. Figure 1 demonstrates that the selectivity of trichloroethylene has a positive relation with the reaction temperature. In the case of the selectivity of perchloroethylene, it has positive relation with the feed rate of HCl, N_2, and O_2 gases. However, the selectivity of perchloroethylene decreases with the increase of reactant pressure. On the other hand, the selectivity of $C_2H_2Cl_2$ decreases with the increase of HCl, N_2, and O_2 gases, thus suggesting that the selectivity of perchloroethylene and $C_2H_2Cl_2$ act in an opposite manner against gas feeds. As a trend, one can see the selectivity of trichloroethylene, CCl_4, C_2HCl_5, C_2Cl_6, CO, and, CO_2 have a positive relation with the feed rate of HCl, N_2, and O_2 gases while the selectivity of perchloroethylene, C_2H_3Cl, $C_2H_2Cl_2$, 1,1,2-$C_2H_3Cl_3$, $C_2H_2Cl_4$ against the feed rate of HCl, N_2, and O_2 have an inverse relation as seen on the correlation coefficient map where red and blue represent proportional and inverse proportional relation, respectively. Thus, one can understand which experimental conditions and selectivities share correlations with each other and better understand how certain factors may influence the outcome of the

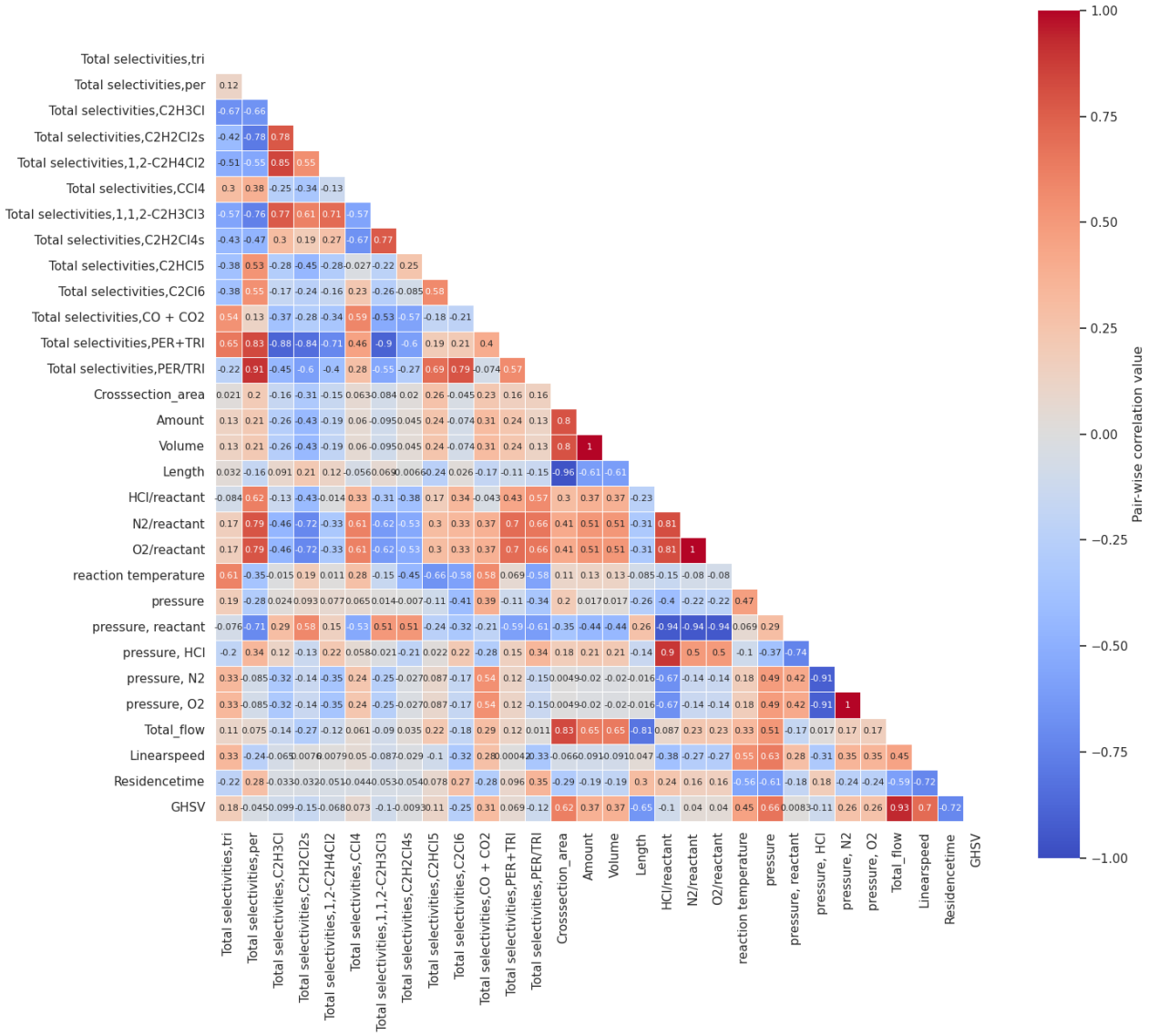

Fig. 1 Pearson correlation coefficient map where red and blue represent proportional and inverse proportional relation, respectively.

reaction.

Machine learning is performed to evaluate the importance of variables presented in the dataset. In particular, random forest regression is used to evaluate the feature importances. Here, two random forest regression models are created where the target variable is set to trichloroethylene and perchloroethylene selectivities, respectively. Calculated features importances for each case is represented in Figure 2. Figure 2 (a) shows that reaction temperature has a major impact on the selectivity of trichloroethylene, thus following the trends seen in the pearson correlation coefficient map shown in Figure 1. Figure 2 (b) shows that pressure of the reactant has a major impact on the selectivity of perchloroethylene, which is also observed in Figure 1. Thus, pearson correlation coefficient and machine learning unveil the key factors responsible for the selectivity. However, the relation involves various products and relation between those products and process conditions are complex matters, therefore, network analysis is implemented.

Reactant	Reactant Tube ID
Reaction Tube Inner Diameter	Catalyst Amount
Catalyst Volume	Catalyst Length
HCl/Reactant	N_2/Reactant
O_2/Reactant	Reaction Temperature
Pressure	Pressure-Reactant
Pressure-HCl	Pressure-N_2
Pressure-O_2	Total Gas Flow
Linear Velocity	Residence Time
GHSV	Selectivity: TRI
Selectivity: PER	Selectivity:C_2H_3Cl
Selectivity: $C_2H_2Cl_2$'s	Selectivity: 1,2-$C_2H_4Cl_2$
Selectivity: CCl_4	Selectivity: 1,1,2-$C_2H_3Cl_3$
Selectivity: $C_2H_2Cl_4$'s	Selectivity: C_2HCl_5
Selectivity: C_2Cl_6	Selectivity: CO+CO_2
Selectivity: PER+TRI	Selectivity: PER/TRI

Table 3 Experimental data categories used for network modeling. Data is collected from previous reports.[1,2]

3.2 Network Analysis

Catalysis network analysis is employed in order to understand the complex relations between process conditions and product selectivities. Networks have been previously used to visualize the relationships between catalyst composition, experimental conditions, and resulting yields, where the resulting information was then used to successfully design high-yield catalysts more efficiently[22]. In this instance, networks are used in order to visualize the relationships between various experimental conditions and the resulting selectivities of a select number of products.

In order to analyse the multidimensional data in further detail, literature experimental data is transformed into a network using Gephi and the ForceAtlas2 algorithm.[24,25] Data preprocessing is carried out to prepare graph edges and nodes for network construction. In order to determine ideal trade-offs for high PER+TRI selectivity and low PER/TRI selectivity and understand the experimental conditions that correlate with them. The categories of experimental data collected from previous studies are listed in Table 3 and were chosen in order to determine ideal trade-offs for high PER+TRI selectivity and low PER/TRI selectivity and understand the experimental conditions that correlate with them. Experimental data points have been rounded for ease of visualization. Nodes represent reactants, catalysts, gas/reactant ratios, experimental conditions, and selectivities of various products. Edges denote that the connected nodes share a relationship within the experimental data. Edge weights are assigned a value of 1. Note that nodes are represented as circles within the network and edges are represented as the lines that connect nodes together. Node color and size are adjusted for visualization purposes. Node colors are defined as the following: light blue (reactant), army green (reaction tube ID, inner diameter of reaction tube), sea green (catalyst amount, catalyst volume, catalyst length), purple (HCl/reactant, N_2/reactant, O_2/reactant), salmon pink (pressure, pressure-reactant, pressure-HCl, pressure-N_2, pressure-O_2), grey (total gas flow rate), moss green (linear velocity), mustard green (residence time), red (GHSV), and yellow (all nodes relating to total selectivities).

Once preprocessed, the data is then transformed into an indirect graph using the Force Atlas 2 algorithm[25]. The constructed network is illustrated in Figure 3 and can also been seen in further detail in Figures 4 through 7. The network consists of 455 nodes and 12,597 edges composing of data ranging from catalyst information, process conditions, and resulting selectivities of various compounds.

From the network visualized in Figure 3, one can see that various experimental conditions and various selectivities have stronger correlations towards reactants C_2H_4 and 1,2-DCE based on their node placement. Additionally, it also becomes possible to visually understand which sets of experimental conditions, catalysts and catalyst-related properties correlate with particular ranges of PER and TRI based on how closely the nodes for experimental conditions, catalysts and catalyst-related properties are found in regards to particular ranges of PER and TRI. In this case, conditions for two outcomes are of interest and optimal scenarios are investigated where PER+TRI selectivity is high and PER/TRI selectivity is low simultaneously. The first investigation involves determining the best conditions for high TRI selectivity and low PER and CO+CO_2 selectivities. The second investigation involves determining the best conditions for high TRI+PER and low PER/TRI. By focusing on these sets of conditions, the effect of experimental conditions upon the selectivities of various products should become easier to understand. Regions showing these conditions are marked in Figure 8 and Figure 9.

From Figure 8, one can see that out of the two reactants, experiments conducted using C_2H_4 tend to result in selectivities that match either one of the two optimal ranges for PER+TRI and PER/TRI selectivities. In particular, there are two regions where both PER+TRI selectivity is high and PER/TRI selectivity is low simultaneously. In Figure 8, for example, closer analysis reveals that experimental conditions such as reaction temperatures 402°C and 433°C or pressure of 0.09 correlate to high PER+TRI and low PER/TRI selectivities. Figure 8 also illustrates that factors

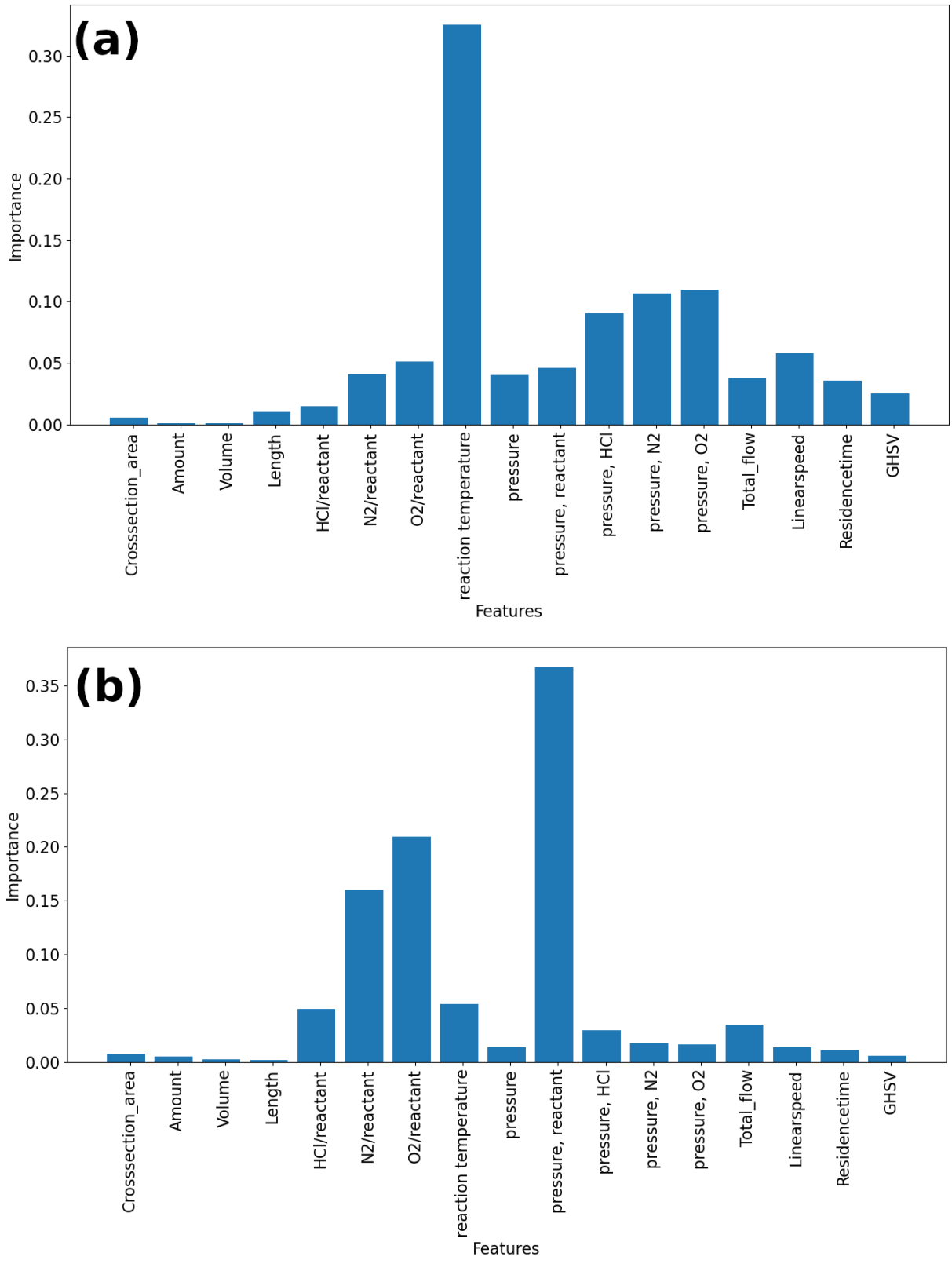

Fig. 2 Features importance for (a)trichloroethylene selectivity and (b)perchloroethylene selectivity by random forest regression. Note that "Crosssection_area" represents cross-section area and "Total_flow" represents total flow.

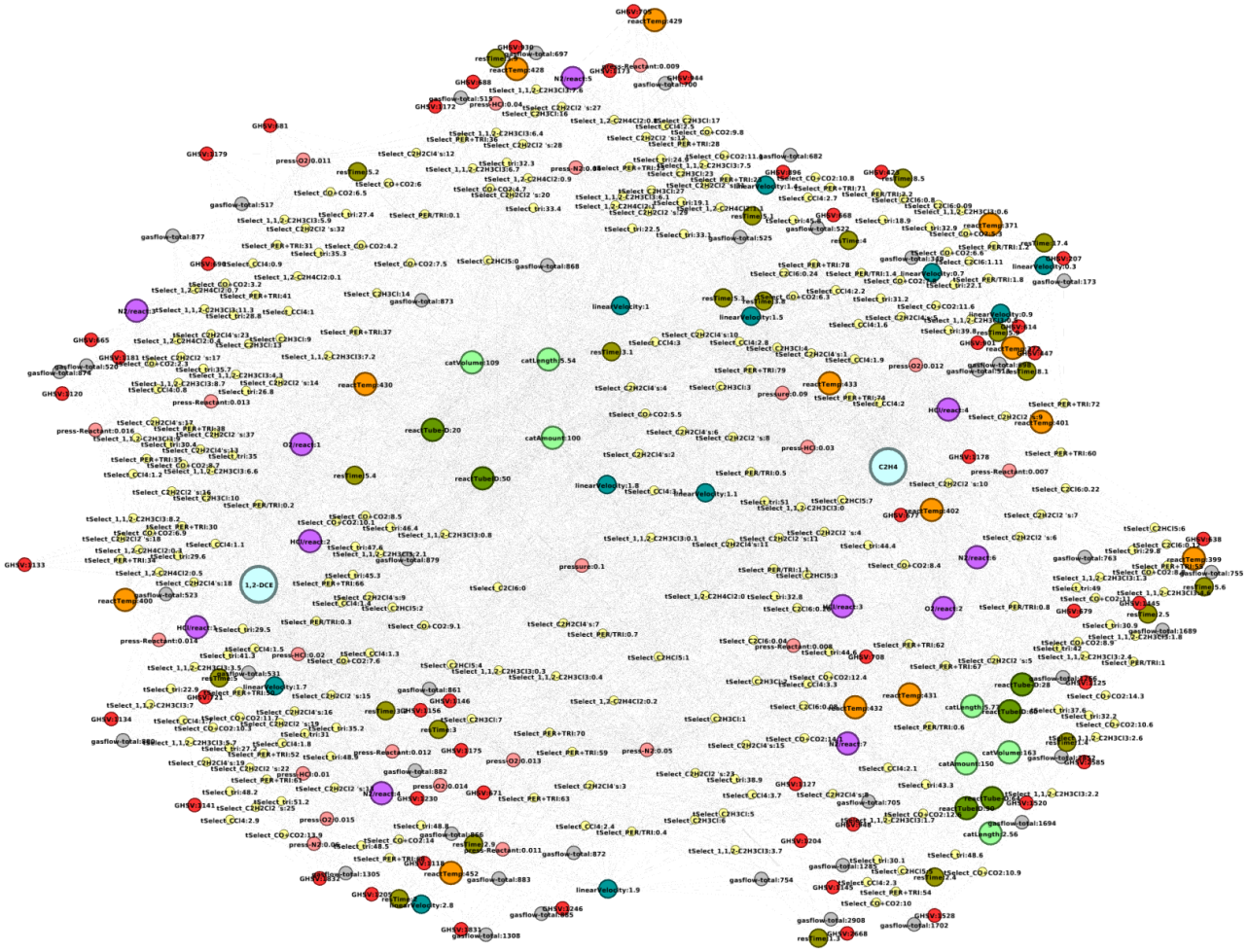

Fig. 3 The full network composed of experimental data as listed in Table 3. A higher resolution of this network can be found in the Supporting Information.

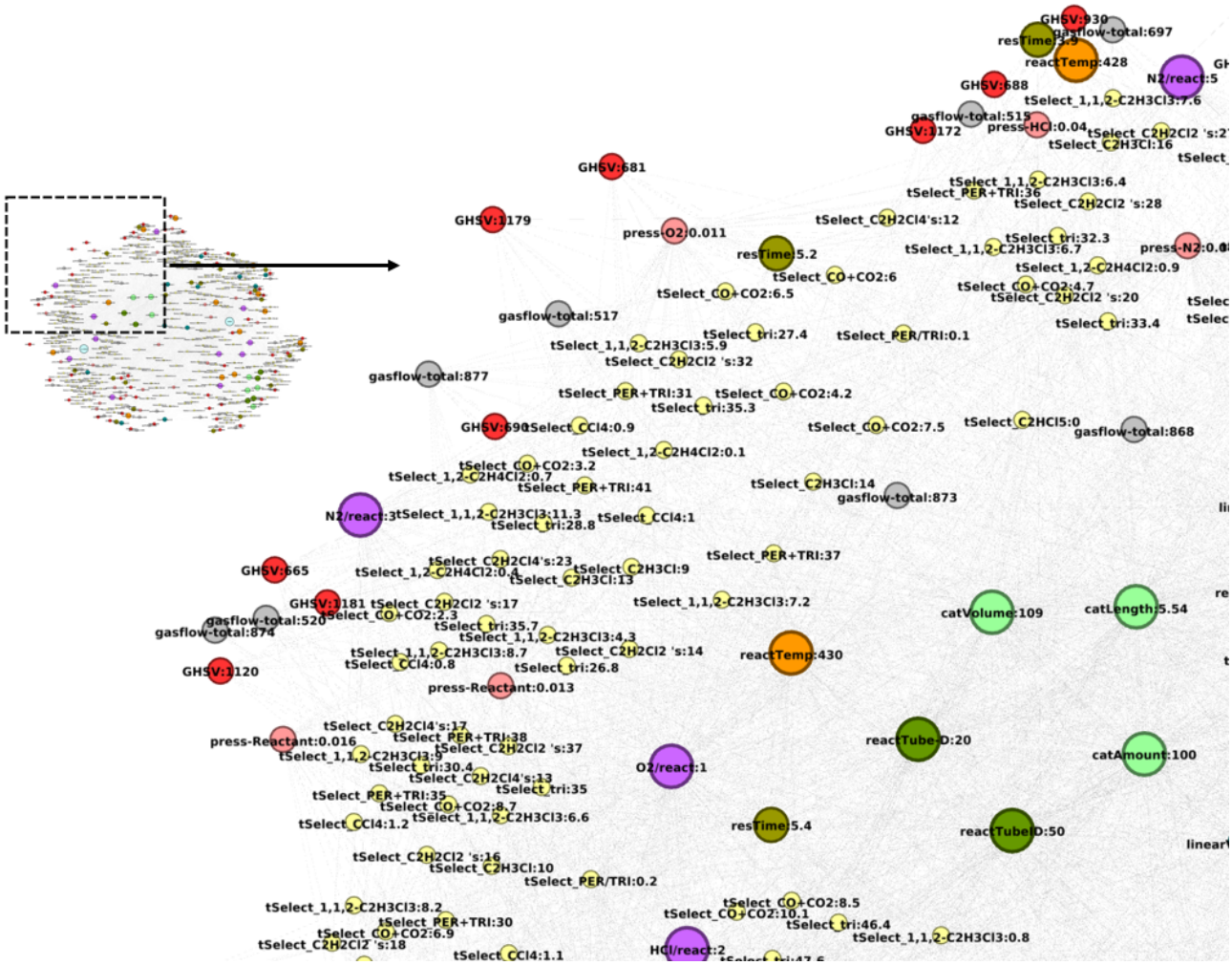

Fig. 4 A more-detailed view of the upper left portion of the network illustrated in Figure 3.

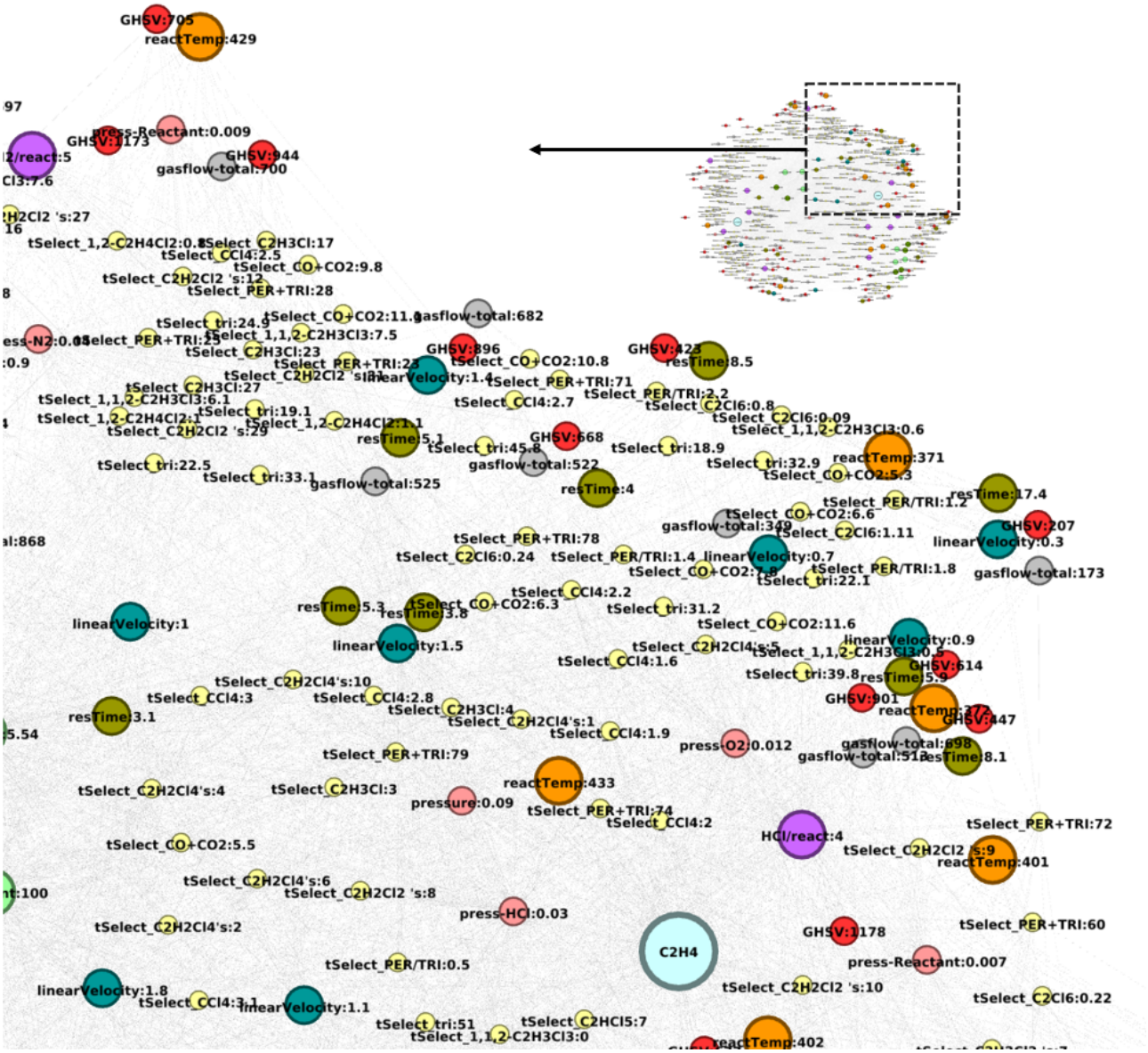

Fig. 5 A more-detailed view of the upper right portion of the network illustrated in Figure 3.

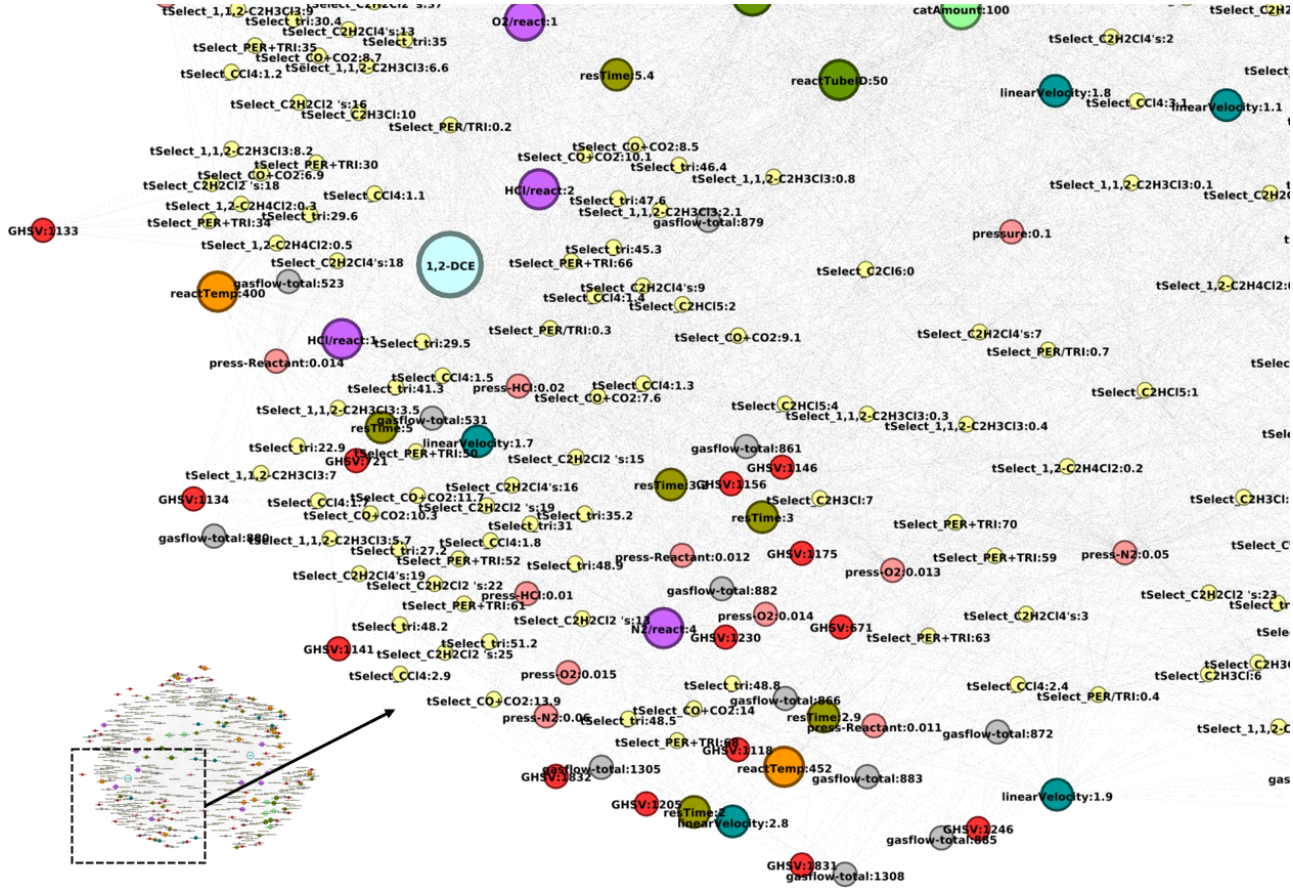

Fig. 6 A more-detailed view of the lower left portion of the network illustrated in Figure 3.

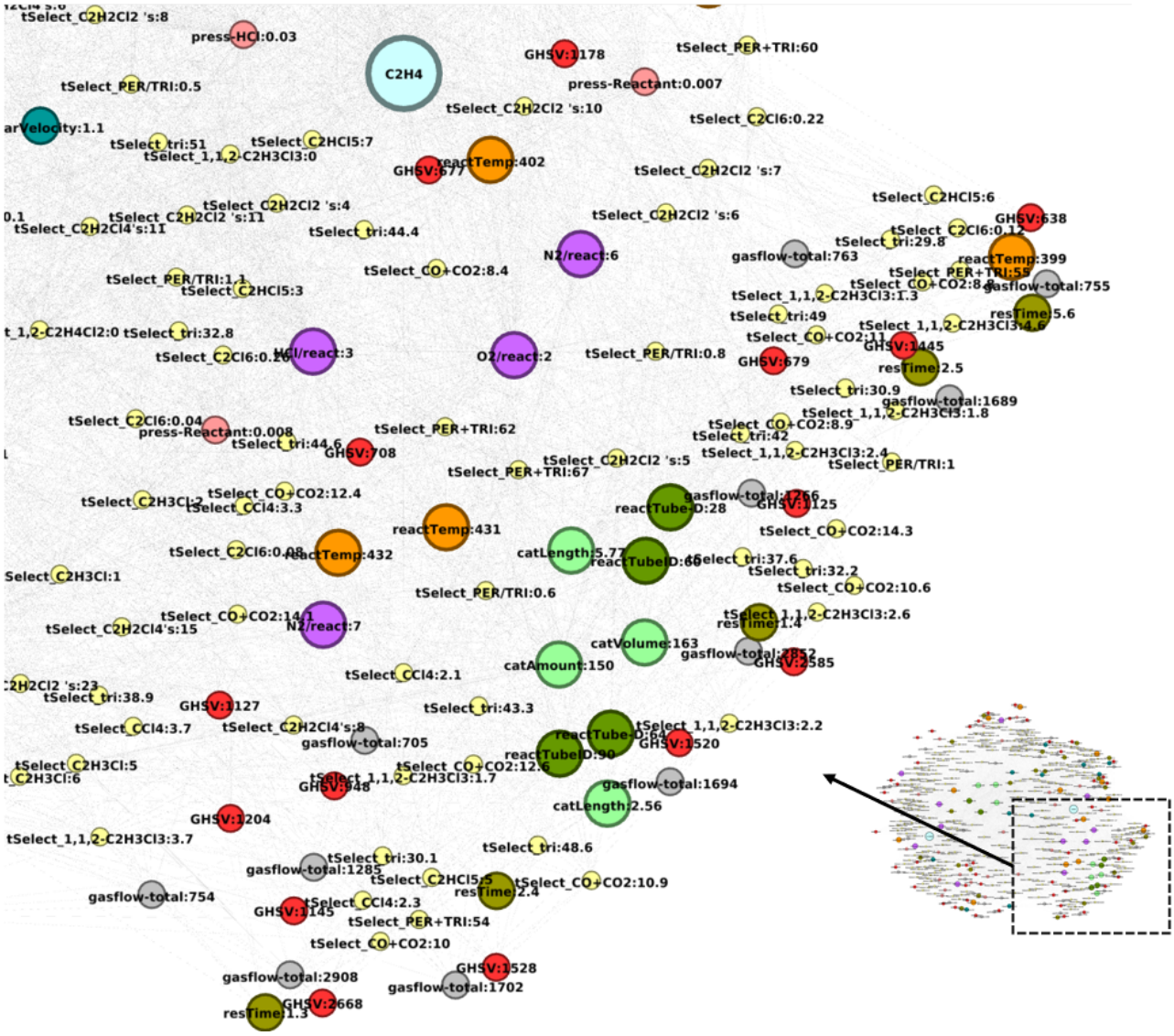

Fig. 7 A more-detailed view of the lower right portion of the network illustrated in Figure 3.

such as catalyst length 5.77 and catalyst amount of 150 also associate with high PER+TRI and low PER/TRI selectivities. Meanwhile, with the case of 1,2-DCE, it becomes clearer that while low PER/TRI selectivities occur, the selectivity of PER+TRI is also low. Thus, it becomes possible to tweak experimental conditions and catalysts in order to more efficiently produce high PER+TRI and low PER/TRI selectivities.

By isolating regions that result in both high PER+TRI selectivity and low PER/TRI selectivity simultaneously, as illustrated in Figure 9, the sets of process conditions that associated with these outcomes become clearer. To start, reaction temperatures that result in these levels of selectivities are found to be 431°C, 432°C, and 433°C; while 402°C also falls in this range, temperatures 401°C and 399°C are found in areas that reflect high PER/TRI ratios, suggesting that the 431-433°C range is ideal. Gas/reactant ratios that associate with high PER+TRI selectivity and low PER/TRI selectivity simultaneously are also made clearer. In particular, HCl/reactant ratios of 3 and 4 are found to correlate with both high PER+TRI selectivity and low PER/TRI selectivity conditions simultaneously. Other gas/reactant ratios such as N_2/reactant ratio of 7 or O_2/reactant ratio of 2 may also lead to simultaneous high PER+TRI selectivity and low PER/TRI selectivity conditions, though may not be as effective given their proximity to areas with higher PER/TRI selectivities. Various pressures can also be optimized; as seen in Figure 8, HCl pressure 0.03, N_2 pressure 0.05, and reactant pressure 0.008 correlate to areas of high PER+TRI selectivity and low PER/TRI selectivities, becoming good candidates for adjustments in plant conditions when attempting to optimize experimental conditions. Other areas that could benefit from adjustments and fine-tuning include catalyst-related conditions. For instance, for cases where C_2H_4 are used as the reactant, researchers may wish to set the catalyst amount to 150 and the length of the catalyst to 5.77 to optimize high PER+TRI selectivity and low PER/TRI selectivity conditions. It should be noted, however, that gas/reactant ratios correspond to the ratio of individual partial pressures of each gas type, making it necessary to consider consistency when using the network to improve experiment conditions. For instance, in cases where the HCl/reactant ratio is 3 and 4, consistency is maintained when HCl pressure is 0.03 and reactant pressure 0.008 due to the HCl/reactant ratio equaling 3.75. Meanwhile, in cases where the N_2/reactant ratio is 7, for example, N_2 pressure of 0.05 and reactant pressure of 0.008 would not be consistent as the N_2/reactant ratio equals 6.25. Instead, the network suggests that a N_2/reactant ratio of 6 may result in high PER+TRI selectivity but also high PER/TRI selectivity. Thus, we can see what factors can be adjusted in order to improve high PER+TRI selectivity and low PER/TRI selectivity conditions.

Meanwhile, from the network it is also possible to understand which conditions may be better to avoid. To start, of the two reactants, 1,2-DCE is more likely to result in undesireable levels of PER+TRI and PER/TRI selectivities. A closer look at the network reveals that reactions involving 1,2-DCE result in low PER/TRI selectivities, a desired condition, yet also results in low PER+TRI selectivity. For instance, when considering gas/reactant ratios that result in high PER+TRI selectivity and low PER/TRI selectivity simultaneously, it may be better to avoid the following ratios: HCl/reactant ratio 1, HCl/reactant ratio of 2, N_2/reactant ratio of 3, N_2/reactant ratio 4, and O_2/reactant ratio 1. Reaction temperatures of 400°C, 430°C, and 452°C are also temperatures that should likely be avoided when attempting to avoid undesireable levels of PER+TRI and PER/TRI selectivities. It should be noted that there may be some cases where temperatures with a single-degree difference may result in very different locations in the network. The measured temperature values are found to correlate with experiments that had different selectivities which are therefore reflected in node placement within the network. In these cases, it is recommended to also consider pairing those temperatures with neighboring experimental condition nodes to help improve the experiment design outcome. Thus, by transforming experimental data into networks, it also becomes possible to better understand particular experiment set-ups or conditions that should be avoided in order to avoid a poor outcome.

4 Conclusion

The relations between reactants, experimental conditions, and selectivities of TRI, PER, and other select products are clarified through the utilization of data science and network analysis. Through applications of data visualization and machine learning, it became clear that certain groups of factors have positive impacts on TRI and PER productions while other factors have an inverse impact on their selectivities. Previous reports have shown that the yield of trichloroethylene can be improved by optimizing certain reaction conditions. In this analysis using the network method, the same trend for the specific reaction conditions reported in the literature were not only obtained, but also succeeded in finding the optimal conditions for other reaction conditions. It was found that not only is PER+TRI selectivity high and PER/TRI selectivity low (i.e., trichloroethylene yield is increased), but also HCl/reactant, O_2/reactant, and reaction temperature are high (i.e., trichloroethylene yield is decreased). In addition, optimal experimental conditions for factors such as N_2/reactant, catalyst length, HCl pressure, N_2 pressure, and reactant pressure are successfully uncovered. By transforming the data into a network, it also becomes possible to see how particular properties associate with various outcomes, making it possible to pinpoint which sets of experimental conditions may be favorable to high PER+TRI selectivity and low PER/TRI selectivity simultaneously while also highlighting sets of conditions that may be best to avoid during experimental set-up. Thus, through the adoption of data science and network analysis, it becomes possible to tailor experimental conditions to influence the selectivities of desireable products in an efficient manner while also reducing the production of undesirable products.

5 Acknowledgment

This work is funded by Japan Science and Technology Agency(JST) CREST Grant Number JPMJCR17P2, ERATO grant number JPMJER1903,JST Mirai Program Grant Number JPMJMI22G4, JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number JP23H01762. The authors wish to thank Hirokazu Takagi, Kazuhiro Suzuki, and Satoshi Kawaguchi from

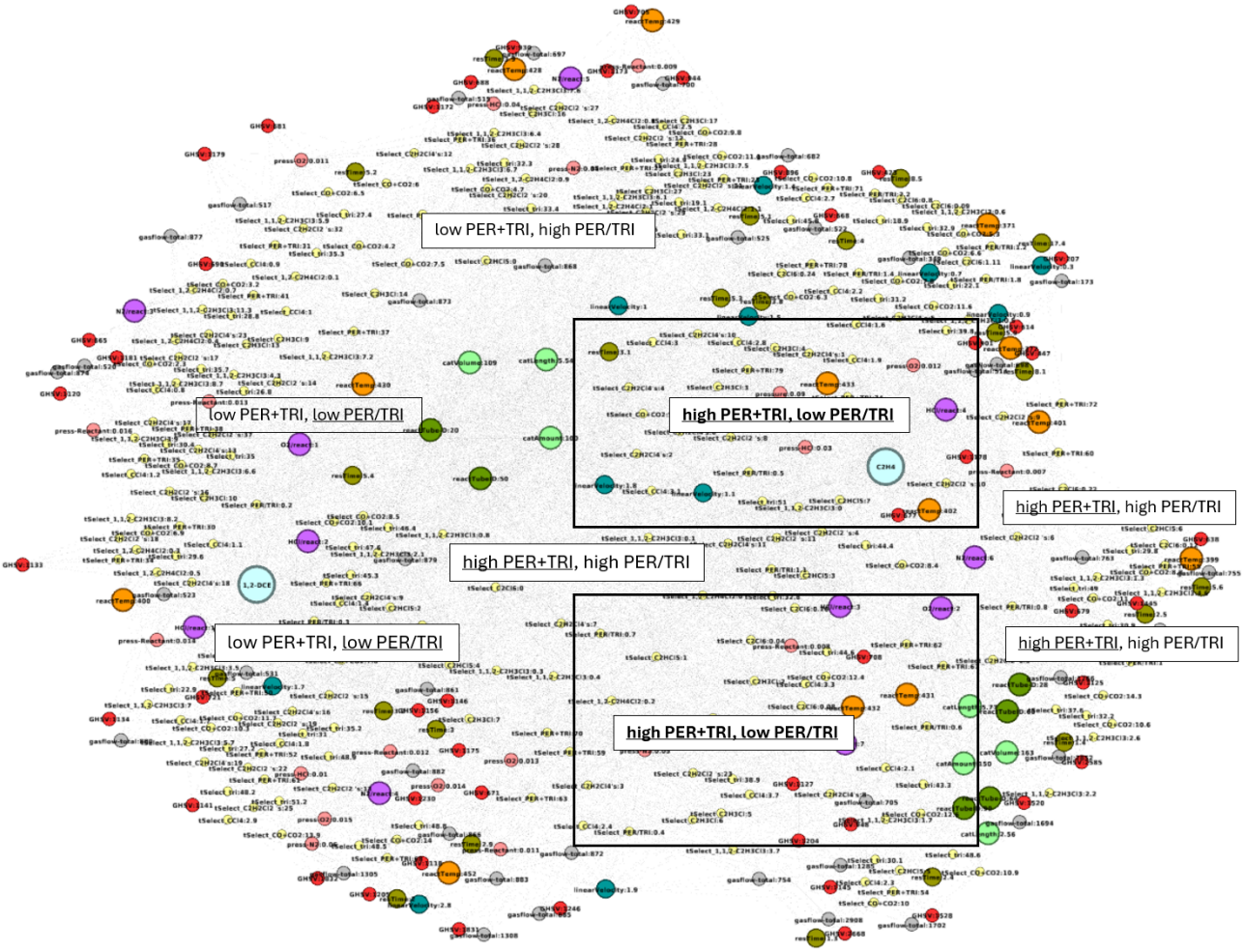

Fig. 8 Select areas of interest of the network illustrated in Figure 3.

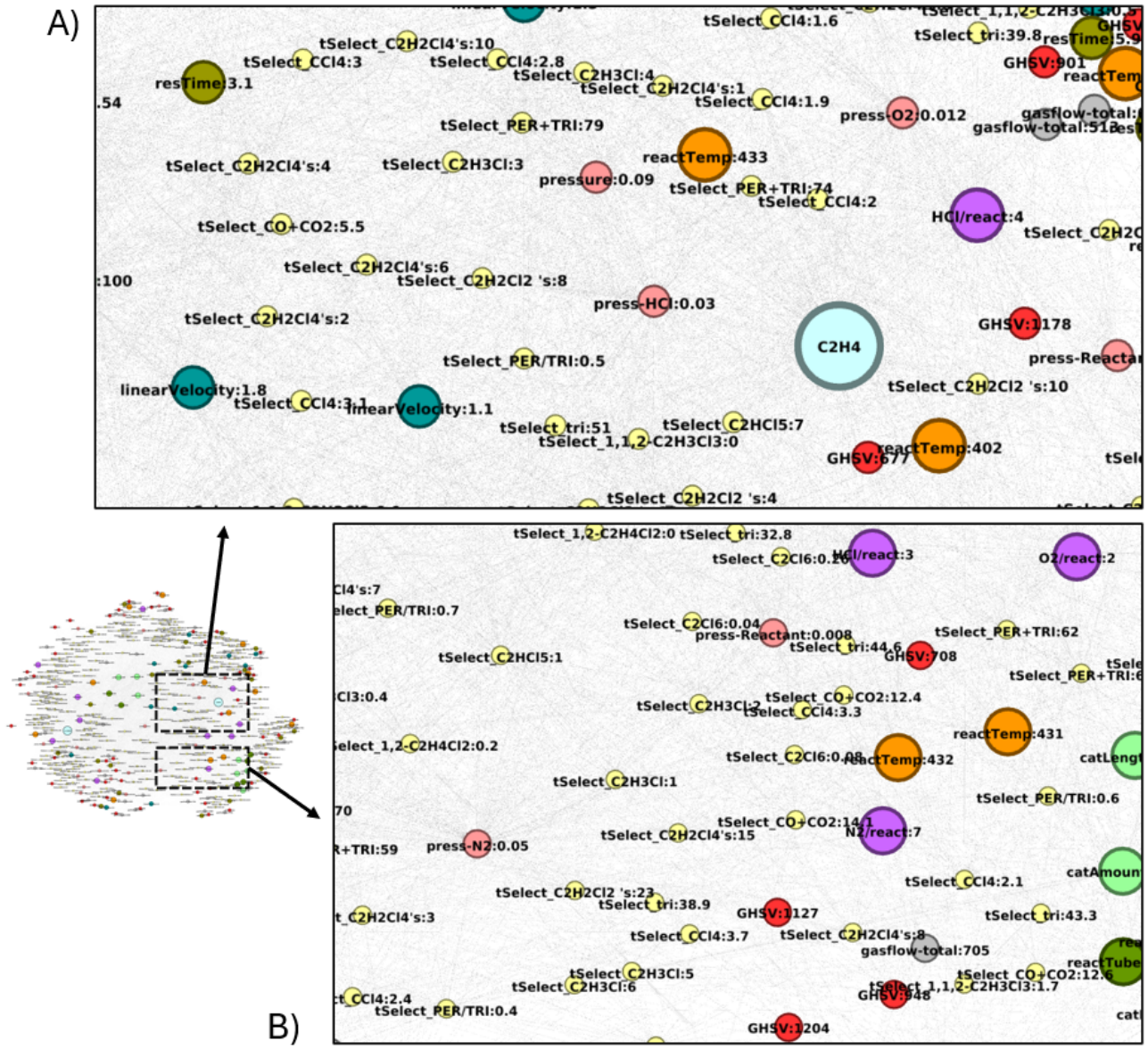

Fig. 9 Areas with high PER+TRI and low PER/TRI, as marked in Figure 8.

AGC Inc. for their support.

Notes and references

1 Arcoya, A.; Cortés, A.; Seoane, X. L. Tri-and Perchloroethylene. 1. Fluid Catalystic Oxyhydrochlorination of Ethylene. *Industrial & Engineering Chemistry Product Research and Development* **1980**, *19*, 77–82.

2 Arcoya, A.; Cortés, A.; Seoane, X. L. Tri-and Perchloroethylene. 2. Fluid Catalytic Oxyhydrochlorination of Dichloroethane. *Industrial & Engineering Chemistry Product Research and Development* **1980**, *19*, 82–86.

3 Smith, G. Trichloroethylene: a review. *British journal of industrial medicine* **1966**, *23*, 249.

4 Doherty, R. E. A history of the production and use of carbon tetrachloride, tetrachloroethylene, trichloroethylene and 1, 1, 1-trichloroethane in the United States: part 1—historical background; carbon tetrachloride and tetrachloroethylene. *Environmental forensics* **2000**, *1*, 69–81.

5 Moore, D. R.; Walker, S.; Ansari, R. *Canadian water quality guidelines for trichloroethylene*; Inland Waters Directorate, 1991.

6 Linak, E.; Leder, A.; Yoshida, Y. C2 chlorinated solvents. *Chemical economics handbook* **1992**, 632–3000.

7 Wolf, K.; Morris, M. Spotting Chemicals: Alternatives to Perchloroethylene and Trichloroethylene in the Textile Cleaning Industry. *Institute for Research and Technical Assistance: Los Angeles, CA, USA* **2007**,

8 Tubich, G.; Davis, I.; Bloomfield, B., et al. Occupational health studies of the shell-molding process. *Arch. Indust. Health* **1960**, *21*, 424–44.

9 Combey, W., et al. Dry Cleaning and Health. *Public Health Inspector* **1965**, *73*, 306–11.

10 Sutherland, I. W.; Hamilton, N. G.; Dudman, C. C.; Jones, P.; Lennon, D.; Winfield, J. M. Chlorination and dehydrochlorination reactions relevant to the manufacture of trichloroethene and tetrachloroethene: Part 1. Reaction pathways. *Applied Catalysis A: General* **2011**, *399*, 1–11.

11 Sutherland, I. W.; Hamilton, N. G.; Dudman, C. C.; Jones, P.; Lennon, D.; Winfield, J. M. Chlorination reactions relevant to the manufacture of trichloroethene and tetrachloroethene; Part 2: Effects of chlorine supply. *Applied Catalysis A: General* **2014**, *471*, 149–156.

12 Clark, J. H. Green chemistry: challenges and opportunities. *Green chemistry* **1999**, *1*, 1–8.

13 Anastas, P. T.; Warner, J. C. *Green chemistry: theory and practice*; Oxford university press, 2000.

14 Anastas, P.; Eghbali, N. Green chemistry: principles and practice. *Chemical Society Reviews* **2010**, *39*, 301–312.

15 Abraham, M. A.; Nguyen, N. "Green engineering: Defining the principles"—resdts from the sandestin conference. *Environmental Progress* **2003**, *22*, 233–236.

16 Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting knowledge from data through catalysis informatics. *Acs Catalysis* **2018**, *8*, 7403–7429.

17 Takahashi, K.; Takahashi, L.; Miyazato, I.; Fujima, J.; Tanaka, Y.; Uno, T.; Satoh, H.; Ohno, K.; Nishida, M.; Hirai, K., et al. The rise of catalyst informatics: towards catalyst genomics. *ChemCatChem* **2019**, *11*, 1146–1152.

18 Takahashi, K.; Ohyama, J.; Nishimura, S.; Fujima, J.; Takahashi, L.; Uno, T.; Taniike, T. Catalysts informatics: paradigm shift towards data-driven catalyst design. *Chemical Communications* **2023**, *59*, 2222–2238.

19 Taniike, T.; Takahashi, K. The value of negative results in data-driven catalysis research. *Nature Catalysis* **2023**, *6*, 108–111.

20 Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

21 Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine learning for catalysis informatics: recent applications and prospects. *Acs Catalysis* **2019**, *10*, 2260–2297.

22 Takahashi, L.; Nguyen, T. N.; Nakanowatari, S.; Fujiwara, A.; Taniike, T.; Takahashi, K. Constructing catalyst knowledge networks from catalyst big data in oxidative coupling of methane for designing catalysts. *Chemical Science* **2021**, *12*, 12546–12555.

23 Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

24 Bastian, M.; Heymann, S.; Jacomy, M. Gephi: an open source software for exploring and manipulating networks. Proceedings of the international AAAI conference on web and social media. 2009; pp 361–362.

25 Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one* **2014**, *9*, e98679.

The data supporting this article have been included as part of the Supplementary Information.