Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Catalysis 223 (2004) 278–289
JOURNAL OF CATALYSIS
www.elsevier.com/locate/jcat

Algebraic system identification for a homogeneous catalyzed reaction: application to the rhodium-catalyzed hydroformylation of alkenes using in situ FTIR spectroscopy

Effendi Widjaja, Chuanzhao Li, and Marc Garland *

Department of Chemical and Environmental Engineering, 4 Engineering Drive 4, National University of Singapore, Singapore 117576, Singapore

Received 28 May 2003; revised 2 September 2003; accepted 4 September 2003

Abstract

An algebraic system identification procedure has been developed and tested on a single semibatch homogeneous rhodium-catalyzed alkene hydroformylation. The goal of this study was to determine the primary algebraic structures which describe the catalytic system in terms of observables. The sole experimental input to the identification is a matrix of in situ spectroscopic data **A** and a matrix of initial moles $\mathbf{N}^0$. Subsequently, we determine (1) the number S of observable species and their characteristic pure-component spectra **a**, (2) the moles of all observable species **N**, (3) the number R of reactions present and their reaction stoichiometries $\boldsymbol{\nu}$, and finally (4) the extent of reactions $\boldsymbol{\xi}$. Meaningful extraction of such algebraic system information (an inverse algebraic problem) is a prerequisite to subsequent detailed kinetic modeling (an inverse kinetics problem). The methodology is successfully applied to this homogeneous transition-metal-catalyzed hydroformylation reaction. The methodological development has clear implications for exploratory studies of new catalytic systems (so-called "gray" reaction systems), in which only in situ spectroscopic reaction data and knowledge of the initial amounts of reagents put into the reactor/system are available.

Keywords: Algebraic system identification; In situ spectroscopy; Rhodium-catalyzed alkene hydroformylation; BTEM; Single semibatch

1. Introduction

Significant progress in the chemical sciences in general, and catalysis in particular, is often highly dependent on extracting meaningful chemical knowledge from large amounts of experimental information. Such experimental measurements can be obtained from a wide range of instruments; however, spectrometers are the true center of most modern chemical research. A systematic and general methodology for the effective analysis and interpretation of *complex-mixture* spectroscopic information to gain physically and chemically meaningful insight into reactive systems is still limited [1]. Although the above system identification concerns are valid for any arbitrary reactive system, the above issues are particularly relevant to the most general and widespread problem in the chemical sciences—namely, system identification in chemical synthesis.

Many liquid-phase chemical syntheses for bulk, fine, and pharmaceutical chemicals are facilitated by the use of transition-metal homogeneous catalysis. Typically, such liquid-phase reactive systems are initiated by introducing various amounts of reactants and organometallics/ligands (catalyst) into a liquid-phase batch reactor. In those cases when in situ observations of the system are performed, one or more on-line spectrometers may be used. The spectral time series obtained during reaction contain information about the time-dependent concentrations of the reactants, intermediates, and products [2].

Given the enormous complexity of the data involved, traditional approaches in chemometrics have often had to rely on some sort of a priori information such as the identities of the major compounds present and their pure-component spectra. Sometimes, a priori information concerning the reaction network also has been incorporated, before multivariate calibration tools can be utilized for evaluating the corresponding concentrations. However, in most real-life research environments, the reaction systems are very complex and their underlying structure essentially unknown (such as

* Corresponding author.
E-mail address: chemvg@nus.edu.sg (M. Garland).

0021-9517/$ – see front matter
doi:10.1016/j.jcat.2003.09.007

many organic and transition-metal homogeneous catalytic syntheses). The experimentalist is therefore faced with a so-called "gray" or even "black" system [3]. A gray system signifies that only spectroscopic data and information concerning the composition and amount of the initial reactants are known; however, further knowledge concerning other species and the reactions present is totally unavailable. In a black system, no knowledge other than the in situ spectroscopic information is available. In other words, the problem of identifying gray or black systems in the chemical sciences belongs to a larger class of mathematical/physical problems typically referred to as "inverse problems" [4].

There are two primary goals that are investigated and developed in this study in order to achieve algebraic system identification. These two goals are:

1. Goal A: Determine the number of observable species s, and the pure-component spectra of all these observable species $\mathbf{a}_{s\times\nu}$ using the spectroscopy measurements alone (*Black System*).
2. Goal B: Given information on the amount of reagents put into the system, determine the time-dependent moles of all observable species $\mathbf{N}_{ke\times s}$, the total number of observable reactions, the reaction stoichiometries $\boldsymbol{\upsilon}_{r\times s}$, and their extent of reactions $\boldsymbol{\xi}_{ke\times r}$ (*Gray System*).

Recently, in this journal, a solution to Goal A called band-target entropy minimization (BTEM) was presented [2]. In the present contribution BTEM is coupled to additional numerical procedures in order to achieve Goal B.

It is worth noting that the present approach is different from recent developments as studied by Furusjo and Danielsson [5], Bijlsma and co-workers [6–8], de Juan et al. [9], and Thurston and Brereton [10]. In these studies, kinetic constraints (a priori knowledge of the reaction orders or reaction models) are imposed at the outset on the multivariate analysis, to determine pure-component spectra, concentration profiles, and reaction rate constants. However, in complex reaction systems such as transition-metal homogeneous catalysis for fine chemicals and pharmaceuticals, such constraints cannot be applied. Almost without exception, homogeneous catalytic reactions are in the gray and black areas; therefore, no information concerning the reaction network is known a priori. In other words, we develop our approach by taking a soft-modeling viewpoint. The final aim is the solution of the algebraic system identification inverse problem as noted in Goal A and Goal B. Such an algebraic system identification is a logical and necessary starting point for later rigorous solution of the differential inverse problem (classic inverse problem in chemical kinetics) in order to determine the kinetics and reaction rate constants [11,12].

A crucial observation is that current soft-modeling techniques fall considerably short of the desired level of sophistication needed to tackle most experimental reactive systems. Therefore, many of the really difficult spectroscopic problems (associated in one way or the other with the algebraic inverse problem) remain unsolved. Difficulties include high spectral overlap and similarity of spectral features, low signal to noise, and—most challenging—changing band shapes and changing band positions (nonstationary signals). Therefore, a new and general approach to overcome these noted difficulties would represent a significant contribution in the field.

In this contribution we present a general methodology which appears to be applicable to a host of reactive systems in the chemical sciences. This methodology is successfully tested on experimental data from one of the most important reactions in transition-metal homogeneous catalysis, the rhodium-catalyzed hydroformylation reaction [13,14]. Numerous detailed in situ spectroscopic studies of this reaction have been reported [15–20]. The present contribution deliberately uses a relatively simple data set to convey the system identification concepts. Much larger systems are currently under investigation in our labs.

2. Computational methods

The rhodium-catalyzed homogeneous hydroformylation of alkenes was run in a semibatch experimental design with in situ FTIR (please refer to Experimental).

Let $\mathbf{A}_{ke\times\nu}$ represent the consolidated spectroscopic data matrix where k denotes the number of spectra taken in one step of a semibatch experiment (each step has an initial condition), e denotes the number of semibatch steps, and ν is the number of data channels associated with the spectroscopic range. $\mathbf{A}_{ke\times\nu}$ can be assumed to result from a linear combination of cell path length $\mathbf{l}_{ke\times ke}$, concentration matrix $\mathbf{C}_{ke\times s}$, and the pure-component spectra matrix $\mathbf{a}_{s\times\nu}$ (where s denotes number of observable species in the chemical mixture), and $\boldsymbol{\varepsilon}_{ke\times\nu}$ is error. The bilinear form in Eq. (1) is assumed to be locally valid. Therefore in larger composition regions, the error term includes experimental and instrumental error as well as *nonlinearities* in the Lamber–Beer–Bouguer model [21].

$$\mathbf{A}_{ke\times\nu} = \mathbf{l}_{ke\times ke}\mathbf{C}_{ke\times s}\mathbf{a}_{s\times\nu} + \boldsymbol{\varepsilon}_{ke\times\nu}. \tag{1}$$

2.1. Pure-component spectra reconstruction and determination number of observable species

The novel algorithm called band target entropy minimization (BTEM) was used to reconstruct each *normalized* observable pure component spectrum $\hat{\mathbf{a}}_{1\times\nu}$ in the data array $\mathbf{A}_{ke\times\nu}$, and repeated and exhaustive use of BTEM determines the number of observable species. The algorithm proceeds using singular value decomposition of $\mathbf{A}_{ke\times\nu}$ followed by appropriate transformation of the basis vectors $\mathbf{V}^{\mathbf{T}}_{z\times\nu}$ ($z \gg s$) using Eq. (2). The seminal concept used is information entropy theory [22]. The BTEM algorithm performs spectral reconstruction one spectrum at a time [2,23–25].

For detailed descriptions of the BTEM algorithm, readers can refer to Refs. [2,23]. BTEM has been successfully used to obtain the pure-component spectra of species at sub-ppm levels, and species whose features contribute less than circa 0.1% of total signal intensity [25]:

$$\hat{\mathbf{a}}_{1\times\nu} = \mathbf{T}_{1\times z}\mathbf{V}^{\mathbf{T}}_{z\times\nu}. \tag{2}$$

2.2. Data renormalization

The aim of numerical renormalization of the data is to eliminate the dependence of the spectroscopic data on volume V and cell path length l variables using the solvent n-hexane as the reference A^{std}_{ω}.

Very briefly, based on the LBBL model, the absorbance data $\mathbf{A}_i$ at the ith wavenumber/channel is a linear combination of optical path length l, molar absorptivity a_{ij}, and molar concentration c_j, and where s = number of species. Clearly, the concentrations c_j are liquid-phase volume dependent:

$$A_i = \sum_{j=1}^{s} a_{ij}c_j l. \tag{3}$$

After numerous substitutions [26], Eq. (4) is obtained, where the values for the absorbance and absorptivity of the internal standard at wavelength ω are used, and the moles n_{std} of the internal standard are known. It is important to note that as indicated by Eq. (4), the dimensionless absorbance A^{Dml}_i is a linear function of the moles of species present in the liquid phase, regardless of spectroscopic cell-thickness changes or liquid-phase volume changes:

$$A^{\text{Dml}}_i = \frac{a^{\text{std}}_{\omega} n_{\text{std}}}{A^{\text{std}}_{\omega}} A_i = \sum_{j=1}^{s} a_{ij} n_j. \tag{4}$$

2.3. Real IR absorptivities and time-dependent moles

Since the reconstructed pure-component spectra $\hat{\mathbf{a}}$ obtained from BTEM (Goal A) are in normalized form, a diagonal weighting matrix $\mathbf{d}_{s\times s}$ is required to scale them to their real magnitudes, i.e., correct absorptivities $\mathbf{a}$,

$$\mathbf{A}^{\text{Dml}}_{ke\times\nu} = \mathbf{N}_{ke\times s}\mathbf{a}_{s\times\nu} = \mathbf{N}_{ke\times s}\mathbf{d}_{s\times s}\hat{\mathbf{a}}_{s\times\nu}. \tag{5}$$

If the dimensionless absorbance matrix $\mathbf{A}^{\text{Dml}}_{ke\times\nu}$ and pure spectral estimates from BTEM have been determined, the time dependence of the moles of all species $\mathbf{N}_{ke\times s}$ can be derived as shown in Eq. (6), where $[\hat{\mathbf{a}}_{s\times\nu}]^{+}$ is the pseudo-inverse matrix of $\hat{\mathbf{a}}_{s\times\nu}$, and $(\mathbf{d}_{s\times s})^{-1}$ is the inverse matrix of $\mathbf{d}_{s\times s}$:

$$\mathbf{N}_{ke\times s} = \mathbf{A}^{\text{Dml}}_{ke\times\nu}[\hat{\mathbf{a}}_{s\times\nu}]^{+}(\mathbf{d}_{s\times s})^{-1}. \tag{6}$$

A material balance criterion can be employed to construct a reaction invariant model to optimize the diagonal matrix $\mathbf{d}_{s\times s}$. The estimated moles of conserved elements (or groups) throughout reaction $\overline{\overline{\mathbf{N}}}_{ke\times E}$ can be calculated as a linear combination of moles of species $\mathbf{N}_{ke\times s}$ and the atomic matrix $\boldsymbol{\upsilon}_{s\times E}$.

$$\overline{\overline{\mathbf{N}}}_{ke\times E} = \mathbf{N}_{ke\times s}\boldsymbol{\upsilon}_{s\times E}. \tag{7}$$

A least-squares approach/sum of square error function minimization between $\overline{\overline{\mathbf{N}}}_{ke\times E}$ and the $\mathbf{N}_{ke\times E}$ (known from the initial reaction conditions) is applied to optimize $\mathbf{d}_{s\times s}$. Accordingly, the real properly scaled IR absorptivities $\mathbf{a}$ and the mole numbers for each species $\mathbf{N}_{ke\times s}$ can be obtained:

$$\text{Min}\,\mathbf{G} = \sum_{ke}\sum_{E}\left(\overline{\overline{\mathbf{N}}}_{ke\times E} - \mathbf{N}_{ke\times E}\right)^2 \quad \text{w.r.t. } \mathbf{d}_{s\times s}. \tag{8}$$

2.4. Determining the number of observable reactions

According to the Jouguet–De Donder equation [27], the change in moles of the observable species, occurring in the time interval $[t_1, t_k]$, can be defined in terms of the extents of reaction $\boldsymbol{\xi}_{ke\times r}$ and the reaction stoichiometries $\boldsymbol{\upsilon}_{r\times s}$. Equation (9) provides the representation of the Jouguet–De Donder equation in the case of the present semibatch experiments,

$$\Delta\mathbf{N}_{ke\times s} = \mathbf{N}_{ke\times s} - \mathbf{N}^{**}_{ke\times s} = \boldsymbol{\xi}_{ke\times r}\boldsymbol{\upsilon}_{r\times s}, \tag{9}$$

where the superscript $**$ explicitly denotes the initial conditions of each eth experiment, i.e., data attained from the $k = 1$ data sets of each e semibatch perturbation.

After imposing the internal standard, the modified data matrix is now free from variable reaction volume and variable path length dependency, and in turn it can be modeled as a linear combination of moles and molar absorptivities only:

$$\mathbf{A}^{\text{Dml}}_{ke\times\nu} = \mathbf{N}_{ke\times s}\mathbf{a}_{s\times\nu}. \tag{10}$$

Therefore,

$$\begin{aligned}\mathbf{A}^{\text{Dml}}_{ke\times\nu} - \left(\mathbf{A}^{\text{Dml}}_{ke\times\nu}\right)^{**} &= \mathbf{N}_{ke\times s}\mathbf{a}_{s\times\nu} - (\mathbf{N}_{ke\times s}\mathbf{a}_{s\times\nu})^{**} \\ &= \Delta\mathbf{N}_{ke\times s}\mathbf{a}_{s\times\nu}.\end{aligned} \tag{11}$$

Substitution gives

$$\mathbf{A}^{\text{Dml}}_{ke\times\nu} - \left(\mathbf{A}^{\text{Dml}}_{ke\times\nu}\right)^{**} = \boldsymbol{\xi}_{ke\times r}\boldsymbol{\upsilon}_{r\times s}\mathbf{a}_{s\times\nu}. \tag{12}$$

Next, a first estimate of the number of observable reactions can be obtained from the left-hand side of Eq. (13). Singular value decomposition will give rise to three primary matrices:

$$\left(\mathbf{A}^{\text{Dml}}_{ke\times\nu} - \mathbf{A}^{\text{Dml}}_{ke\times\nu}\right)^{**} = \mathbf{U}_{ke\times ke}\boldsymbol{\Sigma}_{ke\times\nu}\mathbf{V}^{\mathbf{T}}_{\nu\times\nu}. \tag{13}$$

The singular values in $\boldsymbol{\Sigma}_{ke\times\nu}$ can be subjected to Malinowski's F test to determine the statistically significant number of factors, which are associated with the number of observable reactions. Once the statistically significant values have been attained, the matrix above can be reapproximated by Eq. (14). Because of the nonlinearities present in the spectroscopic data set, this estimate is only approximate and must be checked later using a target-transformation factor analysis approach. If all the observable reactions have been taken into account, i.e., the dimension of r is accurately

determined, then the right-hand side of Eq. (14) closely approximates the left-hand side and a very accurate representation of the Jouguet–De Donder relation for the reactive system has been achieved:

$$\overline{\overline{\left(\mathbf{A}^{\mathrm{Dml}}_{ke\times v} - \mathbf{A}^{\mathrm{Dml}}_{ke\times v}\right)^{**}}} = \mathbf{U}_{ke\times r}\boldsymbol{\Sigma}_{r\times r}\mathbf{V}^{\mathbf{T}}_{r\times s}. \tag{14}$$

2.5. Estimation of reaction stoichiometries

From the estimates of the time-dependent moles (Section 2.3), the mole changes for the observable species $\Delta\mathbf{N}_{ke\times s}$ are calculated and are subjected to SVD:

$$\Delta\mathbf{N}_{ke\times s} = \mathbf{U}_{ke\times ke}\boldsymbol{\Sigma}_{ke\times s}\mathbf{V}^{\mathbf{T}}_{s\times s}. \tag{15}$$

Under an assumption that the spectroscopic measurements of the system are not too nonlinear, the singular values of the $\boldsymbol{\Sigma}_{ke\times s}$ matrix provide a reasonable indication of the number of significant factors in the $\mathbf{U}$ and $\mathbf{V}^{\mathbf{T}}$ matrices. Thus the rank of the reaction system is revealed; in other words, the number of observable reactions r is now known. Some well-known statistical tests such as F test, IE, RPV, or IND can be employed to determine these significant factors [19]. It is also noteworthy that such a system can only be evaluated if the number of observable reactions, r, is smaller than the number of observable species, s, and the number of spectra taken from whole experiment, ke. Based on this number of significant factors, r, $\Delta\overline{\overline{\mathbf{N}}}_{ke\times s}$ is reformulated as follows:

$$\Delta\overline{\overline{\mathbf{N}}}_{ke\times s} = \mathbf{U}_{ke\times r}\boldsymbol{\Sigma}_{r\times r}\mathbf{V}^{\mathbf{T}}_{r\times s}. \tag{16}$$

The matrix product $\mathbf{U}_{ke\times r}\boldsymbol{\Sigma}_{r\times r}$ contains information associated with the observed extent of reaction space, and the matrix $\mathbf{V}^{\mathbf{T}}_{r\times r}$ contains basis vectors for the observed stoichiometric space. Thus, $\boldsymbol{\xi}_{\mathbf{a}ke\times r}$ and $\boldsymbol{\upsilon}_{\mathbf{a}r\times s}$ may be called the abstract extent of reaction matrix and the abstract reaction stoichiometries matrix, respectively,

$$\boldsymbol{\xi}_{\mathbf{a}ke\times r} = \mathbf{U}_{ke\times r}\boldsymbol{\Sigma}_{r\times r}, \tag{17}$$

$$\boldsymbol{\upsilon}_{\mathbf{a}r\times s} = \mathbf{V}^{\mathbf{T}}_{r\times s}. \tag{18}$$

The physically meaningful reaction stoichiometries and extent of reactions lie in the spaces spanned by the rows of $\boldsymbol{\upsilon}_{\mathbf{a}r\times s}$ and the columns of $\boldsymbol{\xi}_{\mathbf{a}ke\times r}$, respectively, and therefore a transformation matrix $\mathbf{T}_{r\times r}$ is required to rotate the abstract factors into the real factors:

$$\begin{aligned}\Delta\overline{\overline{\mathbf{N}}}_{ke\times s} &= \boldsymbol{\xi}_{\mathbf{a}ke\times r}\boldsymbol{\upsilon}_{\mathbf{a}r\times s} = \boldsymbol{\xi}_{\mathbf{a}ke\times r}\mathbf{T}^{-1}_{r\times r}\mathbf{T}_{r\times r}\boldsymbol{\upsilon}_{\mathbf{a}r\times s}\\ &= \boldsymbol{\xi}_{ke\times r}\boldsymbol{\upsilon}_{r\times s}.\end{aligned} \tag{19}$$

This rotation is perhaps best performed one vector at a time using target factor analysis (TFA). In this context, TFA attempts to find a match between the proposed or target stoichiometry, $(\boldsymbol{\upsilon}_{1\times s})_{\mathrm{tar}}$, and the abstract stoichiometries. The projected stoichiometry vector, $(\boldsymbol{\upsilon}_{1\times s})_{\mathrm{proj}}$, and $\mathbf{T}_{1\times r}$ can be generated from a least-squares approach [1]:

$$\mathbf{T}_{1\times r} = (\boldsymbol{\upsilon}_{1\times s})_{\mathrm{tar}}\mathbf{V}^{\mathbf{T}}_{s\times r}(\mathbf{V}^{\mathbf{T}}_{r\times s}\mathbf{V}^{\mathbf{T}}_{s\times r})^{-1}, \tag{20}$$

$$(\boldsymbol{\upsilon}_{1\times s})_{\mathrm{proj}} = \mathbf{T}_{1\times r}\mathbf{V}^{\mathbf{T}}_{r\times s}. \tag{21}$$

According to Bonvin and Rippin, if target stoichiometries are consistent with the observations, they lie in the observed stoichiometric space, and hence it can be seen that the target and projected stoichiometries are identical [28]. On the other hand, if the target is not consistent with the observations, its projection will differ from the target, and the target must be rejected. The superset of possible target stoichiometries can be obtained/deduced from chemical reasoning or from an approach proposed by Yin [29].

Briefly, we note that r degrees of freedom, as obtained from statistical tests only, are usually not sufficient. Due to the nonlinearities noted repeatedly before, we will use more than r right singular vectors for determining the reactions stoichiometries $\boldsymbol{\upsilon}_{r\times s}$ using TFA [similar to Eqs. (20) and (21)]. Details are reported under Results.

2.6. Estimation extent of reactions

Once a full set of stoichiometric vectors $\boldsymbol{\upsilon}_{r\times s}$ has been identified by TFA, the last remaining quantity in Goal B, namely the extent of reaction $\boldsymbol{\xi}_{ke\times r}$, can be obtained. Since the mole changes arise from a linear combination of the extent of reactions and the reaction stoichiometries, the extents of reaction can be calculated using a pseudo-inverse of $\boldsymbol{\upsilon}_{r\times s}$, where the latter is calculated using a Moore–Penrose approach [30]:

$$\boldsymbol{\xi}_{ke\times r} = \Delta\mathbf{N}_{ke\times s}[\boldsymbol{\upsilon}_{r\times s}]^{+}. \tag{22}$$

3. Experimental

3.1. General information

All solution preparations, for the oxygen-moisture-light-sensitive system, were carried out under argon (99.9995%, Soxal, Singapore) using standard Schlenk techniques [31]. The argon was further purified prior to use by passage through deoxy and zeolite columns. All reactions were carried out under carbon monoxide (99.97%, Soxal, Singapore) and hydrogen (99.999%, Soxal, Singapore) after further purification through deoxy and zeolite columns.

The precious metal complex $Rh_4(CO)_{12}$, with a stated purity of 98% min was obtained from Strem Chemicals (Newport, MA) and was used without further purification, although trace quantities of the high nuclearity cluster $Rh_6(CO)_{16}$ are virtually always present. The n-hexane solvent (stated purity > 99.6%, Fluka AG) was refluxed over sodium potassium alloy under argon. 3,3-Dimethylbut-1-ene(99%, Fluka AG, Switzerland) was used as obtained.

Details of the high-pressure experimental equipment used can be found in Refs. [2,24,25]

3.2. In situ spectroscopic studies

A single semibatch experiment was carried out in the following manner. First 150 ml n-hexane was transferred under

Table 1
Experimental design for a single 11-step semibatch hydroformylation

Step	P_{H_2} (MPa)	P_{CO} (MPa)	$Rh_4(CO)_{12}$ (mg)	33DMB (ml)	Hexane (ml)	T (K)	No. of spectra
1	1.0	1.0	36	3.3	200	293	1–15
2	1.0	2.0	36	3.3	200	293	16–28
3	2.0	2.0	36	3.3	200	293	29–45
4	2.0	2.0	83	3.3	250	293	46–59
5	3.0	2.0	83	3.3	250	293	60–75
6	3.0	3.0	83	3.3	250	293	76–91
7	3.0	3.0	83	6.8	300	293	92–106
8	3.0	3.0	83	6.8	300	298	107–120
9	3.0	3.0	83	6.8	350	298	121–135
10	4.0	3.0	83	6.8	350	298	136–149
11	4.0	4.0	83	6.8	350	298	150–158

argon to the autoclave. The total system pressure was raised to 1.0 MPa CO, and the stirrer and high-pressure membrane pump were started. A solution of 3.3 ml 3,3-dimethylbut-1-ene (33DMB) dissolved in 50 ml *n*-hexane was prepared, transferred to the high-pressure reservoir under argon, pressured with CO, and then added to the autoclave. A solution of 36 mg $Rh_4(CO)_{12}$ dissolved in 50 ml *n*-hexane was prepared, transferred to the high-pressure reservoir under argon, pressured with CO, and then added to the autoclave. A 1.0 MPa hydrogen was then added to initiate the synthesis. Spectra were recorded every 4 min at 0.2 cm^{-1} intervals in the range 1000–2500 cm^{-1}.

After 15 spectra were taken in the first step, the partial pressure of carbon monoxide was increased to 2.0 MPa to initiate the second step and then circa 15 spectra were taken. In the following steps, the partial pressures of carbon monoxide and hydrogen, the loadings of precursor ($Rh_4(CO)_{12}$), reagent (33DMB), solvent (*n*-hexane), and reaction temperature were changed according to the experimental design (see Table 1). A total of 11 steps were performed. In each step, circa 15 spectra were taken and a total of 158 spectra were obtained for the spectral analysis.

4. Results and discussion

4.1. Spectral analysis

Since the signal contributions in the region 1000–1550 cm^{-1} are almost entirely due to background moisture and hexane solvent (and not the species participating in the reaction), and since much of this region has very low transmission, the spectral data in this range will not be included for Goal A. However, the quite accurate analytical band for hexane at 1137 cm^{-1} was used for data renormalization purposes, in order to eliminate any dependency on variable reaction volume and variable cell path length.

The truncated data gave rise to an absorbance matrix with the size: rows = 158 and columns = 4751. SVD was directly performed (no preconditioning) [32] to decompose the raw absorbance data matrix, $\mathbf{A}_{158\times4751}$, yielding the orthonormal

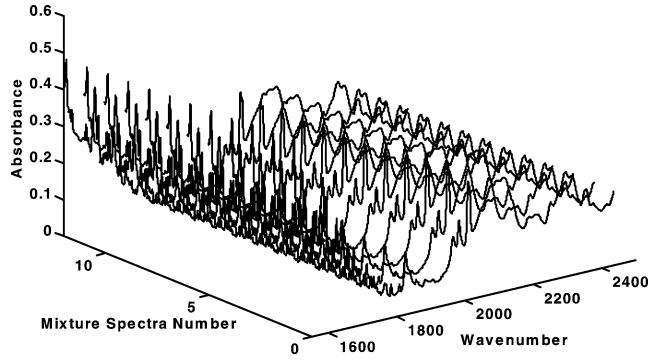

Fig. 1. The first 12 raw experimental reaction spectra from a single 11-step semibatch hydroformylation of 3,3-dimethylbut-1-ene.

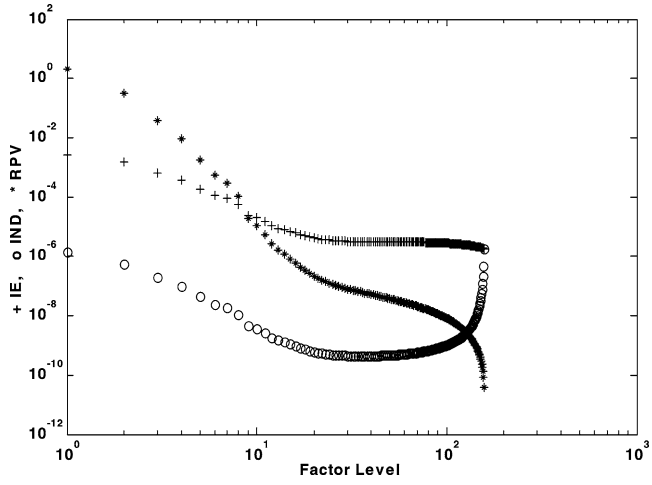

Fig. 2. Attempt at the determination of the number of observable species using IND, IE, RPV statistical tests.

matrices $\mathbf{U}_{158\times158}$ and $\mathbf{V}^{\mathbf{T}}_{4751\times4751}$, and the diagonal singular value matrix $\mathbf{\Sigma}_{158\times4751}$. Since only the first 158 vectors of $\mathbf{V}^{\mathbf{T}}$ matrix are physically meaningful, the absorbance matrix could be approximated using truncated vectors as follows.

$$\mathbf{A}_{158\times4751} = \mathbf{U}_{158\times158}\mathbf{\Sigma}_{158\times158}\mathbf{V}^{\mathbf{T}}_{158\times4751}. \quad (23)$$

The complexity of the raw absorbance spectra without any spectral preprocessing is shown in Fig. 1.

4.2. Determining the number of observable species

A rough first approximation for the number of observable species in a system was obtained using some statistical tests such as IND, IE, RPV, and F test [1]. Statistical tests were performed on the diagonal values of $\mathbf{\Sigma}_{158\times158}$ matrix. However, the results show that IND, IE, and RPV are inconclusive for this real hydroformylation reaction system as shown in Fig. 2. There are subtle inflection points between the eighth and ninth factors, suggesting a demarcation between the primary factors and secondary factors, but at this point it is questionable to conclude that there are only eight observable components in the system.

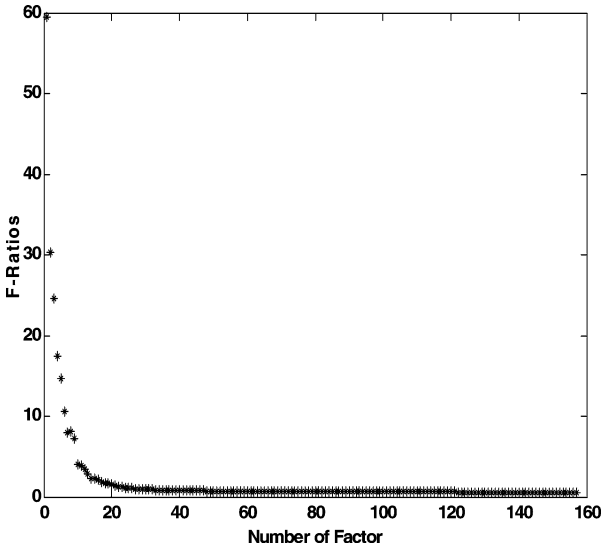

Fig. 3. Attempt at the determination of the number of observable species using Malinowski's *F* test.

A further investigation using the Malinowski *F* test predicted nine observable components. The results are presented in Fig. 3, where the profile shows a clear discontinuity between factors 9 and 10. Thus, at this point in the analysis, at least nine significant factors or nine observable components are assumed present. Furthermore, owing to the known spectral nonlinearities in real systems, the statistical tests should be taken as only a rough guide. As shown below, BTEM will be used to help resolve the issues.

Clearly, the effectiveness of statistical tests for this nonlinear system is questionable. As shown below, Goal A using BTEM was used to resolve the issues.

4.3. Pure-component spectra reconstruction

The right singular vectors of the $\mathbf{V^T}$ matrix were inspected for significant spectral features/extrema to be used as targets in the BTEM algorithm. From the first 14 $\mathbf{V^T}$ vectors, 10 significant extrema labeled 1–10 were selected as shown in Fig. 4. Accordingly these targets were subjected to BTEM for pure-component spectral reconstruction. The selected wavenumber regions, maximum peak absorbance, number of $\mathbf{V^T}$ vectors taken for spectral reconstruction, and the identity of resolved pure-component spectrum are presented in the Table 2.

From Table 2, it can be observed that all reconstructed pure-component spectra are consistent with what we have obtained in previous studies (multiexperiments and three semibatch experiments) [20,23]. Both major components (hexane, dissolved CO, and H_2O) and minor components (44DMP, 33DMB, $Rh_4(\sigma\text{-CO})_9(\mu\text{-CO})_3$, $RCORh(CO)_4$, $Rh_4(\sigma\text{-CO})_{12}$, 2M33DMB, and unknown) are resolved. These pure-component spectral estimates are shown in Fig. 5. The unknown spectrum is probably a deposition on the cell window (organometallics are very sensitive), since a broad spectrum at circa 2330 can be observed (broad features of often due to solids). From this figure, it can be seen that most of the resolved pure spectra are smooth with the exception of the low concentration $Rh_4(\sigma\text{-CO})_{12}$ and 2M33DMB.

Some spectral artifacts, particularly in the last two components noted, are clearly seen in the noisy baseline features. In addition, the estimate of $Rh_4(\sigma\text{-CO})_{12}$ shows some local-

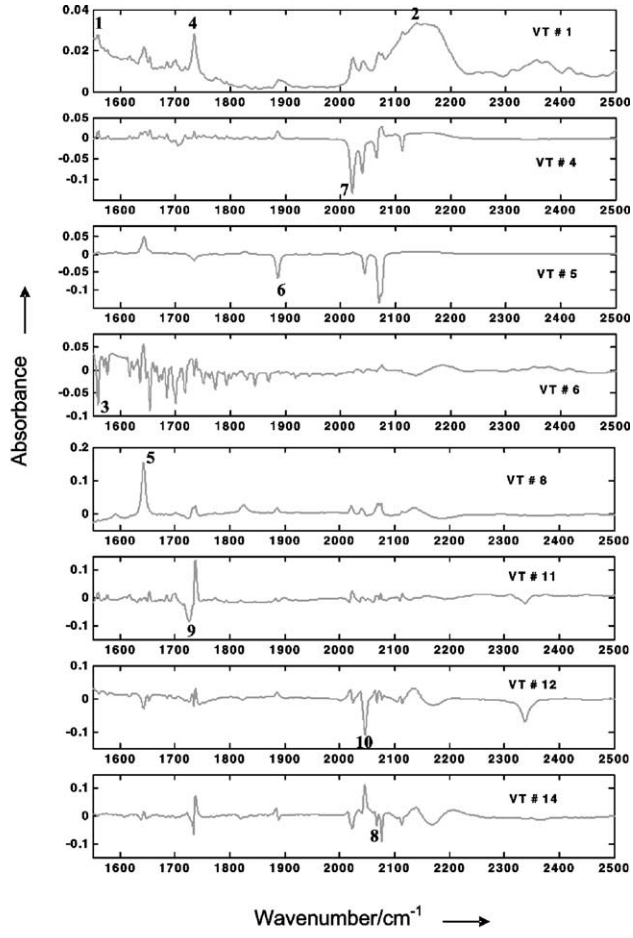

Fig. 4. The first few right singular vectors with 10 marked extrema used to recover the pure-component spectra via BTEM technique.

Table 2
Spectral reconstruction parameters and species identities

No. of species	No. of $\mathbf{V^T}$ vectors, z	Wavenumber region	Maximum absorbance α	Species identification[a]
1	25	1550–1552	1	*n*-Hexane
2	25	2135–2140	1	Dissolved CO
3	25	1558–1560	1	Moisture
4	50	1733–1735	1	44DMP
5	50	1640–1644	1	33DMB
6	50	1885–1887	5	$Rh_4(\sigma\text{-CO})_9(\mu\text{-CO})_3$
7	50	2038–2040	5	$RCORh(CO)_4$
8	50	2067–2069	1	$Rh_4(\sigma\text{-CO})_{12}$
9	50	1720–1726	1	2M33DMB
10	50	2045–2047	1	Unknown

[a] 44DMP, 4,4-dimethylpentanal; 33DMB, 3,3-dimethylbut-1-ene; 2M33DMB, 2-methyl,3,3-dimethylbutanal.

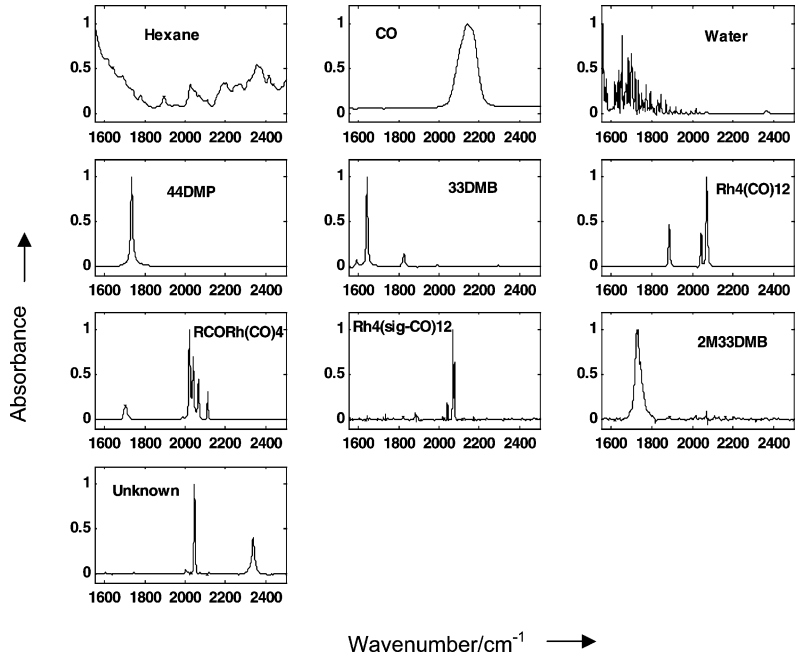

Fig. 5. Ten reconstructed pure-component spectra via BTEM algorithm from the single semibatch hydroformylation reaction system.

ized absorbance at ca. 1642, 1734, 1885, and 2041 cm^{-1}, which are associated with 33DMB, 44DMP, and $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$. The estimate of 2M33DMB also shows localized signal at 2070 cm^{-1}, which corresponds to $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$. Artifacts are frequently observed in spectroscopic studies, particularly when spectral subtractions are performed on nonstationary bands and when spectral reconstruction attempts are made on minor components with very low signal intensity (just above noise level). However, these spectral artifacts often do not seriously prevent identification since the primary bands are usually well resolved. It should be noted that this problem of artifacts is essentially eliminated if more experimental spectra are measured. Indeed, in a previous study with 475 experimental spectra, even trace components representing only 0.15% of the integrated signal were recovered with no noticeable artifacts [25].

From this single 11-step semibatch experiment, reconstruction of pure-component spectra of carbon dioxide (CO_2) and $Rh_6(CO)_{16}$ was also tried. However, the attempts were unsuccessful. To reconstruct CO_2, the target region 2360–2364 cm^{-1} was taken with $z = 50$ vectors and $\alpha = 1$. To reconstruct $Rh_6(CO)_{16}$, the target region 1818–1820 cm^{-1} was taken with $z = 50$ vectors and $\alpha = 5$. The spectral estimates for these two components can be seen in Fig. 6.

These two resolved spectra are clearly distorted, although the two primary bands of $Rh_6(CO)_{16}$ at circa 1819.2 and 2075 cm^{-1} and the one primary band of CO_2 at circa 2360 could be recovered. Many random signals associated with measurement noise are observed. In addition, TFA was performed to confirm the reconstruction quality of CO_2 and $Rh_6(CO)_{16}$. Pure-component spectra of CO_2 and $Rh_6(CO)_{16}$ obtained from real experimental references from this laboratory were used as the target vectors, and were projected onto 50 vectors of $\mathbf{V^T}$ matrix. The sums of square errors between target and projected vectors were 10.60 and 0.140, respectively. The target and projected pure-component spectra are shown in Fig. 7.

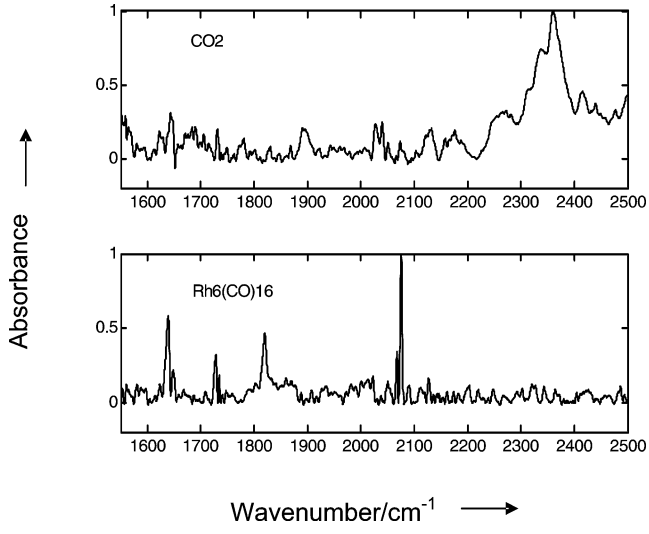

Fig. 6. The pure-component spectral estimates of CO_2 and $Rh_6(CO)_{16}$.

Although the baselines of the projected spectra are very distorted, the expected primary bands could be observed. These results indeed confirm the presence of these two components at very low concentrations. Their signals are certainly imbedded in the noise. The pure-component spectra are hard to resolve. The extremely low concentration of CO_2 was caused by an extra high rate of purging nitrogen before and during the spectroscopic measurements. Furthermore, the exceptionally low signal for $Rh_6(CO)_{16}$ was apparently

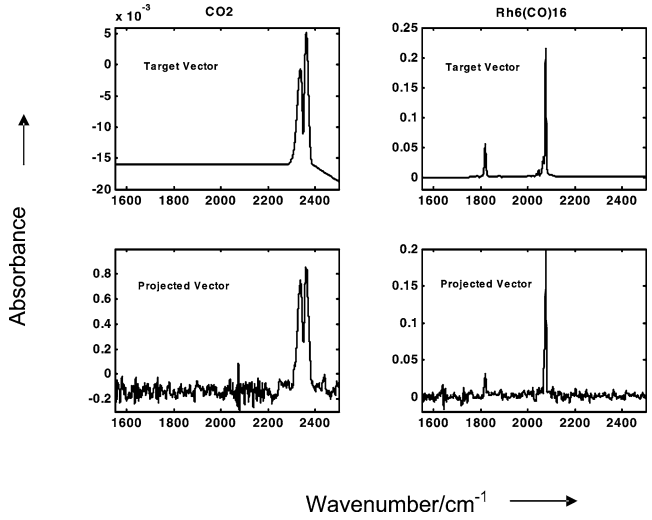

Fig. 7. The target and projected pure-component spectra of CO_2 and $Rh_6(CO)_{16}$.

Table 3
Percentage of integrated absorbance of each component compared to the total original experimental data

Component	Integrated intensity of each component (%)
n-Hexane	65.770
Dissolved CO	28.011
Moisture	1.953
44DMP	2.818
33DMB	1.284
$Rh_4(\sigma\text{-CO})_9(\mu\text{-CO})_3$	0.827
$RCORh(CO)_4$	1.927
$Rh_4(\sigma\text{-CO})_{12}$	0.053
2M33DMB	0.228
Solid dirt	0.317
Total	103.188

due to an exceptionally good experimental procedure (the deoxy columns, made from BASF catalyst and zeolites, had been regenerated the day before) and therefore very low amounts of oxygen/water were present in the reactor system.

The total integrated intensities for the major and minor species are shown in Table 3.

The data above indicate that only circa 8% of the overall spectroscopy signals are due to minor components. Furthermore, $Rh_4(\sigma\text{-CO})_{12}$, 2M33DMB, and unknown contribute to less than 0.6% of the overall signals. This is an independent confirmation that these three components are very minor components in this reaction system. The total intensity ca. 103% is due to the fact that the reconstructed pure-component spectra are mean representations over all observations (bands are usually just a little bit broader). Thus, pure-component spectral estimates have slightly larger integrated intensities than any particular true-component spectrum.

One very important concluding remark is needed for this section. First, there was nothing special about the choice of the 10 extrema identified in the first 14 $\mathbf{V}^{\mathbf{T}}$ vectors, which were used in the BTEM algorithm. As we have shown earlier, BTEM is usually capable of reconstructing a pure-component spectrum, from virtually any choice of targeted band. Thus an exhaustive search of the first 10–15 vectors in $\mathbf{V}^{\mathbf{T}}$, using BTEM on every spectral feature, leads to many redundant reconstructions. Secondly, the maxima present in the 10 reconstructed pure-component spectra are sufficient to account for essentially all the extrema present in the first 10–20 $\mathbf{V}^{\mathbf{T}}$ vectors.

4.4. Data renormalization

As noted in the computational section, before further analysis (Goal B), the raw experimental absorbance data matrix must be renormalized. The aim is to eliminate variable reaction volume and variable path length dependency.

In the present study, the standard A_{ω}^{std} used for each reaction spectrum is obtained from the solvent *n*-hexane's absorbance data at a reference wavenumber $\omega = 1137\ \text{cm}^{-1}$. In addition, the moles of solvent *n*-hexane used n_{std} can be calculated from its initial prepared volume and the known molar absorptivity of *n*-hexane at $\omega = 1137\ \text{cm}^{-1}$.

4.5. Determining the number of observable reactions

SVD was performed on the $[\mathbf{A}_{ke\times\nu}^{\text{Dml}} - (\mathbf{A}_{ke\times\nu}^{\text{Dml}})^{**}]$ matrix, and subsequently the resulting singular values of the $\mathbf{\Sigma}$ matrix were subjected to Malinowski's *F* test. The *F* test result is shown in Fig. 8. The exact number of significant factors is very difficult to infer from this result, since some discontinuities are clearly generated. These can be seen by the gaps between factors 2 and 3, 4 and 5, and 9 and 10. A further confirmation of the number of observable reactions is thus necessary, and it is obtained later via target factor analysis.

4.6. Real IR absorptivities and time-dependent moles

Based on the L^1-normalized pure spectral estimates obtained via BTEM, and knowledge of the amount of initial

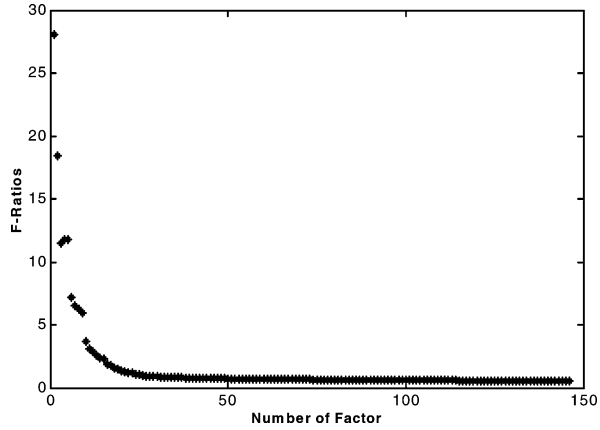

Fig. 8. Attempt at the determination of the number of observable reactions using Malinowski's *F* test.

moles of reagents put into the reactor, a reaction invariant model was set up to calculate the real magnitude of absorptivities according to Eq. (8). In the first attempt, the $Rh_4(\sigma\text{-}CO)_{12}$ and 2M33DMB species were included in the calculation. Because these two components have very low concentrations, the multiple linear regression sensitivity to these two components in the absorbance data is also very low. In the regression and optimization process, the real magnitude absorptivities of these two components were unstable. Hence, in the second attempt, these two components are not included in the regression (their contribution to the mass balances is negligible).

In the present hydroformylation system, the concept of atomic matrix cannot be used alone. Instead, it is modified to include chemical groups. Note that the liquid phase is open to transport of H, C, and O (through H_2 and CO) and these cannot be used. However, fortunately and to a very good approximation, the rhodium (Rh-) is conserved in the liquid phase as is the organic fragment (Org-)$(CH_3)CCHCH_2$. In the solution of Goal B for this hydroformylation, the concept of chemical group is preferred and will be used throughout.

Since there are seven primary observable species in the spectroscopic data—*n*-hexane, dissolved CO, H_2O, 44DMP, 33DMB, $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$, and $RCORh(CO)_4$,—*a* chemical group matrix $\boldsymbol{\upsilon}_{s\times E}$ can be constructed as follows:

$$\boldsymbol{\upsilon}_{s\times E} = \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \end{matrix} \overset{\begin{matrix}\text{Rh-} & \text{Org-}\end{matrix}}{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \\ 4 & 0 \\ 1 & 1 \end{bmatrix}}. \tag{24}$$

Because there are no Rh- nor Org- groups in *n*-hexane, dissolved CO, and H_2O (note also that this last component is not in the liquid phase system), the corresponding matrix elements are zero. Thus, in the present context, it is not possible to calculate the real magnitudes of the absorptivities for these latter three components. However, their inclusion in the $\mathbf{d}_{s\times s}$ optimization is important, since a full spectral range fitting is performed using multiple linear regression to estimate the moles of all species. All primary and significant pure-component spectra are required in the calculation.

The number of decision variables in the diagonal matrix $\mathbf{d}_{s\times s}$ is only 4, and these correspond to 44DMP, 33DMB, $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$, and $RCORh(CO)_4$. The weightings for the first three components $\{d(1,1), d(2,2), \text{and } d(3,3)\}$ (corresponding to *n*-hexane, dissolved CO, and H_2O) were assigned arbitrary nonzero numbers of a proper magnitude to prevent matrix ill-conditioning. The downscaling factor [33] 10^{-2} was also applied to the second column of the $\mathbf{N}_{ke\times E}$ matrix which is associated with the group Org-. The purpose was to adjust the magnitude of both columns since Org- is roughly 100 times larger than Rh-. With this adjustment, it was expected that the numerical sensitivity would be minimized and that good optimal values of $\mathbf{d}_{s\times s}$ could be obtained.

Table 4
The maximum absorptivities for each pure-component spectrum

Species	Max peak (cm^{-1})	Absorbance
44DMP	1734.2	264.7
33DMB	1642.6	37.9
$Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$	2070	15,680
$RCORh(CO)_4$	2021.4	4657.6

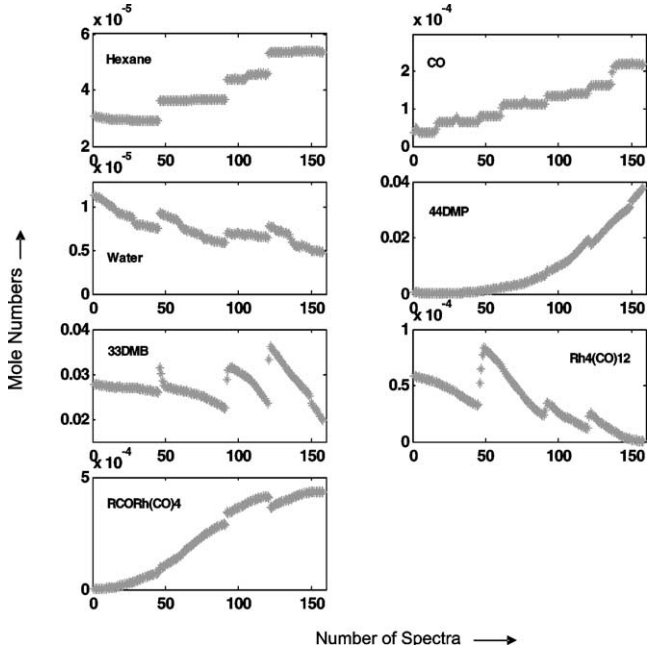

Fig. 9. Mole profiles for hexane, CO, and water (relative) and for the four primary observable components 44DMP, 33DMB, $Rh_4(CO)_{12}$, and $RCORh(CO)_4$ (absolute).

The minimization of the sum of square error function was performed using Corana's SA [34]. Upon attaining the optimum values of $\mathbf{d}_{s\times s}$, the properly scaled absorptivities of these pure-component spectra can be calculated. Furthermore, the moles of these primary species can also be obtained by a pseudo-inverse calculation. The maximum absorbance of the properly scaled pure-component spectra are shown in Table 4, and the moles of all species are presented in Fig. 9. Please note that the moles of *n*-hexane, dissolved CO, and vapor-phase H_2O are in the form of relative concentrations since the real magnitudes of the pure-component spectra of these three components are unknown from this calculation. The mean values of the solute mole fraction concentrations were 0.0049, 0.0136, 1.75E-5, 1.17E-4 for 44DMP, 33DMB, $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$, and $RCORh(CO)_4$, respectively.

4.7. Estimation of reaction stoichiometries

The mole changes $\Delta\mathbf{N}_{ke\times s*}$ for the primary solutes (44DMP, 33DMB, $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$, and $RCORh(CO)_4$)

were calculated using Eq. (25), where s^* denotes these four primary solutes only, and where $\mathbf{N}^{**}_{ke\times s^*}$ denotes the moles of each solute in the first spectrum of each reaction step:

$$\Delta\mathbf{N}_{ke\times s^*} = \mathbf{N}_{ke\times s^*} - \mathbf{N}^{**}_{ke\times s^*}. \tag{25}$$

SVD was performed on $\Delta\mathbf{N}_{ke\times s^*}$, and the diagonal values of the singular matrix $\mathbf{\Sigma}_{ke\times s^*}$ are presented as follows:

$$\operatorname{diag}\mathbf{\Sigma}_{ke\times s^*} = \begin{bmatrix} 0.0341 \\ 0.0149 \\ 0.0003 \\ 0.0000 \end{bmatrix}. \tag{26}$$

It is clearly seen that only the first two factors of the singular matrix are significant, and therefore the number of observable independent reactions is equal to two. This number is consistent with the Gibb's stoichiometric rule proposed by Aris and Mah [35] to predict the number of independent reactions, r. This rule is shown in

$$r \leqslant s - p_1, \tag{27}$$

where s is the number of components and p_1 is the rank of the atomic matrix. But in our case, p_1 is the rank of the chemical group matrix. Thus, for this reduced problem, since $s = 4$ and $p_1 = 2$, r is equal to or less than 2.

Since we know that the system is slightly nonlinear due to the nonstationary absorptivities, a conservative choice of $(r + 1)$ number of significant factors is taken to span the reaction stoichiometries space:

$$\overline{\overline{\Delta\mathbf{N}_{147\times 4}}} = \mathbf{U}_{147\times 3}\mathbf{\Sigma}_{3\times 3}\mathbf{V}^{\mathbf{T}}_{3\times 4}. \tag{28}$$

Target factor analysis is used to check whether the proposed reaction stoichiometries are compatible with the observation space $\mathbf{V}^{\mathbf{T}}_{3\times 4}$. The superset of potential target stoichiometries for the reaction network could be extracted from the chemical group matrix according to the method proposed by Yin [29]. Accordingly, in the current system there are two obvious independent reactions to consider, which in matrix form are as follows:

$$(\boldsymbol{\upsilon}_{r\times s^*})_{\text{target}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & -4 & -1 & 4 \end{bmatrix}. \tag{29}$$

Upon projecting these target vectors onto the observation space $\mathbf{V}^{\mathbf{T}}_{3\times 4}$, the projected stoichiometries are as follows:

$$(\boldsymbol{\upsilon}_{r\times s^*})_{\text{proj}} = \begin{bmatrix} 1.0000 & -1.0000 & 0.0011 & 0.0003 \\ -0.0000 & -4.0000 & -1.0062 & 3.9984 \end{bmatrix}. \tag{30}$$

It can be seen that the sum of square errors between target and projected stoichiometries for the first and second reactions are 1.2×10^{-6} and 4.13×10^{-5}, respectively. By most standards, these errors are very small. Thus, it can be confidently concluded that the target vectors are the true reaction stoichiometries observed in this current system.

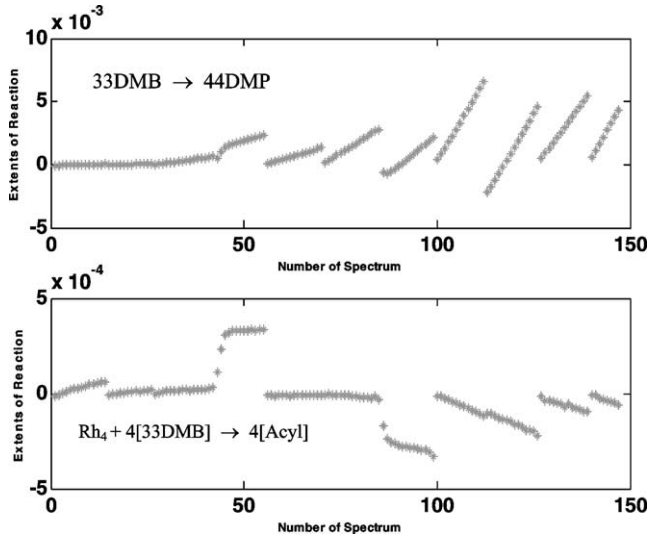

Fig. 10. The extent of reactions for the two independent reactions.

4.8. Estimation for extent of reactions

The extent of reactions was calculated using Eq. (22), and their values for every observation in the semibatch reaction are presented in Fig. 10.

The results show the advancement for the two observable and independent reactions. The first is associated with the conversion of alkene to aldehyde, and the second corresponds to the formation of the organometallic intermediate $RCORh(CO)_4$ from alkene and $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$. A few negative values are seen for the extents of the first reaction. These negative extents are associated with the hexane additions, and this indicates a mixing problem. More specifically, it appears to be a mixing/mass transfer problem. The hexane added at each step is deficient in dissolved CO and hydrogen. Time is required (circa 10 min) before the solution is uniformly saturated with dissolved gases again.

From Fig. 9 jumps can be seen in the mole profiles of alkene and $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$ at some of the reaction steps, such as at steps 3 to 4, when some catalyst was added to the reaction. The jumps were not smooth. Some transients could be observed. After slowly increasing, the system achieves a well-mixed state, and the moles of $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$ decrease again.

As was expected, all extents of the first reaction show a positive gradient, which means that throughout the whole semibatch reaction period aldehyde was formed. However, for the second reaction, inconsistent profiles for the extent of reactions are seen. Some have positive gradients and some have negative gradients. This result is not expected since the moles of $RCORh(CO)_4$ are increasing during each step. The reaction is moving forward and a positive gradient is expected.

Close inspection of the mole time series for these four components indicates that a very ill-conditioned problem is present. The mole ratios of organic compounds (aldehyde and alkene) to organometallic compounds ($Rh_4(\sigma\text{-}CO)_9(\mu\text{-}$

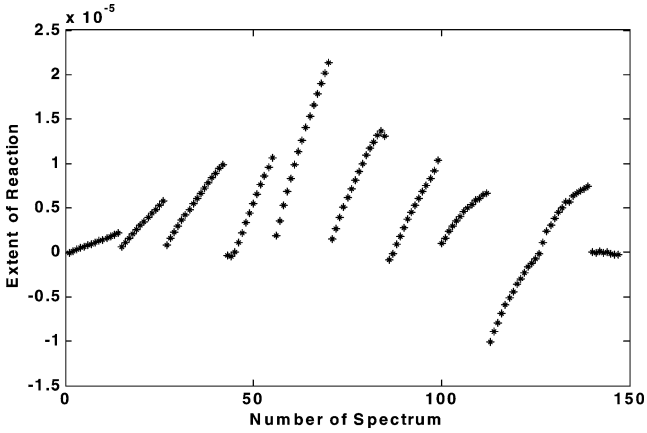

Fig. 11. The extent of reaction for $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3 \rightarrow 4[RCORh(CO)_4]$.

$CO)_3$ and $RCORh(CO)_4$) are very high (more than 100 times). Therefore during the process of regression using multiple linear regression, there is a numerical sensitivity in the extent of reaction for Eq. (31) (note: CO and H_2 are not observable quantities, and are therefore not available for the stoichiometry TFA test):

$$Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3 + 8CO + 2H_2 + 4[33DMB] \rightarrow 4[RCORh(CO)_4]. \quad (31)$$

Further analysis confirms that the error for the extent of reactions is very large, i.e., more than 40%. The analysis was performed using Eqs. (32) and (33). First, the estimated mole changes are calculated using the regressed extent of reactions. These are then compared to the experimental mole changes.

$$\overline{\overline{\Delta \mathbf{N}_{ke\times s^*}}} = \boldsymbol{\xi}_{ke\times r}\boldsymbol{\upsilon}_{r\times s^*}, \quad (32)$$

$$\text{Error} = \frac{\sum\sum\left|\Delta \mathbf{N}_{ke\times s^*} - \overline{\overline{\Delta \mathbf{N}_{ke\times s^*}}}\right|}{\sum\sum\left|\Delta \mathbf{N}_{ke\times s^*}\right|} \times 100\%. \quad (33)$$

Since the negative gradients for the extent of reactions arise from a numerical sensitivity alone, i.e., from the big difference between the alkene and the organo-rhodium concentrations, a simple solution to the problem can be achieved by rewriting the stoichiometric reaction as $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3 \rightarrow 4[RCORh(CO)_4]$. This was done, and the extent of reactions was reevaluated using the mole change data. The results are shown in Fig. 11.

Considerable improvement is obtained. Almost all the data now show a positive advancement. The only spurious data are found in step 9. The reason appears to be that the first spectrum of the step was measured a brief moment too soon. This incorrect "initial" spectrum then leads to a propagation of errors in all subsequent calculations for the step.

The above observations can be generalized to a rather important conclusion. If large mole differences between species exist in a data set (and this is the normal case for catalytic reactions), then the reactions involving both major and minor species should be rewritten to include only the latter. This should solve the problem of abnormal gradients in the extent of reactions for most cases.

This study shows that only a few well-planned perturbations of a reaction system (in this case an 11-step single semibatch) are needed for proper experimental design of a reactive system. The resulting spectra properly span all the vector space of the observables. More importantly, this study shows that algebraic system identification of a reactive system can be achieved. All pure-component spectra, both major and minor components, could effectively be recovered using the BTEM algorithm without any spectral preprocessing. Quantification of primary observable components such as 44DMP, 33DMB, $Rh_4(\sigma\text{-}CO)_9(\mu\text{-}CO)_3$, and $RCORh(CO)_4$ was obtained.

As with any extensive procedure, the developed methodology has a few obvious limitations. Currently, spectral reconstruction from in situ observations is limited to species with ppm concentrations or above. Smaller spectroscopic signals are difficult to recover. Also, it is still not possible to obtain accurate mole time series for the most minor observable species present at ppm concentrations, i.e., in this study $Rh_4(\sigma\text{-}CO)_{12}$, 2M33DMB, and unknown.

This study indicates that statistical rank tests for the determination of the number of observable species could only give a first approximation. F tests predicted the presence of 9 observable species, but spectral reconstruction using BTEM could recover 10 observable species, and the addition of the TFA results confirm the presence of 12 species. Therefore, the first approximation obtained using an F test may be used only as a starting point for the minimum number of species that may be recovered by the BTEM algorithm or combined BTEM and TFA. The best estimate of the number of observable species s is obtained after exhaustive spectral reconstructions have been performed.

5. Conclusion

Algebraic system identification was successfully performed on the homogeneous rhodium-catalyzed 3,3-dimethylbut-1-ene hydroformylation reaction using in situ FTIR spectroscopy. The algebraic chemical information retrieved included the number of observable species and reactions, the pure-component spectra of observable species, time-dependent moles, reaction stoichiometries, and extent of reactions. Thus, subsequent kinetic studies can proceed. In addition, current results also indeed show that a single semibatch experimental reaction is sufficient for rapid and cost-effective spectroscopic system identification of reactive organometallic and homogeneous catalytic systems. The spirit of the methodology presented here appears to be applicable to a very wide range of reactive systems in the chemical sciences.

References

[1] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
[2] C. Li, E. Widjaja, M. Garland, J. Catal. 213 (2003) 126.
[3] Y.Z. Liang, O.M. Kvalheim, R. Manne, Chemom. Intell. Lab. Syst. 18 (1993) 235.
[4] P.C. Sabatier, Applied Inverse Problems, Springer, Berlin, 1978.
[5] E. Furusjo, L.G. Danielsson, Anal. Chim. Acta 373 (1998) 83.
[6] S. Bijlsma, D.J. Louwerse, A.K. Smilde, AIChE J. 44 (1998) 2713.
[7] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta 396 (1999) 231.
[8] S. Bijlsma, PhD thesis, University of Amsterdam, 2000.
[9] A. De Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 49.
[10] T.J. Thurston, R.G. Brereton, Analyst 127 (2002) 659.
[11] K.H. Ebert, P. Deuflhard, W. Jager, Modelling of Chemical Reaction Systems, Springer, Berlin, 1981.
[12] P. Deulfhard, E. Hairer, Numerical Treatment of Inverse Problems in Differential and Integral Equations, Birkhauser, Boston, 1983.
[13] Frohning, C.W. Kohlpaintner, in: B. Cornils, W.A. Herrmann (Eds.), Applied Homogeneous Catalysis with Organometallic Compounds: A Comprehensive Handbook, vol. 1, Wiley–VCH, Weinheim, 1996, Chap. 2.
[14] P.W.N.M. Van Leeuwen, in: P.W.N.M. Van Leeuwen, C. Claver (Eds.), Rhodium Catalyzed Hydroformylation, Kluwer Academic, Boston, 2000, pp. 1–15.
[15] M. Garland, G. Bor, Inorg. Chem. 28 (1989) 410.
[16] M. Garland, P. Pino, Organometallics 10 (1990) 1693.
[17] M. Garland, Organometallics 12 (1993) 535.
[18] C. Fyhr, M. Garland, Organometallics 12 (1993) 1753.
[19] J. Feng, M. Garland, Organometallics 21 (1993) 417.
[20] G. Liu, R. Volken, M. Garland, Organometallics 18 (1999) 3429.
[21] M. Garland, E. Visser, P. Terwiesch, D.W.T. Rippin, Anal. Chim. Acta 351 (1997) 337.
[22] C.E. Shannon, Bell System Tech. J. 3 (1948) 379.
[23] E. Widjaja, C.Z. Li, M. Garland, Organometallics 21 (2002) 1991.
[24] W. Chew, E. Widjaja, M. Garland, Organometallics 21 (2002) 1982.
[25] C.Z. Li, E. Widjaja, W. Chew, M. Garland, Angew. Chem. 41 (2002) 3785.
[26] W. Chew, PhD thesis, National University of Singapore, 2003.
[27] T. De Donder, Lecons de Thermodynamique et de Chimie Physique, Gauthier–Villars, 1920.
[28] D. Bonvin, D.W.T. Rippin, Chem. Eng. Sci. 45 (1990) 3417.
[29] F. Yin, Ind. Eng. Chem. Res. 29 (1990) 34.
[30] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Univ. Press, New York, 1985.
[31] D.F. Shriver, M.A. Drezdzon, The Manipulation of Air-Sensitive Compounds, Wiley, New York, 1986.
[32] L. Chen, M. Garland, Appl. Spectrosc. 56 (2002) 1422.
[33] G.W. Stewart, SIAM J. Sci. Statis. Comput. 5 (1984) 403.
[34] A. Corana, M. Marchesi, C. Martini, S. Ridella, ACM Trans. Math. Softw. 13 (1987) 262.
[35] R. Aris, R.H.S. Mah, Ind. Eng. Chem. Fundam. 2 (1963) 90.