Multivariate analysis of homogeneous nucleation rate measurements. Nucleation in the \( p \)-toluic acid/sulfuric acid/water system

Robert McGraw
Environmental Sciences Department, Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, New York 11973, USA

Renyi Zhang
Department of Atmospheric Sciences, Texas A&M University, College Station, Texas 77843, USA

(Received 16 May 2007; accepted 6 December 2007; published online 13 February 2008)

Recent kinetic extensions of the nucleation theorem suggest that the logarithm of the steady-state nucleation rate has strong multilinear dependence on the log concentrations of condensable species present in the vapor phase. A further remarkable result is that the coefficients of this linear dependency provide a direct determination of the molecular content of the critical nucleus itself. Building on these results, the powerful utility of multivariate statistical methods is demonstrated here for physically based parametrization and interpretation of nucleation rate measurements. The new approach is applied to recent measurements by Zhang et al. [Science 304, 1487 (2004)] on the \( p \)-toluic acid/sulfuric acid/water ternary vapor system. A linear minimum variance parametrization for nucleation rate dependence on vapor composition, accurate over the range of the measurements, is obtained. Estimates of critical nucleus molecular composition are also presented. These suggest that a single molecule of \( p \)-toluic acid present in the critical nucleus is sufficient to trigger a ternary nucleation event. Efforts under way to apply the new methods to analysis of new particle formation in the atmosphere are discussed. © 2008 American Institute of Physics. [DOI: 10.1063/1.2830030]

I. INTRODUCTION

Increasingly, the need to model nucleation processes thought to be important for applications exceeds the predictive capability of phenomenological nucleation theories including, most notably, classical nucleation theory (CNT). Examples from the field of atmospheric science include current efforts to model frequent and widespread observations of new particle formation, which occur through various mechanisms of gas to particle conversion. Early models were limited almost exclusively to binary sulfuric acid-water nucleation. More recent studies support the binary mechanism as consistent with the observations of new particle formation in the upper troposphere; however, measured rates substantially in excess of the binary rate are often found, especially in the marine boundary layer and over continental sites. One explanation for the discrepancy is the participation of a third vapor component that, in concert with sulfuric acid and water, results in enhancement of the nucleation rate over the binary rate. Suggested third species include ammonia and condensable organic acids.

Classical nucleation theory has been applied to the ternary ammonia/sulfuric acid/water system, but model predictions have not been compared with the experiment in other than a qualitative way due to the lack of quantitative measurements. Initial laboratory measurements were reported for this system, and significant enhancement of the nucleation rate over the binary rate was found, but estimates of the ammonia vapor concentrations present in the experimental nucleation region were reported as uncertain. Nevertheless, a recent comparison of the Ball et al. data with a CNT-based parametrization of the ternary nucleation rate suggests that CNT greatly overestimates the ammonia enhancement by predicting even larger enhancement of nucleation rate. Laboratory measurements are available for the ternary organic acid/sulfuric acid/water system for several organic acids, but inadequate knowledge of bulk solution properties, such as surface tension and partial vapor pressures over these ternary mixtures, would seem to preclude application of phenomenological approaches, including CNT, to any quantitative analysis of these potentially important atmospheric systems. In this paper, we will examine the \( p \)-toluic acid/sulfuric acid/water system using a new approach based on the recent development of nucleation theorems and application of multivariate statistical methods.

Even when the bulk properties of a condensate are known, classical nucleation theory requires correction to ensure satisfaction of the law of mass action, internal consistency, and proper accounting of translational and rotational degrees of freedom during the mapping of molecular clusters to capillary drops. Furthermore, it is known that with surface active systems such as ethanol/water, the theory sometimes produces unphysical predictions (negative occupation numbers of water molecules in the critical nucleus). It is generally held that all of these difficulties would vanish in a fully molecular nucleation theory.

The present paper introduces an alternative approach using the nucleation theorem to establish optimal coordinates for characterizing the rate of nucleation in multicomponent vapor systems. Rooted in the fundamental law of mass action
and principle of detailed balance, nucleation theorems inherit a molecular character not found in CNT. When used in conjunction with experimental measurements, nucleation theorems provide molecular-level information on nucleus composition, hence, on nucleation mechanism. Unlike CNT, however, nucleation theorems make no a priori prediction of absolute nucleation rate—additional input from modeling or measurement is required. Section III applies linear regression, using coordinates set by the nucleation theorem, to provide a fully multivariate description of a ternary nucleation data set. This approach allows one to describe relative sensitivity of nucleation rate to simultaneous changes in vapor composition and temperature, including changes in concentration of multiple vapor components. Temperature dependence is addressed in a future publication. By applying the new methods, we determine in Sec. III that a single p-toluic acid molecule in the critical nucleus is the best estimate for the number needed to initiate a ternary nucleation event. Finally, the new methods are shown to yield a simple yet accurate parametrization for the ternary nucleation rate.

II. KINETIC EXTENSION OF THE NUCLEATION THEOREM FOR MULTICOMPONENT VAPORS: SELECTION OF COORDINATES

The nucleation theorem (NT) is a thermodynamic result relating the sensitivity of the nucleation barrier height to changes in log vapor species concentration. A clear limitation of the NT is that the barrier height cannot be measured directly, unlike the nucleation rate itself. This situation was improved greatly through the development of kinetic nucleation theorems (KNTs) beginning with the work of Ford. KNTs provide a direct calculation of the rate sensitivity by incorporating the full Becker-Döring molecular cluster summation for the nucleation rate, thus, including contributions from the multistate kinetics as well as thermodynamics to the rate. More recently, the KNT has been shown to follow very generally from the law of mass action and principle of detail balance and extended to several models of multicomponent nucleation rate. The KNT is next briefly reviewed as originally developed for single component vapors. This is followed with the introduction of a simple transition state model designed to motivate its extension to systems having multiple components.

A. Composition variation at constant temperature

Constant temperature differentiation of the Becker-Döring expression for the steady-state nucleation rate \( J \) gives

\[
\left( \frac{\partial \ln J}{\partial \ln n_1} \right)_T = \left( \frac{\partial \ln J}{\partial \ln S_1} \right)_T = 1 + \bar{g} \approx 1 + g^*. 
\]  

(2.1)

Here \( n_1 \) is the concentration of condensable monomer in the vapor phase, \( S_1 = n_1/n_1^0 \) is the saturation ratio, and \( n_1^0 = n_1^{eq}(T) \) is the vapor concentration in equilibrium with bulk phase. The average is defined as

\[
\bar{g} = \frac{\sum_{g=1}^{G} \frac{1}{\beta_g n_g^*} g}{\sum_{g=1}^{G} \frac{1}{\beta_g n_g}} = \sum_{g=1}^{G} g P(g), \tag{2.2}
\]

where the summation runs from monomer through clusters well beyond the critical cluster size and \( P(g) \) is the normalized distribution of reciprocal cluster growth rates defined by the second equality. \( n_g \) is the constrained equilibrium concentration of clusters of size \( g \) (clusters containing \( g \) monomeric units) and \( \beta_g \) is the rate constant for monomer addition to a single \( g \) cluster,

\[
\beta_g = \frac{\alpha \bar{c}_1}{4} s_g n_1 = \frac{kT}{2 \pi m_1} s_g n_1, \tag{2.3}
\]

where \( \alpha \) is the accommodation coefficient, \( s_g \) is the surface area of a \( g \) cluster and \( \bar{c}_1 \) is the mean molecular speed of a molecule of mass \( m_1 \). The approximate equality of Eq. (2.1) holds whenever \( P(g) \) is sharply peaked at a critical cluster size \( g^* \). For the present analysis, we will use the leading partial derivative in Eq. (2.1), expressing results in terms of measured vapor number concentrations, so as to avoid having to compute saturation ratios in cases for which equilibrium vapor pressures are not well known.

B. Extensions to multicomponent systems

Early applications of the thermodynamic nucleation theorem to multicomponent systems suggested that measurements of the relative sensitivity of nucleation rate with respect to change in the saturation ratio of each vapor component provide direct information on critical nucleus composition. Multicomponent extension of the KNT, on the other hand, requires an explicit expression for the nucleation rate. Unfortunately, a suitable extension of the closed-form Becker-Döring rate, used to obtain Eqs. (2.1)–(2.3), to multicomponent systems is not available without further approximation. Thus, rigorous extensions of the KNT have only been achieved for special model cases. These include the Shugard-Heist-Reiss (SHR) binary nucleation model, shown to be a very accurate approximation to the full nucleation kinetics for sulfuric acid-water mixtures and, for the multicomponent case under the approximation that the free-energy saddle surface has quadratic form. Here, we motivate the multicomponent extension using an even simpler transition state model for which a well-defined critical cluster serves as the transition state, and the nucleation rate is determined from the barrier crossing rate, which is equal to the sum of the net fluxes contributed by each condensable vapor species to growth beyond the critical cluster size. The model is largely thermodynamic in its focus on the population of clusters of critical size, but it does include a kinetic prefactor (one consistent with the law of mass action) which gives rise to departures from the thermodynamic result [e.g., the \( \delta \) terms present in Eqs. (2.8) and (2.9) below].

Illustrating for two components (the model is readily extended to any number of components), let
be the overall nucleation rate. Taking the logarithm and expressing the result in differential form gives

\[
\frac{d \ln J}{d \ln n_1} = \frac{J_1}{J_1 + J_2}, \quad \frac{d \ln J}{d \ln n_2} = \frac{J_2}{J_1 + J_2}.
\]

Here,

\[
J_i = \kappa_i \beta_i (g_{1i}^*, g_{2i}^*) n_i (g_{1i}^*, g_{2i}^*)
\]

is the net flux for growth of critical clusters through addition of species \(i\). \(\beta_i (g_{1i}^*, g_{2i}^*)\) is the rate at which molecules of species \(i\) add to a single cluster having \(g_{1i}^*\) molecules of species 1 and \(g_{2i}^*\) molecules of species 2, and \(n_i (g_{1i}^*, g_{2i}^*)\) is the constrained equilibrium cluster concentration where parenthesis are used in place of subscripts when more than one component is present. For a single component vapor \(\beta_i (g_{1i}^*)\) is identical to \(\beta_n\) of Eq. (2.3) with \(s_i\) replaced by the surface area of the critical cluster. \(\kappa_i\) is a barrier transmission coefficient, assumed here to be constant, along the growth coordinate of species \(i\). In the absence of recrossing, \(\kappa_i\) has the value unity, as in transition state theory, and is otherwise a correction to that result. (See Ref. 16 for a discussion of \(\kappa\) in the context of classical nucleation theory.)

With these definitions, the differentiation in Eq. (2.5) is readily carried out. First, applying the law of mass action to the concentration of critical clusters:

\[
\frac{\partial \ln n_i (g_{1i}^*, g_{2i}^*)}{\partial \ln n_1} = g_{1i}, \quad \frac{\partial \ln n_i (g_{1i}^*, g_{2i}^*)}{\partial \ln n_2} = g_{2i}.
\]

Note that this result applies to clusters of any size, not just critical size. Combining Eqs. (2.5)–(2.7), gives the constant temperature KNT for this model

\[
\frac{\partial \ln J}{\partial \ln n_1} = \frac{J_1}{J_1 + J_2} (g_{1i}^* + 1), \quad \frac{\partial \ln J}{\partial \ln n_2} = \frac{J_2}{J_1 + J_2} g_{2i}^*.
\]

in agreement with each of the above-mentioned models for which the KNT has been developed. Thus, we obtain agreement with the SHR model for \(J_2/J_1 = 1, J_1/J_0 = 0\), where component 2 is sulfuric acid. Equation (2.8) also agrees with the quadratic free-energy surface model, wherein \(J_1\) and \(J_2\) represent components of the net nucleation flux through the saddle point region. Multicomponent generalization of the binary case is easily carried out and gives a similar result

\[
\frac{\partial \ln J}{\partial \ln n_i} = g_i^* + \delta_i.
\]

where \(\delta_i = J_i/J\) and \(n_i\) is the monomer number concentration of species \(i\). Note that the \(\{J_i\}\), giving the net currents for molecular addition to the critical cluster, are positive quantities. Thus, \(0 < \delta_i < 1\) and \(\Sigma \delta_i = 1\).

Equation (2.9) suggests a local multilinear expansion for \(\ln J\) about some arbitrary reference steady-state condition

\[
\ln J = \ln J_0 + \sum_i \left( \frac{\partial \ln J}{\partial \ln n_i} \right) d \ln n_i,
\]

where the summation is over the number of condensable components present in the vapor phase. On integration, and neglect of higher-order terms, Eq. (2.10) gives

\[
\ln J = \ln J_0 + \sum_i (g_i^* + \delta_i) (\ln n_i - \ln n_i^0).
\]

A sub- or superscript “0” refers to a specific reference condition (e.g., conditions at the centroid of an experimental data set). Equation (2.11) suggests a multilinear form for the logarithmic rate in composition coordinates \((\ln n_i)\), expected to be accurate over a limited range of composition for which the coefficients can be treated as constants. The full nucleation rate surface will, of course, be quite complicated and nonlinear as critical cluster size varies with larger changes in environmental conditions. The value of Eq. (2.11) is that, as found empirically, there appears to be a strong propensity for the linear regime to hold over a surprisingly wide rate range. (Exceptionally cases exhibiting more complex behavior are described in Sec. III E.) Application of Eq. (2.11) to analysis and parametrization of nucleation rate measurements for the ternary system will now be demonstrated.

### III. MULTIVARIATE ANALYSIS OF A TERNARY SYSTEM AT CONSTANT TEMPERATURE

In this section, we analyze measurements from Ref. 3 for the \(p\)-toluic acid/sulfuric acid/water system. As already mentioned, nucleation rate measurements tend to support a strong propensity to linear behavior when results are plotted in the \(\log(J)\) versus either \(\log(S)\) or \(\log(n_i)\) coordinates suggest by the NT/KNT. An analysis of higher derivatives of the nucleation rate at constant temperature shows that this will be the case when the variance of the distribution \(P(g)\) [Eq. (2.2)] is small, as this variance turns out to be proportional to the rate of change of \(g\) with change in vapor concentration, which will then also be small.\(^{13}\) This suggest writing the ternary nucleation rate for the organic acid system, \(J_T\) in the multilinear form of Eq. (2.11), which on exponentiation yields

\[
J_T = J_T^0 \left( \frac{[\text{H}_2\text{SO}_4]}{[\text{H}_2\text{SO}_4]_0} \right)^{a + \delta_i} \left( \frac{\text{RH}}{\text{RH}_0} \right)^{b + \delta_i} \left( \frac{[\text{Org}]}{[\text{Org}]_0} \right)^{c + \delta_i}.
\]

Here, RH is water relative humidity (saturation ratio \(\times 100\)), \([\text{H}_2\text{SO}_4]\) is the concentration of sulfuric acid (cm\(^{-3}\)), and \([\text{Org}]\) is the concentration of \(p\)-toluic acid in parts per \(10^6\).
(ppb) in the vapor phase. A sub- or superscript 0 again refers to the reference condition. Consistency with Eq. (2.11) follows where \( a, b, \) and \( c \) are the number of molecules of sulfuric acid, water, and organic acid, respectively, in the critical nucleus and \( \delta_a, \delta_b, \) and \( \delta_c \) are the small quantities (between 0 and 1) that relate to the direction of the nucleation flux through the transition zone.

The experimental measurements are shown in Fig. 1. Zhang et al.\(^3\) also include nucleation rate measurements for the binary sulfuric acid-water subsystem (solid triangles in the figure). Although the effect is small here (as discussed further in Sec. III B), the ternary measurements (circles) should be corrected by subtracting off the binary rate under the presumption that, when all three components are present, nucleation can occur through either the binary or ternary pathways. We will assume these pathways are additive and independent (see Sec. III B). Accordingly, we subtract the binary rate \( J_B \) measured for the sulfuric acid-water subsystem [and fit using Eq. (3.9) below] from the measured total rate \( J \), when all three components are present, to obtain the nucleation rate specifically for the ternary mechanism:

\[
J_T = J - J_B.
\]

The measurements are expressed in \( \{z_i, x_i, y_i\} \) coordinates, with the dependent variable listed first, for \( i = 1, \ldots, N \). Here, \( N = 18 \) is the number of ternary measurements. We use the nucleation theorem-motivated logarithmic coordinates

\[
z = \log_{10}(J_T \text{ cm}^{-3} \text{ s}^{-1}), \quad x = \log_{10}[\text{H}_2\text{SO}_4, \text{ molecules cm}^{-3}], \quad \text{and} \quad y = \log_{10} [\text{Org}, \text{ ppb}].
\]

Temperature (298 ± 2 K), RH (5%), and total pressure (760 torr) were held constant during the measurements.

Analysis of the covariance matrix provides both an algebraic framework suitable for conventional least-squares regression and a geometric framework suitable for visualization of data sets in terms of principal coordinates.\(^3\) Matrix elements are generated through the following sequence of steps: (1) Compute the coordinate means \( \{\mu_z, \mu_x, \mu_y\} \), and the centered coordinates, \( \bm{w}(i) = \{z_i - \mu_z, x_i - \mu_x, y_i - \mu_y\} = \{\tilde{z}_i, \tilde{x}_i, \tilde{y}_i\} \), (2) form the \( 3 \times 3 \) covariance matrix, \( \Sigma \) [see Eq. (3.4) below] whose elements are the variances and covariances obtained by summing over the measurements

\[
\begin{align*}
\langle z \tilde{z} \rangle & = N^{-1} \sum_{i=1}^{N} (z_i - \mu_z)(z_i - \mu_z), \\
\langle z \tilde{x} \rangle & = N^{-1} \sum_{i=1}^{N} (z_i - \mu_z)(x_i - \mu_x), \\
\langle z \tilde{y} \rangle & = N^{-1} \sum_{i=1}^{N} (z_i - \mu_z)(y_i - \mu_y),
\end{align*}
\]

etc.

This construction is sufficient for linear regression analysis of the measurements described below. A full principal components analysis (PCA) requires the additional step of solving the eigenvalue problem associated with \( \Sigma \). The normalized eigenvectors of \( \Sigma \) are the principal component basis vectors \( \{\bm{v}_1, \bm{v}_2, \bm{v}_3\} \) and the corresponding sorted eigenvalues \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \) give the variances of the data set along the directions of the principal components.

The transformed (principal) coordinates, shown in Fig. 2, are obtained as the scalar products: \( \eta_1(i) = \bm{w}(i) \cdot \bm{v}_1 \), etc. According to the KNT [Eq. (2.11)] these points should lie on a first order be coplanar, lying in the \( (\eta_1, \eta_2) \) plane (orthogonal to the principal coordinate of smallest variance, \( \eta_3 \)). Apart from a small amount of apparently random noise, this indeed appears to be the case with no sign of any systematic curvature present (Fig. 2). The \( (\eta_1, \eta_2) \) plane is close to but not the optimal planar fit to the data in the least-squares sense. The optimal plane (hyperplane in higher dimension) has minimum error variance and is known as the linear minimum variance estimator (LMVE) for the data set.\(^8\) We will use the LMVE, also obtained from the coordinate means and covariance matrix, to parametrize the ternary nucleation rate. For this purpose it is useful to write vectors and matrices in block form, separating the dependent variable \( z \) from the independent variable set \( \{x, y\} \). Thus, for the mean

\[
\mu = \begin{pmatrix} \mu_z \\ \mu_x \\ \mu_y \end{pmatrix} = \begin{pmatrix} \mu_z \\ \mu_x \\ \mu_{xy} \end{pmatrix}
\]

with \( \mu_{xy} = (\mu_2, \mu_1)^T \) and, similarly, for the covariance matrix

\[
\Sigma = \begin{pmatrix} \sigma_z^2 & \rho_zx & \rho_zy \\ \rho_xz & \sigma_x^2 & \rho_xy \\ \rho_yz & \rho_yx & \sigma_y^2 \end{pmatrix}
\]
\[
\Sigma = \begin{pmatrix}
(\bar{z} \bar{z}) & (\bar{z} \bar{x}) & (\bar{z} \bar{y}) \\
(\bar{x} \bar{z}) & (\bar{x} \bar{x}) & (\bar{x} \bar{y}) \\
(\bar{y} \bar{z}) & (\bar{y} \bar{x}) & (\bar{y} \bar{y})
\end{pmatrix} = \begin{pmatrix}
\Sigma_{zz} & \Sigma_{zx} & \Sigma_{zy} \\
\Sigma_{xz} & \Sigma_{xx} & \Sigma_{xy} \\
\Sigma_{yz} & \Sigma_{xy} & \Sigma_{yy}
\end{pmatrix},
\]

where \( \Sigma_{xy} \) is the lower right \( 2 \times 2 \) block of \( \Sigma, \Sigma_{zx}, \Sigma_{zy} = \langle \bar{z} \bar{x} \rangle \), \( \Sigma_{xz} = \langle \bar{x} \bar{z} \rangle \), \( \Sigma_{xy} = \langle \bar{x} \bar{y} \rangle \), and \( \Sigma_{yy} = \langle \bar{y} \bar{y} \rangle \). The LMVE for the ternary nucleation rate, \( z_T \), takes the form

\[
z_T = \mu_z + \Sigma_{c,xy} \Sigma_{xy}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} - \mu_{xy}.
\]

(3.5)

Evaluating the means and covariance matrix elements for the ternary data set, and substituting these into Eq. (3.5) gives the result

\[
z_T(\text{LMVE}) = -76.75 + 8.12x + 1.86y.
\]

(3.6)

The subscript refers specifically to the ternary mechanism and Eq. (3.6) provides an especially compact parametrization of the ternary nucleation rate. A measure of the quality of fit is provided by \( R^2 \), the coefficient of determination or fraction of variance explained. The standard formula is (e.g., Faraway)

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (z_i - \bar{z}_T)^2}{\sum_{i=1}^{N} (z_i - \bar{z}_L)^2} = 0.98.
\]

(3.7)

Equation (3.6) could have been obtained directly from the data set using any standard multilinear regression package, and the analysis here, based on the covariance matrix, could be considered “overkill.” Nevertheless, PCA provides a more general framework that can also be used to identify the most important species combinations participating in a nucleation process, thereby potentially reducing the dimensionality of the problem for cases that the number of species present is large. It is anticipated that these considerations will be important in future studies of atmospheric nucleation processes—e.g., in the correlation of field measurements of new particle formation rate with trace species concentrations—and for this reason we have described the more general method.

**A. Nucleus composition**

The nucleation theorem, applied now to Eq. (3.6), takes the especially simple form, \( \frac{\partial z_T}{\partial x} = 8.12z_{\text{H}_2\text{SO}_4} + \delta_{\text{H}_2\text{SO}_4} \), \( \frac{\partial z_T}{\partial y} = 1.86z_{\text{Org}} + \delta_{\text{Org}} \). Taking into account that the \( \delta \) values are positive and less than unity, we obtain the critical nucleus molecular composition: approximately eight sulfuric acid molecules and a single molecule of the organic acid.

Assuming the measurement errors are Gaussian and evenly distributed over the measurements, we determine the confidence region associated with these estimates of nucleus composition. The 95% confidence intervals for the coefficients in Eq. (3.6) are

\[
\begin{align*}
g^*_{\text{H}_2\text{SO}_4} &+ \delta_{\text{H}_2\text{SO}_4} = 8.12 \pm 0.57, \\
g^*_{\text{Org}} &+ \delta_{\text{Org}} = 1.86 \pm 0.66.
\end{align*}
\]

(3.8)

Although the best estimator points to a single molecule of the organic acid present in the nucleus, the confidence intervals of Eq. (3.8) do not exclude the (less likely) possibility that two are present. Without variation of RH, the number of water molecules present in the critical cluster cannot be determined.

**B. Total nucleation rate from the binary and ternary pathways**

Measurements for the binary sulfuric acid-water reference system (triangles) are also shown in Fig. 1. The least-squares regression line is

\[
z_B = 1 - 9.11 + 9.17y
\]

(3.9)

indicating approximately nine molecules of sulfuric acid in the critical nucleus along the binary nucleation path.

Regarding the binary and ternary nucleation rates as occurring through independent pathways, the total nucleation rate will, as noted previously, be a sum of these: \( J = J_B + J_T \). Substitution from Eqs. (3.9) and (3.6), respectively, gives

\[
J(x,y) = 10^{-9.11+9.17y} + 10^{-76.75+8.12x+1.86y}.
\]

(3.10)

Equation (3.10), with \( y \) set to the logarithms of each of the two measured p-toluic acid vapor concentrations, \( \log_{10}(0.2 \text{ppb}) \) and \( \log_{10}(0.4 \text{ppb}) \), gives the (nearly parallel) surface projections shown by the dashed curves in Fig. 1. The experimental data points are also shown on this log-log coordinate scale suggested by the KNT and are seen to be in excellent agreement with the fit. The solid line is the binary fit obtained either from Eq. (3.9) or from Eq. (3.10) with \( y = -\infty \). The parametrized model rate from Eq. (3.10) is compared with the totality of measurements, both binary and ternary, in Fig. 3 together with the 1-1 line drawn for comparison. The agreement shows that Eq. (3.10) is an excellent predictor of nucleation rate over the range of the measurements.
Figure 4 (solid curve) shows the total nucleation rate from Eq. (3.10) for the constant values of vapor phase sulfuric-acid concentration and RH indicated in the figure. The figure shows binary behavior, characterized by insensitivity of the nucleation rate to p-tolue acid concentration below about 10 ppt (parts per trillion), a minimum detection threshold for the organic of about 10–30 ppt, and abrupt crossover to ternary behavior at higher concentrations. The solid curve represents a cross section of the total rate surface [Eq. (3.10)]. Although this surface is more complicated than the planar, single-path result, an extension of the nucleation theorem gives the slope at any point along the solid curve in terms of the flux-weighted critical nucleus composition over the two distinct nucleation pathways

\[
\frac{\partial \ln J}{\partial \ln [\text{org}]}_{\text{TRH}{[\text{H}_2\text{SO}_4]}} = \left( \frac{\partial \ln (J_B + J_T)}{\partial \ln [\text{org}]} \right)
= \frac{1}{J_B + J_T} \left( J_B \frac{\partial \ln J_B}{\partial \ln [\text{org}]} + J_T \frac{\partial \ln J_T}{\partial \ln [\text{org}]} \right)
= \frac{J_B}{J_B + J_T} (0) + \frac{J_T}{J_B + J_T} (1.86). \quad (3.11)
\]

The quantities in parenthesis to the right of the last equality give the coefficients of \( y \), from Eqs. (3.6) and (3.9), for the binary and ternary pathways and these are weighted by the flux ratios. Extension of the nucleation theorem to include the idea of flux weighting over multiple paths was introduced in Ref. 13. A similar two-pathway dependence was observed by Wagner and Strey in a study of nucleation in supersaturated water/n-nonane vapor mixtures.\(^{19}\)

Analysis of the ternary measurements directly, i.e., without correcting for the binary rate, yields in place of Eq. (3.8) the slightly modified result

\[
g^{\text{H}_2\text{SO}_4} + g^{\delta_{\text{H}_2\text{SO}_4}} = 8.24 \pm 0.54, \tag{3.8'}
\]

having a slightly smaller error range and a slightly higher value for \( R^2 \). Equation (3.10), which is based on inclusion of both the binary and ternary channels, has the advantage of reducing to the binary rate in the limit that the organic acid concentration is small. On the other hand, correction for the binary rate requires combining two different sets of measurements, which in itself might account for the slightly larger uncertainty.

Concentrations of sulfuric acid (SA) were measured by cluster ion mass spectroscopy (CIMS) and errors by as much as a factor of two could have affected the results. Systematic error that results in, say, a simple overestimate of SA by a constant factor will only shift the data points in Fig. 1 uniformly to the left with no change in slope or quality of fit expected. The high \( R^2 \) value associated with the planar fit to the data of Fig. 1 would seem to rule out the presence of large random errors in SA concentration measurement, but cannot be used to rule out systematic errors of the kind described.

C. Quasiunary nucleation rate

The LMVE plane described by Eq. (3.6) can be further made to collapse to a single line by simply combining the terms in \( x \) and \( y \) into an effective homomolecular monomer concentration: \( n^*_i = [\text{H}_2\text{SO}_4]^i [\text{Org}]^j \). Here, \( x_1 = 8.12/(8.12 + 1.86) \) and \( x_2 = 1.86/(8.12 + 1.86) \) are the relative fractions of sulfuric acid and \( p \)-tolue acid in the critical nucleus (apart from the \( \delta \) terms included here). Equation (3.6) thus becomes

\[
\log_{10} J_T = -76.75 + (8.12 + 1.86) \log_{10} [n^*_1]. \tag{3.12}
\]

Figure 5 shows the ternary data set in \( \{\log_{10} J_T, \log_{10} [n^*_1]\} \) coordinates (circles) and comparison with Eq. (3.12) (solid line). The use of an effective homomolecular saturation, similar to our effective monomer concentration, for con-
D. Other ternary organic acid systems

Nucleation rate measurements have recently been carried out by Zhang et al. for the cis-pinonic acid/sulfuric acid/water system.\textsuperscript{23} Here, measurements at three distinct cis-pinonic acid concentrations provide the strongest evidence to date that just a single molecule of the organic acid present in the critical nucleus is sufficient to initiate a ternary nucleation event. Details of the cis-pinonic acid measurements will be reported in a future publication.

Measurements on several other ternary organic acid systems were also reported in Ref. 3. Besides the p-toluic measurements analyzed above, the most extensive data set is for the benzoic acid/sulfuric acid/water system. Nucleation rates were reported as a functions of sulfuric acid concentration at relative humidities of 4.6 and 5% RH, in each case with benzoic acid concentration=0.1 ppb; and at 9.5% RH, with benzoic acid concentration=0.04 ppb. Variable RH and acid concentrations suggests extending the present methods using four coordinates so as to obtain the complete molecular content of the critical nucleus, including the occupation number for water molecules. We attempted such an analysis again finding approximately eight molecules of sulfuric acid in the critical nucleus. Uncertainties in the organic acid and water occupation numbers, however, were found to be too large to draw meaningful conclusions. Nevertheless, the difficulties of analysis are instructive: In addition to the RH values of 4.6 and 5% being close together, vis-à-vis measurement uncertainty, there is the important consideration of sulfuric acid hydration in the vapor phase and the effect that this hydration has on nucleation rate.\textsuperscript{14,15,24}

CIMS, used in the experiments to determine sulfuric acid vapor concentration, measures total sulfuric acid including both its hydrated and unhydrated forms. Strictly, it is the concentration of the unhydrated (free) acid that should appear in the law of mass action and in Eq. (3.1), as it is the free acid that relates to sulfuric acid activity in the vapor phase. A further complication is that the extent of hydration increases markedly from 5% RH, where there is no significant hydration, to 9.5% RH where only about 30% of the sulfuric acid molecules remain unhydrated. These estimates were previously obtained\textsuperscript{23} through a reanalysis of CIMS measurements of total sulfuric acid concentration over bulk acid-water solutions of varying composition for which the vapor phase concentrations of free acid and RH are reliably known.\textsuperscript{25} The basic idea is that subtraction of the free acid concentration from the CIMS measurements of total acid yields an experimental-based estimate of the fraction of sulfuric acid vapor tied up in hydrate form as a function of RH, thus, providing a way to correct the CIMS measurements to free sulfuric acid concentration if required.\textsuperscript{24}

Assuming 30% free acid at 9.5% RH and 100% free at 5% RH (there is some uncertainty in these values due to the reported error bars of the CIMS measurements) gives a linear relation between the occupation numbers of water and organic acid in the critical nucleus. Neither occupation number can be determined separately without a broader range of measurements, however, if one occupation number is known the other is determined by the linear relation. For example, assuming that only one molecule of benzoic acid is present in the critical nucleus (as was found for the p-toluic acid system) yields an approximate occupation number for water molecules of 17. This very rough analysis, together with the eight molecules of sulfuric acid more firmly established, in principle, determines the full nucleus composition. (Note that even if there were hydration in the p-toluic acid system at 5% RH, the analysis of the previous subsections would still hold because the RH was constant. The abscissa of Fig. 1 would then refer to total acid concentration, as measured by CIMS. Any replot of the data in terms of free acid would simply result in a uniform horizontal shift of points and lines with no change in slope.)

E. Limitations of multilinear analysis

The simple transition state model introduced in Sec. II B, used as the foundation for the multilinear analysis of Secs. II and III, works under the assumption that only a single critical cluster is present. That argument was extended in Sec. III B to allow for the possibility of two types of critical nuclei—a binary nucleus with no p-toluic acid present and a ternary nucleus containing sulfuric acid, water, and p-toluic acid. There it was shown that two planar surfaces could be merged in such a way [cf. Eq. (3.10)] so as to accurately describe the case that the overall nucleation rate is a sum of the binary and ternary rates (Fig. 3). A simple extension of the nucleation theorem [Eq. (3.11)] gave the relative rate sensitivity in terms of the flux weighting for the two distinct nucleation paths.

Examples of binary and ternary nucleation systems can be found where the measurements cannot be represented satisfactorily using either the single plane or the two merged plane configurations. Strey et al.\textsuperscript{26} show a curved binary nucleation rate surface, $J(a_1, a_2)$, for water and n-pentanol plotted in terms of the vapor phase activities, $a_1$ and $a_2$, respectively, of these two components. Viisanen and Strey\textsuperscript{27} show a highly curved three-dimensional rate surface: a constant rate cross section of the four-dimensional nucleation rate hypersurface, $J(a_1, a_2, a_3)$, for the ternary system water/ n-nonane/n-butanol. Two-dimensional projection of that constant rate surface onto the water/n-nonane coordinate plane appears to be well described by the two-pathway model with critical nuclei consisting of either pure none or pure water, but similar projections for the water/n-butanol and n-nonane/ n-butanol subsystems exhibit more complicated behavior.

By extension of the two-path analysis, systems with multiple nucleation pathways may be expected to show this kind of more complicated behavior. In their recent paper, Nellas et al. applied Monte Carlo simulation to a systematic study of binary nucleation in the n-nonane/1-alcohol series.\textsuperscript{28} Experimental activity curves of $a_2$ versus $a_1$ at constant
nucleation rate, where \( a_1 \) refers to the \( n \)-nonane vapor phase activity and \( a_2 \) the other component, were compared with simulation results and the curvature properties for each mixture interpreted using contour maps calculated for the nucleation free energy surface from Monte Carlo simulation. For the \( n \)-nonane/water and \( n \)-nonane/methanol systems, just two favorable nucleation paths are evident from the contour maps. However for the higher alcohols, including for the \( n \)-nonane/1-butanol mixture, a subsystem of the ternary case cited above, highly elongated saddle regions were found. According to the authors, the indication here is that nucleation can proceed through multiple channels in contrast to the usual one or two channels seen for other binary systems. These unusual results were interpreted as arising from “a nearly perfect balance between the tendency to mix and the tendency to demix among the \( n \)-nonane and the alcohol molecules.”

These studies show that simple multilinear methods will not always work, and provide foundation and motivation for future extensions to curved nucleation rate surfaces beyond the scope of the present study. We anticipate that nucleation theorems, generalized to include flux weighting along the lines of Eq. (3.11), will continue to provide theoretical foundation for analysis of these more complex systems.

**IV. SUMMARY AND DISCUSSION**

The present study shows how the nucleation theorem and multivariate statistical methods can be effectively combined for interpretation and parameterization of laboratory measurements of nucleation rate involving multiple vapor species. Rooted in the law of mass action and principle of detailed balance, through the nucleation theorem, the new approach incorporates very general statistical-thermodynamic properties of molecular clusters without requiring the capillarity drop approximation of classical nucleation theory. Thus, macroscopic drop properties such as density, surface tension, and equilibrium partial vapor pressures over the condensed phase are not required.

The new methods were illustrated using recent measurements on a mixed organic-inorganic system of potential importance to atmospheric new particle formation: \( p \)-toluic acid/sulfuric acid/water. A key product of the analysis was generation of the LMVE of nucleation rate from the first and second-order moments entering the covariance matrix of the data set. This was shown to provide an especially compact measurement technology. Additional work is needed to account for the influence of background aerosol on nucleation rate and to sort out postnucleation processes including particle growth, by condensation and coagulation, and scavenging of nucleated particles and vapor during the growth stage before the particles reach detectable size.

**ACKNOWLEDGMENTS**

This research was supported by the DOE Atmospheric Sciences Program. The authors thank Professor Ari Laaksonen and Dr. Stephen E. Schwartz for their insightful comments. This manuscript has been authorized by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH1-886 with the U. S. Department of Energy.

---