JOURNAL OF SPACECRAFT AND ROCKETS
Vol. 34, No. 4, July–August 1997

Air-Quality Monitoring and Detection of Air Contamination in an Enclosed Environment

Mikhail Skliar*
University of Utah, Salt Lake City, Utah 84112
and
W. Fred Ramirez†
University of Colorado, Boulder, Colorado 80309

We report on the development of an air-quality monitoring and early detection system for an enclosed environment with specific emphasis on manned spacecraft. The proposed monitoring approach is based on a distributed parameter model of contaminant dispersion and real-time contaminant concentration measurements. Kalman filtering is identified as a suitable method for generating on-line estimation of the spatial contamination profile, and an implicit Kalman filtering algorithm is shown to be preferable for real-time implementation. The identification of the contaminant concentration profile allows for a straightforward solution of the early detection of an air contamination event and provides information that enables potential automatic diagnosis of an unknown contamination source.

Nomenclature

$\boldsymbol{A}$	= discrete analog of $\mathcal{L}_{\boldsymbol{x}}$
$c(\boldsymbol{x}, t)$	= noise intensity
D_M	= molecular diffusivity
D_T	= eddy diffusivity
E	= expectation operator
F	= source function
F_i^{si}	= capacity of the ith point sink
F_i^{so}	= capacity of the ith point source
H	= maximum value of z
$h(q, t)$	= measurement operator
$\boldsymbol{I}$	= unit matrix
$\mathcal{L}_{\boldsymbol{x}}$	= spatial operator, $\mathcal{L}_x + \mathcal{L}_y + \mathcal{L}_z$
$\mathcal{L}_x, \mathcal{L}_y, \mathcal{L}_y$	= spatial operators acting along x, y, z
l	= number of sensors
N_{si}	= number of point sinks
N_{so}	= number of point sources
$Q(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$	= covariance function of the model noise
q	= instantaneous contaminant concentration
$q_{2\text{D}}$	= two-dimensional approximation of q
$\boldsymbol{q}$	= finite difference or finite element approximation of q, $[q_1(t), \ldots, q_n(t)]$
$R(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$	= covariance function of the measurement noise
$\boldsymbol{U}$	= air velocity vector
u, v, w	= Cartesian velocity components in x, y, z direction
$v(\boldsymbol{x}, t), w(\boldsymbol{x}, t)$	= independent zero-mean white Gaussian processes
x, y, z	= Cartesian coordinates
$\boldsymbol{x}$	= vector of Cartesian coordinates x, y, z
$\boldsymbol{x}_i^{\text{si}}$	= spatial location of the ith point sink
$\boldsymbol{x}_i^{\text{so}}$	= spatial location of the ith point source
$\boldsymbol{x}_j$	= spatial location of the jth sensor
δ	= Dirac delta function
Ω	= spatial domain of the problem
∇	= vector differential operator

Subscript

m	= time index

Superscripts

$'$	= instantaneous fluctuation
$\bar{}$	= time-smoothed value

Received Aug. 19, 1996; revision received April 21, 1997; accepted for publication April 23, 1997.

*Assistant Professor, Department of Chemical and Fuels Engineering. Member AIAA.
†Professor, Department of Chemical Engineering.

Introduction

SAFE air is the most vital environmental resource. Its quality is especially critical in an enclosed environment such as a spacecraft, a submarine, and an airliner cabin. The importance of the closely related issue of indoor (workplace) air quality has been repeatedly emphasized in the past 20 years.[1,2] It is known that only a slight degree of air contamination can have an acute health and performance impact on humans.[3]

In addition to background contamination resulting from material off-gassing, microbial growth, and everyday human activities,[4] there is a risk of an accidental contamination release. In the case of spacecraft and other enclosed environments, an unexpected event can lead to mission failure or even incapacitation and death because there may be a significant time lag before personnel evacuation becomes possible. An early detection of the contamination event can provide the critical time necessary to take mitigating actions or to arrange for safe evacuation.

Recent history of space exploration presents some examples of possible sources of accidental air contamination.[5] One such source is thermodegradation of electronic components in flight hardware.[6,7] It was reported that, during Space Shuttle flight STS-28, 0.1 g of polytetrafluoroethylene (PTFE, or Teflon® by its trademark), used as a wiring insulation, was pyrolyzed within 1.5 s. The crew experienced no adverse health effects. However, it was subsequently estimated that a pyrolysis of 2 g of PTFE would have affected the health and performance of the crew.

Another example of accidental contamination is the fire extinguishant (CO_2) release due to a containment failure and microbial growth resulting in the release of volatile pollutants into the air.[8] Based on the experience with the Space Shuttle, it is estimated that minor contamination events will occur with a frequency of once in 30 days; moderate accidents (those causing symptoms in the crew) will occur once per year. Details are not available on the accidental releases into Russian spacecraft. However, it is known that air contamination problems have nearly ended a mission[9] and that formaldehyde exposures are a concern.[10]

Currently utilized spacecraft air revitalization systems are designed to maintain equilibrium concentrations of contaminants within limits prescribed by spacecraft maximum allowable concentrations.[11,12] The monitoring of the Space Shuttle air is done off line and involves sampling of the spacecraft air immediately before

launch and late in the mission,[5] with additional samples taken if the crew suspects a contamination problem. The samples are analyzed after the spacecraft landing. The analysis of taken samples shows that air quality is consistently within the target region. At the same time, this monitoring technique failed to reflect some known contamination problems (such as the aforementioned PTFE thermodegradation) or provided inconclusive results. This technique is also insensitive to the temporal and spatial variations in contaminant concentrations, even though it is known that temporal variation of some compounds, such as carbon dioxide, can be quite large. The spatial variation of the contaminant concentration to our knowledge has not been previously estimated, but one can predict that it can be significant due to nonuniformity of the airflow within the habitat. This assumption is supported by information on large spatial variations of the contaminant concentrations throughout the Mir Station under normal operating conditions.

NASA's plans for extended human presence in a space station and bases on the moon and Mars in the 21st century[13] raise serious questions about the adequacy of current air revitalization and monitoring systems for long-duration missions. Added to the risk of accidental contaminations, there is a concern[14] that a space station would become an example of runaway tight (or sick) building syndrome[15,16] because there will not be the luxury of frequent returns to Earth for thorough cleaning characteristic to the Space Shuttle missions. It becomes apparent that new-generation air revitalization systems should support such advanced functions as real-time air-quality monitoring, health-risk evaluation, an early detection of a contamination event, which in turn allows for preventive mitigation of a health hazard, and isolation of the contamination source to facilitate repair and cleanup and to predict future health risk by extrapolating the identified trends. Essential components of such a system are gas sensors, high-performance onboard computers, and algorithms and methods that provide required system functionality. The analysis of existing gas sensors technology[17,18] indicates that the instrumentation needed for on-line multicomponent gas analysis is already available. In fact, a prototype combustion-products analyzer has been flown as part of the NASA program to develop the capability to detect combustion products.[4] More advanced instrumentation can be expected to become available in the near future, including a real-time particle detector suitable for space flight.[19] At the same time, the computational capabilities of available computers are constantly increasing.

This research has been motivated by a need for system development and integration of a new generation air-quality control system. The functional structure of such a system is depicted in Fig. 1. It consists of the on-line concentration measurement system of preselected contaminants, a data filter, which utilizes the measurement information and prior knowledge about the contaminant transport process to generate the real-time estimates of the contaminant concentration evolution in space and time, and the detection system, which uses the measurements and the estimations of the contaminant concentration to carry out an early detection of the air contamination event. The contaminant concentration estimates are also the basis for evaluating the crew contaminant intake. Coupled with toxicological dose-response information, concentration estimates allow for classification of a current health risk into the health hazard classes and for prediction of a long-term effect of the contaminated breathing air.

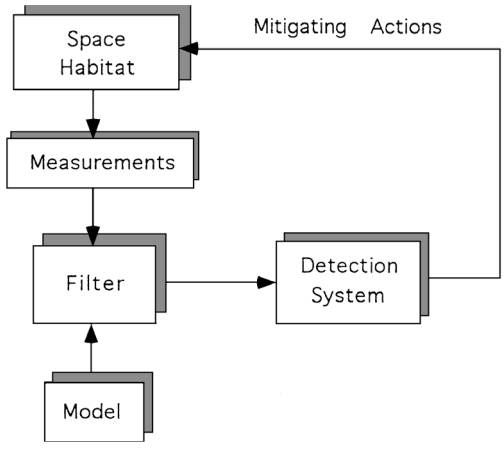


Fig. 1 Integrated air-quality monitoring system.

In the next section, we discuss different approaches to the modeling of air contamination in an enclosed environment and suggest a simple distributed parameter air contaminant dispersion model, applicable to both laminar and turbulent transport. Implementation aspects are also briefly considered. We then utilize the well-established Kalman filtering approach to generate on-line contaminant concentration estimates and propose an implicit Kalman filtering (IKF) algorithm suitable for on-line implementation. Finally, we show that real-time concentration estimates provide necessary information for early detection and isolation of the contamination events.

Model of Air Contamination Transport

Lumped and Distributed Models

A model of the spread, introduction, and removal of airborne contaminants creates the basis for monitoring and environmental control of a space habitat. Generally, there are two different approaches to simulation. In the first one, each separate volume of a habitat, such as a cabin or a room, is represented as a well-mixed tank with uniform distribution of all process variables. Mathematical models of this approach are a system of ordinary differential equations, resulting from the application of the macroscopic mass balances.[20] The result of simulation is the predicted time change of volume-averaged concentration of gaseous and airborne particulate contaminants, which depends on the rate of contaminant generation and removal. These models are relatively easy to develop and apply but are characterized by low resolution and accuracy.

An alternative approach is adopted in this paper. We base the computer model on the differential conservation laws of continuum. The resulting mathematical description of the contaminant transport process in the form of partial differential equations with appropriate initial and boundary conditions gives an exhaustive prediction of the distribution of contaminants throughout the habitat. Detailed information on the contaminant distribution can be used to assess the effects of chronic inhalation of low-concentration contaminants. A distributed parameter model is also required to precisely identify an unknown source of contamination.[21,22]

Three-Dimensional Model of Contaminant Dispersion

According to Fick's law of molecular diffusion,[23] binary mass transport process with isotropic diffusion can be described by the following convection-diffusion equation:

$$\frac{\partial q}{\partial t} + (\nabla \cdot q\boldsymbol{U}) = (\nabla \cdot D_M \nabla q) + F, \qquad \boldsymbol{U} = (u, v, w) \quad (1)$$

with appropriate initial and boundary conditions.

If the contaminant fraction in the air is low, Eq. (1) can be used to describe multicomponent transport. In this case, the spread of each contamination component is governed by a model in the form of Eq. (1). Transport models for different contaminants can be coupled through a generation term F, and molecular diffusivity D_{Mi} in general depends on the particular type of the contaminant species.

Assuming constant density, Eq. (1) in rectangular coordinates yields

$$\frac{\partial q}{\partial t} + \frac{\partial uq}{\partial x} + \frac{\partial vq}{\partial y} + \frac{\partial wq}{\partial z} = (\nabla \cdot D_M \nabla q) + F \quad (2)$$

Appropriate boundary conditions for Eq. (2) correspond to an impermeable wall, air duct, open hatch, and a completely or partially permeable wall (Fig. 2). For all practical purposes, the source function F can be adequately described by the combination of pointwise contaminant sources and sinks:

$$F = \sum_{i=1}^{N_{so}} F_i^{so}(t)\delta\left(\boldsymbol{x} - \boldsymbol{x}_i^{so}\right) - \sum_{i=1}^{N_{si}} F_i^{si}(t)\delta\left(\boldsymbol{x} - \boldsymbol{x}_i^{si}\right) \quad (3)$$

Equation (2) is directly applicable to the laminar transport of an airborne contaminant. Though Ref. 24 suggests that space station airflows will be mainly laminar, turbulent transport can play

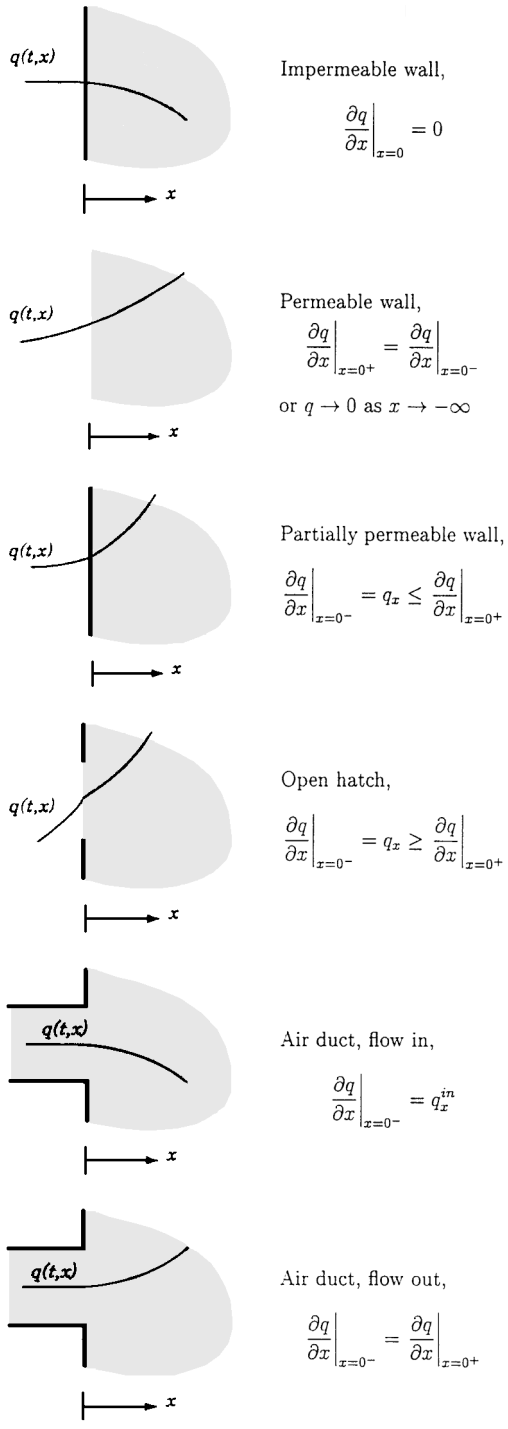


Fig. 2 Boundary conditions of the contaminant transport process.

a significant role at least in some parts of the spacecraft habitat. Thus, it is important to modify Eq. (2) to accommodate the case of turbulent transport.

For the case of turbulent flow, both the flow velocity $\boldsymbol{U} = (u, v, w)$ and the concentration q must be treated as stochastic quantities. Following the framework of semi-empirical theory of turbulent diffusion,[23] we introduce the following notation:

$$\boldsymbol{U} = \bar{\boldsymbol{U}} + \boldsymbol{U}' = (\bar{u}, \bar{v}, \bar{w}) + (u', v', w'), \qquad q = \bar{q} + q' \quad (4)$$

Because we are interested in the meaningful trend of contaminant concentration, rather than its stochastic fluctuations, we average the Fickian model (2) over a time interval ΔT long enough for the integral of instantaneous fluctuations to vanish. This yields

$$\frac{\partial \bar{q}}{\partial t} + \nabla \cdot (\overline{q'\boldsymbol{U}'}) + \nabla \cdot (\bar{q}\bar{\boldsymbol{U}}) = (\nabla \cdot D_M \nabla \bar{q}) + \bar{F} \quad (5)$$

Let

$$\overline{q'\boldsymbol{U}'} = -D_T \nabla \bar{q} \quad (6)$$

Because

$$D_M \ll D_T \quad (7)$$

molecular diffusion can often be ignored, and the resulting three-dimensional turbulent transport model takes the following form:

$$\frac{\partial \bar{q}}{\partial t} + \nabla \cdot (\bar{q}\bar{\boldsymbol{U}}) = (\nabla \cdot D_T \nabla \bar{q}) + \bar{F} \quad (8)$$

The models for laminar transport, Eq. (1), and turbulent transport, Eq. (8), are of the same mathematical form. This allows us to use the same computer model for laminar, turbulent, and mixed contaminant transport cases.

Two-Dimensional Approximation

In an attempt to develop a model that gives sufficient details about the contaminant concentration profile and at the same time is simple enough to be run on line, a two-dimensional approximation of the transport equation (2) was previously suggested.[6,7] The transport model can be simplified by averaging Eq. (2) over the least important spatial coordinate. Assuming the averaging over z, define the height-averaged concentration:

$$q_{2D} = \frac{1}{H}\int_0^H q \, dz \quad (9)$$

We further assume that D_M does not change with z and that (u, v) is a z-averaged air velocity vector. Integrating Eq. (2) with respect to z and accounting for boundary condition $w(0) = w(H) = 0$, we obtain the following two-dimensional approximation of the laminar transport model:

$$\frac{\partial q_{2D}}{\partial t} + \frac{\partial u q_{2D}}{\partial x} + \frac{\partial v q_{2D}}{\partial y} = \frac{\partial}{\partial x} D_M \frac{\partial q_{2D}}{\partial x} + \frac{\partial}{\partial y} D_M \frac{\partial q_{2D}}{\partial y} + f \quad (10)$$

where

$$f = \int_0^H F \, dz + D_M \left[\left.\frac{\partial q}{\partial z}\right|_H - \left.\frac{\partial q}{\partial z}\right|_0 \right] \quad (11)$$

Similarly, a turbulent dispersion is approximated by the following two-dimensional equation:

$$\frac{\partial \bar{q}_{2D}}{\partial t} + \frac{\partial \bar{u}\bar{q}_{2D}}{\partial x} + \frac{\partial \bar{v}\bar{q}_{2D}}{\partial y} = \frac{\partial}{\partial x} D_T \frac{\partial \bar{q}_{2D}}{\partial x} + \frac{\partial}{\partial y} D_T \frac{\partial \bar{q}_{2D}}{\partial y} + \bar{f} \quad (12)$$

where $\bar{q}_{2D}$ is height averages and time-smoothed contaminant concentration.

This approach provides the flexibility of using different two-dimensional approximations, resulting from the averaging along different spatial coordinates, depending on current needs, or of simultaneously running more than one approximation to obtain a more detailed prediction of the spread of the contaminant.

Computer Implementation

Two-Dimensional Model

The transport equation with appropriate boundary conditions is used to develop a computer model of the transport process. First, using the finite differences or finite elements method, we find an approximation of partial derivatives with respect to spatial coordinates. Let matrix $\boldsymbol{A}$ be a discrete analog of the spatial operator

$$\mathcal{L}_x(\cdot) = \frac{\partial}{\partial x} D_M \frac{\partial}{\partial x} + \frac{\partial}{\partial y} D_M \frac{\partial}{\partial y} - \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \quad (13)$$

obtained using finite differences, where the dimension n of the sparse matrix $\boldsymbol{A}$ can be very large. In the vector form, the semidiscrete analog of Eq. (10) can be written as

$$\frac{d\boldsymbol{q}}{dt} = \boldsymbol{A}\boldsymbol{q} + \boldsymbol{f} \quad (14)$$

where $\boldsymbol{f}(t)$ is an approximation of the source function $f(x, y, t)$ plus the contribution from the boundary conditions of Eq. (10). Temporal discretization of Eq. (14) concludes the development of the discrete analog of the transport model. To ensure numerical stability of the computer implementation, it is advisable to use an implicit approximation of the time derivative in Eq. (14). For instance, the Crank–Nicolson scheme results in the following discrete implicit model:

$$\boldsymbol{A}_1\boldsymbol{q}_{m+1} = \boldsymbol{A}_2\boldsymbol{q}_m + \boldsymbol{f}_m \tag{15}$$

where $\boldsymbol{f}_m = \boldsymbol{f}(m\Delta t)$, $\boldsymbol{q}_m$ approximates $\boldsymbol{q}(m\Delta t)$, and

$$\boldsymbol{A}_1 = (\boldsymbol{I}/\Delta t) - \tfrac{1}{2}\boldsymbol{A}, \qquad \boldsymbol{A}_2 = (\boldsymbol{I}/\Delta t) + \tfrac{1}{2}\boldsymbol{A}$$

The sparsity of the system (15) can be exploited to yield an efficient implementation of the computer model.

The finite elements method leads to similar results in that the resulted discrete model is in the form of Eq. (15).

Three-Dimensional Model

The described two-dimensional design can be used to discretize the partial differential equation (1) or (8) to obtain a straightforward computer implementation of the three-dimensional contaminant transport model. However, the dimension n of the resulting discrete model will increase substantially, and the structure of the sparse matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ will become more complicated. Consequently, the direct approximation of a three-dimensional transport model can be quite computationally intensive.

A significant reduction in computational complexity can be achieved following the paradigm of the alternating direction implicit approach, which embodies the idea of operator or time splitting. Such splitting reduces the original problem into a series of simplified subproblems that must be solved consecutively on each time step. For example, application of the Douglas method[25] to Eq. (2) leads to the following numerical scheme:

$$\begin{aligned}
[-\boldsymbol{A}_x - (2/\Delta t)]\boldsymbol{q}^* &= [\boldsymbol{A}_x + 2\boldsymbol{A}_y + 2\boldsymbol{A}_z - (2/\Delta t)]\boldsymbol{q}_m + 2\boldsymbol{f}_m \\
[-\boldsymbol{A}_y - (2/\Delta t)]\boldsymbol{q}^{**} &= -\boldsymbol{A}_y\boldsymbol{q}_m - (2/\Delta t)\boldsymbol{q}^* \\
[-\boldsymbol{A}_z - (2/\Delta t)]\boldsymbol{q}_{m+1} &= -\boldsymbol{A}_z\boldsymbol{q}_m - (2/\Delta t)\boldsymbol{q}^{**}
\end{aligned} \tag{16}$$

where $\boldsymbol{A}_x$ approximates $\mathcal{L}_x(\cdot) = (\partial/\partial x)D_M(\partial/\partial x) - (\partial u/\partial x)$, $\boldsymbol{A}_y$ and $\boldsymbol{A}_z$ approximate $\mathcal{L}_y$ and $\mathcal{L}_z$, respectively, and $\boldsymbol{q}^*$ and $\boldsymbol{q}^{**}$ are intermediate variables. Thus, instead of solving one large and complex system (15), we need to solve three tridiagonal equations (16).

Air-Quality Monitoring Using Kalman Filtering

The model of the introduction, dispersion, and removal of airborne contaminants in an enclosed environment summarizes our knowledge about the transport process and is especially valuable during the design stage because it allows us to analyze the influence of different factors on the spread of contaminants and to simulate performance and limitations of air revitalization system. At the same time, the model is merely a reflection of our current a priori knowledge about the physics of the air contaminants transport process, geometry of the habitat, and emission sources and removal devices. As a result, the quality of simulation depends solely on the assumptions employed in the development of the model and the completeness and accuracy of the input data. In reality, the model is often in question, and the information about some system parameters (such as eddy diffusivity) is inherently uncertain and incomplete.

The most important information about air quality in a habitat is the concentration distribution of the airborne contaminant preselected for monitoring. The task of generating real (or near real) time contaminant concentration estimates can only be accomplished using on-line concentration measurements. However, one should be aware of the limitations of the measurement data alone. Only a limited number of sensors can be deployed. Each sensor usually produces spatially localized, i.e., pointwise, measurements, which must be extrapolated over the entire spatial volume. The measurements from different sensors may not be in complete agreement, requiring data reconciliation. And, finally, the effect of the measurement noise must be taken into account.

All of these factors motivate the development of the air-quality monitoring system based on both on-line measurements and a verified model of the contaminant transport process. The integration of measurement data and the model can be achieved within the framework of the well-established Kalman filtering approach. According to the Kalman filtering paradigm, model and measurement uncertainties are represented by the additive stochastic white Gaussian perturbations. First, let us consider the modification of the transport model. The modified model of the contaminant transport has the following form:

$$\begin{aligned}
&\frac{\partial q}{\partial t} + \frac{\partial uq}{\partial x} + \frac{\partial vq}{\partial y} + \frac{\partial wq}{\partial z} \\
&\quad = (\nabla \cdot D_M \nabla q) + F + c(\boldsymbol{x}, t)w(\boldsymbol{x}, t) \\
&\qquad q(\boldsymbol{x}, 0) = q^0(\boldsymbol{x}), \qquad E[q^0(\boldsymbol{x})] = 0 \\
&\quad E\left[q^0(\boldsymbol{x}_1)q^0(\boldsymbol{x}_2)\right] = p_0(\boldsymbol{x}_1, \boldsymbol{x}_2), \qquad \boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega
\end{aligned} \tag{17}$$

where

$$E[w(\boldsymbol{x}_1, t)w(\boldsymbol{x}_2, \tau)] = Q(\boldsymbol{x}_1, \boldsymbol{x}_2, t)\delta(t - \tau) \tag{18}$$

The process model is supplemented with a model of the measurement system:

$$z(\boldsymbol{x}, t) = h(q, t) + v(\boldsymbol{x}, t), \qquad \boldsymbol{x} \in \Omega \tag{19}$$

where $z(\boldsymbol{x}, t)$ is the output of the measurement system and

$$E[v(\boldsymbol{x}_1, t)v(\boldsymbol{x}_2, \tau)] = R(\boldsymbol{x}_1, \boldsymbol{x}_2, t)\delta(t - \tau), \qquad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega \tag{20}$$

Sensors usually provide only pointwise readings. The model of the measurement system with pointwise sensors takes the following form:

$$z(\boldsymbol{x}_j, t) = h[q(\boldsymbol{x}_j, t), t] + v(\boldsymbol{x}_j, t), \qquad j = 1, \ldots, l \tag{21}$$

The levels of the model and the measurement uncertainties are determined by the covariance functions Q and R. The choice of the measurement noise covariance matrix R is made based on a careful study of the sensing instrumentation under controlled conditions. The covariance of the model noise is chosen at the model validation stage. After initial determination, Q is often viewed as a design parameter and is adjusted to obtain the desired estimates.

The discretization of the stochastic model, Eq. (17), results in either a single implicit equation, such as Eq. (15), or the system of implicit equations, such as the system of Eqs. (16). The case of Eq. (15) is considered first. After discretization of Eqs. (17) and (19), we obtain

$$\boldsymbol{A}_1\boldsymbol{q}_{m+1} = \boldsymbol{A}_2\boldsymbol{q}_m + \boldsymbol{f}_m + \boldsymbol{C}(m)\boldsymbol{w}_m \tag{22}$$

$$\boldsymbol{z}_{m+1} = \boldsymbol{H}(m+1)\boldsymbol{q}_{m+1} + \boldsymbol{v}_{m+1} \tag{23}$$

where $\boldsymbol{C}(m)$ and $\boldsymbol{H}(m+1)\boldsymbol{q}_{m+1}$ are discrete approximations of $c(\boldsymbol{x}, m\Delta t)$ and $h[q, (m+1)\Delta t]$ and $\boldsymbol{z}_{m+1}$ is a measurement vector corresponding to $z(\boldsymbol{x}, t)$; $\boldsymbol{w}_m$ and $\boldsymbol{v}_{m+1}$ are uncorrelated zero-mean white Gaussian sequences such that

$$E\left[\boldsymbol{w}_j\boldsymbol{w}_m^T\right] = \boldsymbol{Q}(m)\delta(j - m) \tag{24}$$

$$E\left[\boldsymbol{v}_{j+1}\boldsymbol{v}_{m+1}^T\right] = \boldsymbol{R}(m+1)\delta(j - m) \tag{25}$$

where $\boldsymbol{Q}(m)$ and $\boldsymbol{R}(m+1)$ approximate functions $Q(\boldsymbol{x}_1, \boldsymbol{x}_2, m\Delta t)$ and $R[\boldsymbol{x}_1, \boldsymbol{x}_2, (m+1)\Delta t]$ and $\delta(j - m) = 1$ if $j = m$ and is zero otherwise.

Traditional Kalman filtering requires the model of the process to be specified in an explicit form. If matrix $\boldsymbol{A}_1$ is nonsingular for all m, the implicit model, Eq. (22), can be rewritten in an equivalent explicit form, and the traditional Kalman filter can be applied to generate the optimal estimates of $\boldsymbol{q}_{m+1}$. However, there is a strong

motivation to avoid the matrix inversion step in the filtering algorithm. If, for some m, matrix $\boldsymbol{A}_1$ is ill-conditioned, its inverse is calculated with significant error, unless special measures are built into the filtering algorithm. Furthermore, the system matrix $\boldsymbol{A}_1^{-1}\boldsymbol{A}_2$ of the equivalent explicit representation must be treated as a general (full) matrix, because matrix inversion destroys matrix sparsity. Therefore, the traditional Kalman filtering is an inefficient algorithm when the system is sparse and implicit. It can also lead to significant computational errors should the matrix $\boldsymbol{A}_1$ be ill-conditioned.

We have previously proposed an IKF method,[26] which does not require matrix inversion. When applied to a sparse implicit system, IKF is an order of magnitude faster than the traditional Kalman filter, making it a superb method for on-line estimation of the contaminant concentration. According to the IKF approch, the optimal estimation $\hat{\boldsymbol{q}}_{m+1|m+1}$ of the contaminant concentration is given by the following implicit Kalman filter equation (notation $\boldsymbol{a}_{b|c}$ means that the estimation of $\boldsymbol{a}$ at time b is based on the measurements up to time c, $c \le b$):

$$\boldsymbol{A}_1\hat{\boldsymbol{q}}_{m+1|m+1} = \hat{\boldsymbol{y}}_{m+1|m} + \boldsymbol{L}_y(m+1)[\boldsymbol{z}_{m+1} - \boldsymbol{H}_1(m+1)\hat{\boldsymbol{y}}_{m+1|m}] \tag{26}$$

for $m = 0, 1, \ldots,$ where

$$\hat{\boldsymbol{y}}_{m+1|m} = \boldsymbol{A}_2\hat{\boldsymbol{q}}_{m|m} + \boldsymbol{f}_m \tag{27}$$

with $\hat{\boldsymbol{q}}_{0|0} = \hat{\boldsymbol{q}}^0$. The modified measurement matrix $\boldsymbol{H}_1(m+1)$ is determined from the linear equation

$$\boldsymbol{H}_1\boldsymbol{A}_1 = \boldsymbol{H} \tag{28}$$

The implicit Kalman filter gain $\boldsymbol{L}_y(m+1)$ is equal to

$$\begin{aligned}\boldsymbol{L}_y(m+1) &= \boldsymbol{P}^y_{m+1|m+1}\boldsymbol{H}_1^T(m+1)\boldsymbol{R}^{-1}(m+1)\\ &= \boldsymbol{P}^y_{m+1|m}\boldsymbol{H}_1^T(m+1)\Big[\boldsymbol{H}_1(m+1)\\ &\quad\times\boldsymbol{P}^y_{m+1|m}\boldsymbol{H}_1^T(m+1) + \boldsymbol{R}(m+1)\Big]^{-1}\end{aligned} \tag{29}$$

where the predicted estimation error covariance matrix $\boldsymbol{P}^y_{m+1|m}$ of an auxiliary variable $\boldsymbol{y}_{m+1} = \boldsymbol{A}_1\boldsymbol{q}_{m+1}$ is found as a result of the time propagation of the estimation error covariance $\boldsymbol{P}^q_{m|m}$ of the concentration $\boldsymbol{q}_m$ according to the following equation:

$$\boldsymbol{P}^y_{m+1|m} = \boldsymbol{A}_2\boldsymbol{P}^q_{m|m}\boldsymbol{A}_2^T + \boldsymbol{C}(m)\boldsymbol{Q}(m)\boldsymbol{C}^T(m) \tag{30}$$

The estimation error covariance matrix $\boldsymbol{P}^y_{m+1|m+1}$ satisfies the covariance measurement update equation

$$\begin{aligned}\boldsymbol{P}^y_{m+1|m+1} &= \boldsymbol{P}^y_{m+1|m} - \boldsymbol{P}^y_{m+1|m}\boldsymbol{H}_1^T(m+1)\Big[\boldsymbol{H}_1(m+1)\boldsymbol{P}^y_{m+1|m}\\ &\quad\times\boldsymbol{H}_1^T(m+1) + \boldsymbol{R}^{-1}(m+1)\Big]^{-1}\boldsymbol{H}_1(m+1)\boldsymbol{P}^y_{m+1|m}\end{aligned} \tag{31}$$

or, equivalently,

$$\boldsymbol{P}^y_{m+1|m+1} = [\boldsymbol{I} - \boldsymbol{L}_y(m+1)\boldsymbol{H}_1(m+1)]\boldsymbol{P}^y_{m+1|m} \tag{32}$$

The estimation error covariance matrix $\boldsymbol{P}^q_{m+1|m+1}$ is related to the covariance matrix $\boldsymbol{P}^y_{m+1|m+1}$ by the following linear matrix equation:

$$\boldsymbol{P}^y_{m+1|m+1} = \boldsymbol{A}_1\boldsymbol{P}^q_{m+1|m+1}\boldsymbol{A}_1^T \tag{33}$$

The IKF generates the minimal variance estimations of $\boldsymbol{q}_{m+1}$ and is theoretically equivalent to traditional Kalman filtering, provided the inverse of matrix $\boldsymbol{A}_1$ exists for all k. However, it provides a basis for the following new state estimation algorithm of the implicit system, Eq. (22), which does not require matrix inversion. Given $\boldsymbol{z}_{m+1}$, $\hat{\boldsymbol{q}}_{m|m}$, and $\boldsymbol{L}_y(m+1)$, follow these steps.

1) Compute $\hat{\boldsymbol{y}}_{m+1|m}$ by propagating $\hat{\boldsymbol{q}}_{m|m}$ according to Eq. (27).

2) Solve the linear matrix equation (28) for the modified measurement matrix $\boldsymbol{H}_1(m+1)$.

3) Solve the linear equation

$$\boldsymbol{A}_1\hat{\boldsymbol{q}}_{m+1|m+1} = \hat{\boldsymbol{y}}_{m+1|m} + \boldsymbol{L}_y(k+1)[\boldsymbol{z}_{m+1} - \boldsymbol{H}_1(m+1)\hat{\boldsymbol{y}}_{m+1|m}] \tag{34}$$

for the optimal estimate $\hat{\boldsymbol{q}}_{m+1|m+1}$.

Note that, if matrices $\boldsymbol{A}_1$ and $\boldsymbol{H}$ are time invariant, matrix $\boldsymbol{H}_1$ needs to be calculated only once.

The Kalman gain $\boldsymbol{L}_y(m+1)$ can be calculated in the following ways. Given $\boldsymbol{P}^q_{m|m}$, follow these steps.

1) Compute $\boldsymbol{P}^y_{m+1|m}$ by propagating $\boldsymbol{P}^q_{m|m}$ according to Eq. (30).

2) Compute the implicit Kalman filter gain using Eq. (29).

3) Calculate $\boldsymbol{P}^y_{m+1|m+1}$ according to Eq. (32).

To initiate the gain calculation on the next time step, we need to find $\boldsymbol{P}^q_{k+1|k+1}$ from the linear Eq. (33), using direct or iterative methods.[26]

Turning our attention to a discrete analog of the stochastic transport model in the form of the system of implicit equations, we immediately observe that, after appropriate modifications, Eqs. (16) can be written as a single implicit equation:

$$\boldsymbol{A}_1\boldsymbol{Q}_{m+1} = \boldsymbol{A}_2\boldsymbol{Q}_m + \begin{bmatrix}2\boldsymbol{f}_m\\ 0\\ 0\end{bmatrix} + \begin{bmatrix}2\boldsymbol{C}(m)\\ 0\\ 0\end{bmatrix} w_m \tag{35}$$

where

$$\boldsymbol{Q}_m = \begin{bmatrix}\boldsymbol{q}^*\\ \boldsymbol{q}^{**}\\ \boldsymbol{q}_m\end{bmatrix}$$

$$\boldsymbol{A}_1 = \left\{\boldsymbol{A}_1^{ij}\right\} = \begin{bmatrix}-\boldsymbol{A}_x - (2/\Delta t) & 0 & 0\\ (2/\Delta t) & -\boldsymbol{A}_y - (2/\Delta t) & 0\\ 0 & (2/\Delta t) & -\boldsymbol{A}_z - (2/\Delta t)\end{bmatrix}$$

$$\boldsymbol{A}_2 = \left\{\boldsymbol{A}_2^{ij}\right\} = \begin{bmatrix}0 & 0 & \boldsymbol{A}_x + 2\boldsymbol{A}_y + 2\boldsymbol{A}_z - (2/\Delta t)\\ 0 & 0 & -\boldsymbol{A}_y\\ 0 & 0 & -\boldsymbol{A}_z\end{bmatrix}$$

and

$$\boldsymbol{z}_{m+1} = [0 \quad 0 \quad \boldsymbol{H}(m+1)]\boldsymbol{Q}_{m+1} + \boldsymbol{v}_{m+1} \tag{36}$$

The system of Eqs. (35) and (36) is in the same form as Eqs. (22) and (23), and the implicit Kalman filter is directly applicable. The resulting algorithm can be further simplified if a special structure of Eq. (35) is taken into account. After transformations, the estimation of the contaminant concentration is determined from the sequential solution of the following tridiagonal equations:

$$\begin{aligned}&[-\boldsymbol{A}_x - (2/\Delta t)]\boldsymbol{q}^* = [\boldsymbol{A}_x + 2\boldsymbol{A}_y + 2\boldsymbol{A}_z - (2/\Delta t)]\hat{\boldsymbol{q}}_{m|m}\\ &\quad + 2\boldsymbol{f}_m + \boldsymbol{L}_1[\boldsymbol{z}_{m+1} - \boldsymbol{H}_1\hat{\boldsymbol{y}}_{m+1|m}]\\ &[-\boldsymbol{A}_y - (2/\Delta t)]\boldsymbol{q}^{**} = -\boldsymbol{A}_z\hat{\boldsymbol{q}}_{m|m} - (2/\Delta t)\boldsymbol{q}^{**}\\ &\quad + \boldsymbol{L}_2[\boldsymbol{z}_{m+1} - \boldsymbol{H}_1\hat{\boldsymbol{y}}_{m+1|m}]\\ &[-\boldsymbol{A}_z - (2/\Delta t)]\hat{\boldsymbol{q}}_{m+1|m+1} = -\boldsymbol{A}_z\hat{\boldsymbol{q}}_{m|m} - (2/\Delta t)\boldsymbol{q}^{**}\\ &\quad + \boldsymbol{L}_3[\boldsymbol{z}_{m+1} - \boldsymbol{H}_1\hat{\boldsymbol{y}}_{m+1|m}]\end{aligned} \tag{37}$$

where the predicted estimation of the auxiliary variable $\boldsymbol{y}$ is

$$\hat{\boldsymbol{y}}_{m+1|m} = \begin{bmatrix}\boldsymbol{A}_2^{13}\\ \boldsymbol{A}_2^{23}\\ \boldsymbol{A}_2^{33}\end{bmatrix}\hat{\boldsymbol{q}}_{m|m} + \begin{bmatrix}2\boldsymbol{f}_m\\ 0\\ 0\end{bmatrix} \tag{38}$$

The modified measurement matrix $\boldsymbol{H}_1 = \{\boldsymbol{H}_{1j}\}$ is found from the following linear equation:

$$[\boldsymbol{H}_{11} \quad \boldsymbol{H}_{12} \quad \boldsymbol{H}_{13}]\boldsymbol{A}_1 = [0 \quad 0 \quad \boldsymbol{H}] \tag{39}$$

the solution of which reduces to the solution of the following three linear tridiagonal equations:

$$\boldsymbol{H}_{13}\boldsymbol{A}_1^{33} = \boldsymbol{H}, \qquad \boldsymbol{H}_{12}\boldsymbol{A}_1^{22} = -\boldsymbol{H}_{13}\boldsymbol{A}_1^{32}, \qquad \boldsymbol{H}_{11}\boldsymbol{A}_1^{11} = -\boldsymbol{H}_{12}\boldsymbol{A}_1^{21} \tag{40}$$

The implicit Kalman gain $\boldsymbol{L}_{m+1} = [\boldsymbol{L}_1^T \;\; \boldsymbol{L}_2^T \;\; \boldsymbol{L}_3^T]^T$ is equal to

$$\boldsymbol{L}_{m+1} = \boldsymbol{P}^y_{m+1\,|\,m}\boldsymbol{H}_1^T\left[\boldsymbol{H}_1\boldsymbol{P}^y_{m+1\,|\,m}\boldsymbol{H}_1^T + \boldsymbol{R}(m+1)\right]^{-1} \tag{41}$$

where

$$\boldsymbol{P}^y_{m+1\,|\,m} = \begin{bmatrix} \boldsymbol{A}_2^{13}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{13^T} + 4\boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}^T & \boldsymbol{A}_2^{13}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{23^T} & \boldsymbol{A}_2^{13}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{33^T} \\ \boldsymbol{A}_2^{23}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{13^T} & \boldsymbol{A}_2^{23}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{23^T} & \boldsymbol{A}_2^{23}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{33^T} \\ \boldsymbol{A}_2^{33}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{13^T} & \boldsymbol{A}_2^{33}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{23^T} & \boldsymbol{A}_2^{33}\boldsymbol{P}^q_{m\,|\,m}\boldsymbol{A}_2^{33^T} \end{bmatrix} \tag{42}$$

As before,

$$\boldsymbol{P}^y_{m+1\,|\,m+1} = [\boldsymbol{I} - \boldsymbol{L}_{m+1}\boldsymbol{H}_1]\boldsymbol{P}^y_{m+1\,|\,m} \tag{43}$$

and

$$\boldsymbol{P}^y_{m+1\,|\,m+1} = \boldsymbol{A}_1\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\boldsymbol{A}_1^T$$

where $\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1} = \{[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}]^{ij}\}$, $i, j = \{1, 2, 3\}$ is the covariance matrix of the estimation error of $\boldsymbol{Q}_{m+1}$. Taking advantage of the block-matrix structure of $\boldsymbol{A}_1$, the solution of the last equation is reduced to a sequential solution of the following six equations:

$$\boldsymbol{A}_1^{11}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{11}\boldsymbol{A}_1^{11^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{11}$$

$$\boldsymbol{A}_1^{11}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{12}\boldsymbol{A}_1^{22^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{12} - \boldsymbol{A}_1^{11}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{11}\boldsymbol{A}_1^{21^T}$$

$$\boldsymbol{A}_1^{11}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{13}\boldsymbol{A}_1^{33^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{13} - \boldsymbol{A}_1^{11}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{12}\boldsymbol{A}_1^{32^T}$$

$$\boldsymbol{A}_1^{22}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{22}\boldsymbol{A}_1^{22^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{22} - \boldsymbol{A}_1^{21}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{11}\boldsymbol{A}_1^{21^T}$$

$$-\boldsymbol{A}_1^{21}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{12}\boldsymbol{A}_1^{22^T} - \left\{\boldsymbol{A}_1^{21}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{12}\boldsymbol{A}_1^{22^T}\right\}^T \tag{44}$$

$$\boldsymbol{A}_1^{22}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{23}\boldsymbol{A}_1^{33^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{23} - \boldsymbol{A}_1^{21}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{12}\boldsymbol{A}_1^{32^T}$$

$$-\boldsymbol{A}_1^{22}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{22}\boldsymbol{A}_1^{32^T} - \boldsymbol{A}_1^{21}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{13}\boldsymbol{A}_1^{33^T}$$

$$\boldsymbol{A}_1^{33}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{33}\boldsymbol{A}_1^{33^T} = \left[\boldsymbol{P}^y_{m+1\,|\,m+1}\right]^{33} - \boldsymbol{A}_1^{32}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{22}\boldsymbol{A}_1^{32^T}$$

$$-\boldsymbol{A}_1^{32}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{23}\boldsymbol{A}_1^{33^T} - \left\{\boldsymbol{A}_1^{32}\left[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}\right]^{23}\boldsymbol{A}_1^{33^T}\right\}^T$$

where $[\boldsymbol{P}^{\mathcal{Q}}_{m+1\,|\,m+1}]^{33} = \boldsymbol{P}^q_{m+1\,|\,m+1}$.

We are now in the position to formulate an algorithm for the estimation of the contaminant concentration $\boldsymbol{q}_{m+1}$. Given $\boldsymbol{z}_{m+1}$, $\hat{\boldsymbol{q}}_{m\,|\,m}$, and $\boldsymbol{L}_{m+1}$, follow these steps.

1) Compute $\hat{\boldsymbol{y}}_{m+1\,|\,m}$ by propagating $\hat{\boldsymbol{q}}_{m\,|\,m}$ according to Eq. (38).

2) Successively solve three tridiagonal matrix equations (40) for the modified measurement matrix $\boldsymbol{H}_1$.

3) Successively solve three tridiagonal equations (37) for the optimal estimation $\hat{\boldsymbol{q}}_{m+1\,|\,m+1}$.

The calculation of the gain $\boldsymbol{L}_{m+1}$ of the implicit Kalman filter follows the following algorithm.

1) Calculate $\boldsymbol{P}^y_{m+1\,|\,m}$ according to Eq. (42).

2) Calculate the IKF gain using Eq. (41).

3) Calculate $\boldsymbol{P}^y_{m+1\,|\,m+1}$ according to Eq. (43).

4) To initiate the gain calculation algorithm on the next time step, sequentially solve matrix Eqs. (44) to find the error covariance matrix $\boldsymbol{P}^q_{m+1\,|\,m+1}$.

Despite the complex appearance, the developed implicit filtering algorithm is very effective for large-scale systems and is convenient for computer implementation. If the number of sensors $l \ll n$, then the computer implementation of this algorithm requires on the order of n^2 floating point operations. Furthermore, all linear systems that need to be solved in the course of implicit filtering have the same simple tridiagonal form and can be solved using the same solution engine.

It is well known that for real-time applications of the Kalman filtering it is advisable to use a square root filtering algorithm, which has a superior numerical stability. The details on the square root implementation of the IKF can be found in Ref. 27.

Example: Estimation of Contaminant Concentration Using an Implicit Kalman Filter

Consider a two-dimensional height-averaged approximation of the contaminant transport process in an enclosed habitat. Assuming laminar airflow, the parabolic input and output air velocities, and the additional parameters of Table 1, the z-averaged air velocity distribution in the cabin was calculated and is shown in Fig. 3. Here, the arrows follow the streamlines of the airflow, and their length is proportional to the magnitude of the air velocity at a particular spatial point. The maximum air velocity is equal to 1 m/s and occurs in the center of the air duct. Note that due to large velocity gradients we have assumed $D_T = 23\,\mathrm{cm}^2/\mathrm{s}$, which is the eddy diffusivity of carbon dioxide in air.

The height-averaged contaminant concentration is probed by 12 identical gas sensors. Spatial location of sensors is shown in Fig. 4. Sensor location is identified by the symbol $\otimes$. The interior point identified by $\oplus$ is referred to later. For reference purposes, each location is associated with the corresponding mesh node number.

Table 1 Input data used in simulation

Parameters	
Diffusivity	$D_T = 23\,\mathrm{cm}^2/\mathrm{s}$
Temperature	$T = 20^\circ\mathrm{C}$
Density	$\rho = 1.200\,\mathrm{g/l}$
Viscosity	$\mu = 1.834 \times 10^{-5}\,\mathrm{Pa}\cdot\mathrm{s}$
Kinematic viscosity	$\nu = 1.528 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$
Bulk velocity U	Fig. 3
Spatial domain	
Geometry	Fig. 4
Mesh	62 by 44
Discretization in x direction	$\Delta x = 0.2$ m
Discretization in y direction	$\Delta y = 0.2$ m
Time step	$\Delta t = 1$ s
Boundary conditions (b.c.)	
Impermeable wall	Fig. 2
Inlet duct	"flow in" b.c., $q_x^{\mathrm{in}} = u(y)q^{\mathrm{in}}$, where $q^{\mathrm{in}} = 0.04\,\mathrm{mg/m}^3$
Outlet duct	"flow out" b.c.
Noise	
Measurement noise, $\boldsymbol{v}_{j+1}$	Zero-mean white Gaussian with $E[\boldsymbol{v}_{j+1}\boldsymbol{v}^T_{m+1}] = \sigma_s^2\boldsymbol{I}\delta(j-m)$, $\sigma_s = 0.001$
Model noise, $\boldsymbol{w}_j$	Zero-mean white Gaussian with $E[\boldsymbol{w}_j\boldsymbol{w}_m^T] = \sigma_m^2\boldsymbol{I}\delta(j-m)$, $\sigma_m = 0.0001$

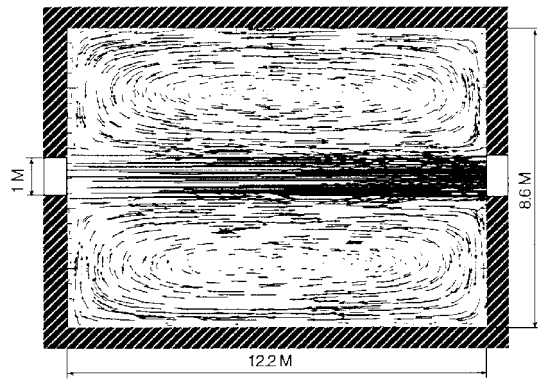


Fig. 3 Height-averaged air velocity distribution.

Assume that the standard deviation of measurement noise is approximately equal to 2.5% of the maximum value of the contaminant concentration. The value of the model noise is given in Table 1.

The following contamination scenario has been adopted. The initial contaminant concentration in the cabin is zero. At time $t = 0$, a contaminant is introduced into the cabin with the inlet airstream at a known rate (Table 1). The IKF is used to estimate contaminant concentration based on the noisy measurements and the uncertain model. Figure 5 depicts the evolution of filtered contaminant concentrations (dashed line) in the spatial point 50, where one of the sensors is located. The dotted line is used to represent the measured contaminant concentration. The solid line depicts the true value of contaminant concentration. Figure 6 shows the spatial distribution of the contaminant concentration at $t = 4$ min estimated by the IKF.

Detection of a Contamination Event

On-line information on the contaminant concentration distribution, generated by the monitoring system, is a basis for the solution of a wide range of air-quality control problems. Among them are 1) the early detection of a contamination event, such as an accidental release or a creeping leak; 2) the localization of the detected contamination source to facilitate the mitigating actions, such as repair and cleanup; 3) an identification of the dynamic characteristics of the contamination source to provide critical information for projecting future air contamination; 4) the health-risk evaluation of the current air contamination based on the estimation of human exposure given the contaminant concentration distribution and the prediction of the immediate and long-term health and performance effects due to an estimated contaminant intake; and 5) a prediction of the future health risk based on the projection of current trends.

On-line automatic health-risk evaluation must rely on a database of health hazard effects of preselected contaminants and the toxicology dose-response information. It must also be outfitted with a decision-making subsystem that would allow us to classify the current contamination situation into the health hazard classes. The classification of current air contamination should invoke an appropriate health-risk mitigating action ranging from "no action" to "immediate evacuation," corresponding to "tolerable" or "extremely dangerous" contamination.

The most important inference that can be drawn from the analysis of on-line concentration measurements is the early detection of an acute or creeping contaminant release. The detection process consists of two steps: residual generation and classification. Ideally, a residual signal is equal to zero during normal operation and deviates substantially from the zero value when a contamination event happens. A decision logic is then used on the residual signal to make a classification into either the fault or no-fault operation. Depending on how a residual signal is generated, detection methods can be

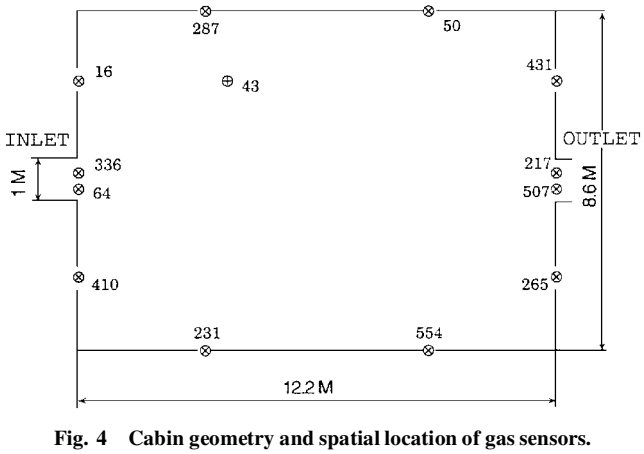


Fig. 4 Cabin geometry and spatial location of gas sensors.

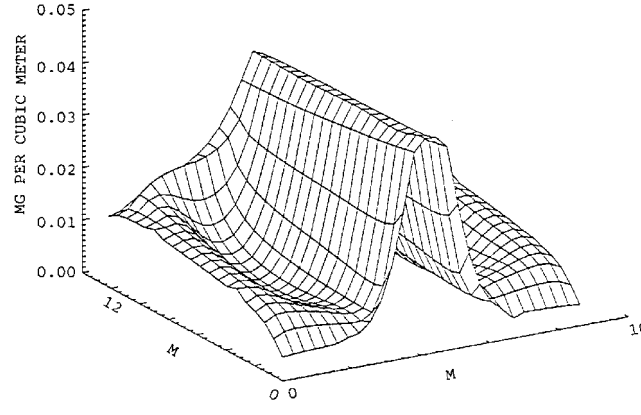


Fig. 6 Estimated contaminant concentration based on stochastically perturbed model and real-time noisy measurements.

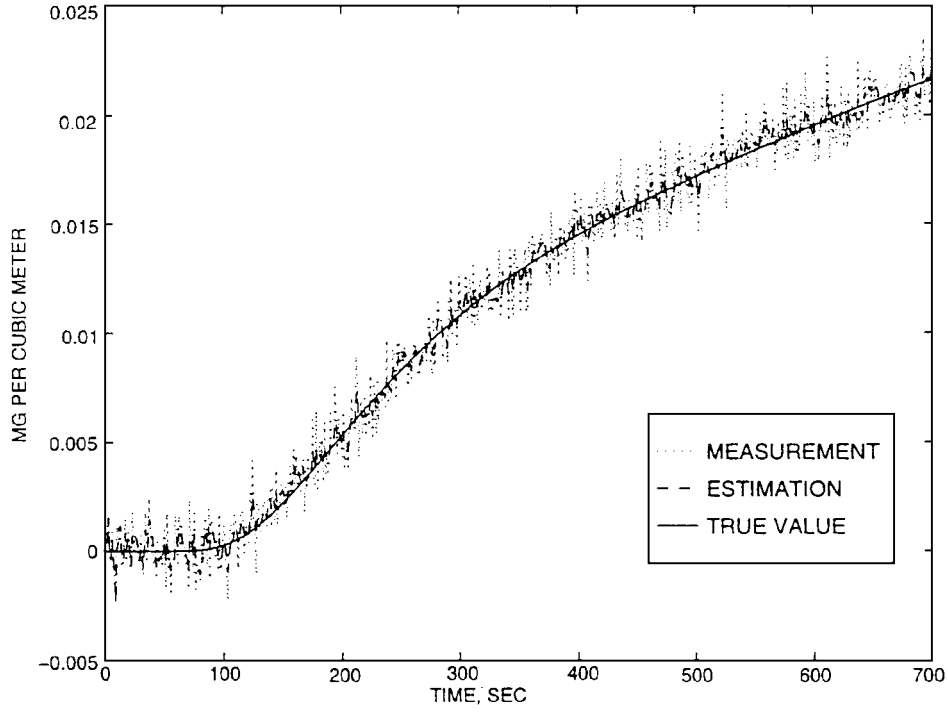


Fig. 5 Estimation of contaminant concentration at spatial point 50.

either model based or purely statistical. If the model is not used, the implementation of fault detection is based solely on the analysis of a change in statistical properties of the measurement data. One classical statistical detection method is a Shewhart control chart, which in essence is a continuous inspection algorithm of a change in the mean value of a measurement signal. Another example is a cumulative sum algorithm, which detects the change in the measurement signal based on a simple integration of the signal with an adaptive threshold. An extensive account of the statistical fault detection methods can be found in Ref. 28.

Fault detection based on the statistical analysis of the measurement time series from multiple sensors is sensitive to the sensor degradation. In fact, using the statistical approach, the symptoms of a faulty primary sensor, such as bias or fluctuation, are often indistinguishable from the symptoms induced by the contamination events. To isolate a sensor failure, a statistical detection requires the utilization of hardware redundancy with majority voting. In this case, at least three identical sensors should be used to acquire a single measurement reading. If the readings from all sensors are the same within the tolerance limit, then the normal instruments operation is assumed. On the contrary, when the reading from one of the sensors differs from the rest of the measurements, its output is ignored, and the sensor is considered faulty. Clearly, in spacecraft applications where a premium for weight reduction is high, it is desirable to avoid hardware redundancy whenever possible. In all applications, as the number of sensors, transmitters, and cables increases, the cost goes up, and the increased number of elements leads to higher maintenance requirements. Furthermore, statistical analysis of the measurement data alone does not provide sufficient information for fault diagnosis, with an exception of the rare case of directly measurable contamination events.

An alternative approach to fault detection is composed of model-based detection schemes. The foundation of this approach is a concept of analytical (or functional) redundancy, which states that given the knowledge about a process in the form of its mathematical model, it should in principle be possible to use measurable characteristics of a process to draw a conclusion about unmeasurable characteristics and events such as an instrument failure or contamination event. Though some model-based methods do not require the distinction between the type of the fault affecting the system, i.e., sensor, actuator, or process fault, we begin with sensor-failure detection. Our objective is limited. We will only show how the state estimations, generated by the monitoring system, can be used to solve a fault detection problem without reverting to hardware redundancy. Survey papers give reviews of instrument fault detection.[29,30]

An obvious way to detect instrument fault is to compare the measurement estimations $\hat{z} = H\hat{q}$, produced by the bank of Kalman filters, each driven by a single component of the measurement vector z. After a sensor fails, the corresponding Kalman filter will generate estimates that differ statistically from the estimates produced by the rest of the Kalman filters, indicating both the fact that the failure occurred and the affected sensor. This scheme is known as a dedicated filter (observer) method[31] and can be used to detect and to isolate multiple simultaneous sensor faults. Its drawback is a high computational cost associated with the propagation of multiple state estimates and error covariance matrices.

A modification of the dedicated filter scheme known as a generalized filter (observer) method[32] associates each sensor with a dedicated filter driven by all available measurements except the measurement from the respective sensor. The generalized filter allows the system to reliably detect and to isolate a single sensor fault, but at a higher computational cost. A single fault limitation is usually insignificant, because multiple simultaneous sensor failures have low probability.

To lower the computational cost of detection and isolation of the instrument fault, we recommend the modification of the dedicated filter approach, which requires only three concurrently run Kalman filters to detect and to isolate a single failed sensor. The measurement vector z on every time step $m+1$ is first partitioned as

$$z_{m+1} = \begin{pmatrix} z^1_{m+1} \\ z^2_{m+1} \end{pmatrix} \tag{45}$$

The vector z and its components z^1 and z^2 can now be used to drive three different Kalman filters. To be more specific, we assume that the model of the process is given in the form of Eqs. (22) and (23). Then the monitoring system, governed by the implicit filter equation (26), is supplemented by the following two filters, driven by z^1 and z^2:

$$A_1\hat{q}^{z^1}_{m+1\,|\,m+1} = \left[A_2 - L^{z^1}_y H^{z^1}_1 A_2\right]\hat{q}^{z^1}_{m\,|\,m} + \left[I - L^{z^1}_y H^{z^1}_1\right]f_m + L^{z^1}_y z^1_{m+1} \tag{46}$$

and

$$A_1\hat{q}^{z^2}_{m+1\,|\,m+1} = \left[A_2 - L^{z^2}_y H^{z^2}_1 A_2\right]\hat{q}^{z^2}_{m\,|\,m} + \left[I - L^{z^2}_y H^{z^2}_1\right]f_m + L^{z^2}_y z^2_{m+1} \tag{47}$$

where the superscripts in $\hat{q}^{z^1}_{m\,|\,m}$ and $\hat{q}^{z^2}_{m\,|\,m}$ are used to specify which subset of measurements is used to obtain the concentration estimate, $H_1 = [H^{z^1}_1 \;\; H^{z^2}_1]$ is the appropriately partitioned modified measurement matrix, and $L^{z^1}_y$ and $L^{z^2}_y$ are the appropriate IKF gains.

The three filters, Eqs. (26), (46), and (47), are now used to generate the following residual signals:

$$r_0 = z_{m+1} - H\hat{q}_{m+1\,|\,m+1} \tag{48}$$

$$r_1 = z_{m+1} - H\hat{q}^{z^1}_{m+1\,|\,m+1} \tag{49}$$

and

$$r_2 = z_{m+1} - H\hat{q}^{z^2}_{m+1\,|\,m+1} \tag{50}$$

A decision logic of the instrument fault detection is based on the fact that, if no fault has occurred, then all residuals r_i are the white noises with zero mean and known covariances. Thus a statistical test of whiteness, change in mean, or variance can be used to detect a fault. Note that, due to the modeling errors, the mean value of r_i may not be zero during normal operation and the detection logic must be based on the increment of a mean value of the residual above the threshold level rather than on its deviation from zero. The detection sensitivity to the modeling errors also raises an important problem of false alarms and motivates the research into robust detection and isolation methods.

The residuals, Eqs. (48–50), also carry information about an unknown contamination event. A contaminant release will affect all three residuals, provided that both sets of measurements z^1 and z^2 are influenced by the release. At the same time, a failure of a single sensor will affect only two residuals. Thus the residual analysis can be used to make a distinction between instrument fault and the contamination event. Suppose that a reading from a faulty gas sensor is a component of z^2 and, therefore, r_0 and r_2 are affected. The inconsistency of r_1 with both r_0 and r_2 is an indication of an instrument fault, and it also points to the subset z^2 as a source of erroneous measurement.

After the instrument failure is detected, the monitoring system must be reconfigured to temporarily ignore the readings from all sensors contributing to z^2 until a precise isolation of a failed sensor is made. One possible isolation approach is to further subdivide z^2 into

$$z^2 = \begin{pmatrix} z^{21} \\ z^{22} \end{pmatrix}$$

A comparison of the residuals from two Kalman filters driven by z^{21} and z^{22} with the reference residual r_1 will limit the source of the instrument fault to a smaller subset of the sensors. The unaffected group of measurements should now be added to z^1, whereas the faulty subset must be further subdivided until a single failed sensor is identified.

Depending on the particular situation, it may be important to estimate a type of the instrument fault. Willsky[33] distinguishes a hard failure, when the sensor reading has nothing to do with reality, and a degradation of the sensor performance, such as a biased reading or increased noise. Clearly, the readings from a hard failed sensor must be ignored. At the same time, a degraded sensor can still be used as a source of useful information, provided its degradation is

taken into account. This requires an automatic instrument fault estimation, including identification of a sensor bias and a new variance of measurement noise. Such a complication of the detection and isolation algorithm can be justified if there is no hardware redundancy to substitute for a failed sensor and if its loss leads to a significant deterioration of the overall performance of the monitoring system.

The hardware and analytical redundancies are not mutually exclusive. Deckert et al.[34] designed an instrument fault detection system for a fly-by-wire aircraft based on the dual redundant sensors and analytical redundancy. Their approach is to detect a sensor failure using the comparison of the output from two redundant sensors and, once a failure is detected, to trigger an analytical procedure to determine which of two sensors has actually failed. Another example of a combined hardware and analytical redundancy is reported in Ref. 35.

An advantage of the combined hardware and analytical redundancy approach is that only doubly redundant instrumentation is required and at the same time the computational burden of analytical fault detection and isolation is reduced. The combined approach integrates a high fidelity of the hardware redundancy and the ability for complete system recovery after the failed instrument is identified with the inference power of model-based methods.

The isolation of an unknown contamination event, which includes the identification of the spatial location and the emission rate of the contaminant source, is considered in Refs. 21, 22, 36, and 37.

Example: Detection of an Unknown Contamination Event

Consider the cabin described in the previous example. The air-quality monitoring system, which uses two-dimensional approximation of the contaminant transport, is described by the implicit Kalman filter, Eq. (26). Assume that of 12 available gas sensors, the locations of which are depicted in Fig. 3, only 10 sensors are actually used in the IKF. The outputs of the two remaining sensors, located at the spatial points 231 and 287, are used to generate the following residuals:

$$r^{231}_{m+1} = z^{231}_{m+1} - \hat{q}^{231}_{m+1|m+1} \tag{51}$$

$$r^{287}_{m+1} = z^{287}_{m+1} - \hat{q}^{287}_{m+1|m+1} \tag{52}$$

where z^{i}_{m+1} is the measured contaminant concentration at point i and $\hat{q}^{i}_{m+1|m+1}$ is the IKF estimate of the contaminant concentration at point i, $i = \{231, 287\}$. The detection strategy is based on the fact that in the absence of unknown contamination sources the residuals r^{231}_{m+1} and r^{287}_{m+1} have a zero mean. The variance of r^{231}_{m+1} is equal to

$$(\sigma_{231})^2 = E\left[\left(r^{231}_{m+1}\right)^2\right] = R_{231} + \left[\boldsymbol{P}^{q}_{m+1|m+1}\right]_{231} \tag{53}$$

where R_{231} is the variance of the measurement noise of the sensor 231 [equal to an appropriate diagonal component on the covariance matrix, Eq. (25)] and $[\boldsymbol{P}^{q}_{m+1|m+1}]_{231}$ is the 231st diagonal component of the IKF estimation error covariance matrix, governed by Eq. (33). A similar equation yields the variance $(\sigma_{287})^2$ of the residual r^{287}_{m+1}.

The continuous inspection of the mean of the residuals r^{231}_{m+1} and r^{287}_{m+1} is used to detect the change in the process. The alarm is triggered for the first time when any of the elements of the decision vector

$$\bar{r}^{i}(K) = \left| \frac{1}{N^i} \sum_{j=N^i(K-1)+1}^{N^i K} r^{i}_{j} \right| \tag{54}$$

exceeds or equals the value of the corresponding component of the threshold vector

$$LIM^{i} = \kappa^{i}\left(\sigma_i \big/ \sqrt{N^i}\right), \qquad i = \{231, 287\} \tag{55}$$

The algorithm described by Eqs. (54) and (55) can be summarized in the following form:

$$\begin{aligned} &\text{normal operation if} \quad \bar{r}^{i}(K) < LIM^{i} \\ &\text{alarm triggered if} \quad \bar{r}^{i}(K) \geq LIM^{i} \end{aligned} \qquad i = \{231, 287\} \tag{56}$$

and is known as the Shewhart control chart. The size of the sliding average window N^i and parameter κ^i are used as the tuning parameters. They control the tradeoff between the sensitivity of the detection algorithm and the frequency of false alarms. Additional improvement of the detection reliability at the expense of a longer detection time can be achieved if the alarm is triggered after the residual consecutively exceeds the threshold an assigned number of times.

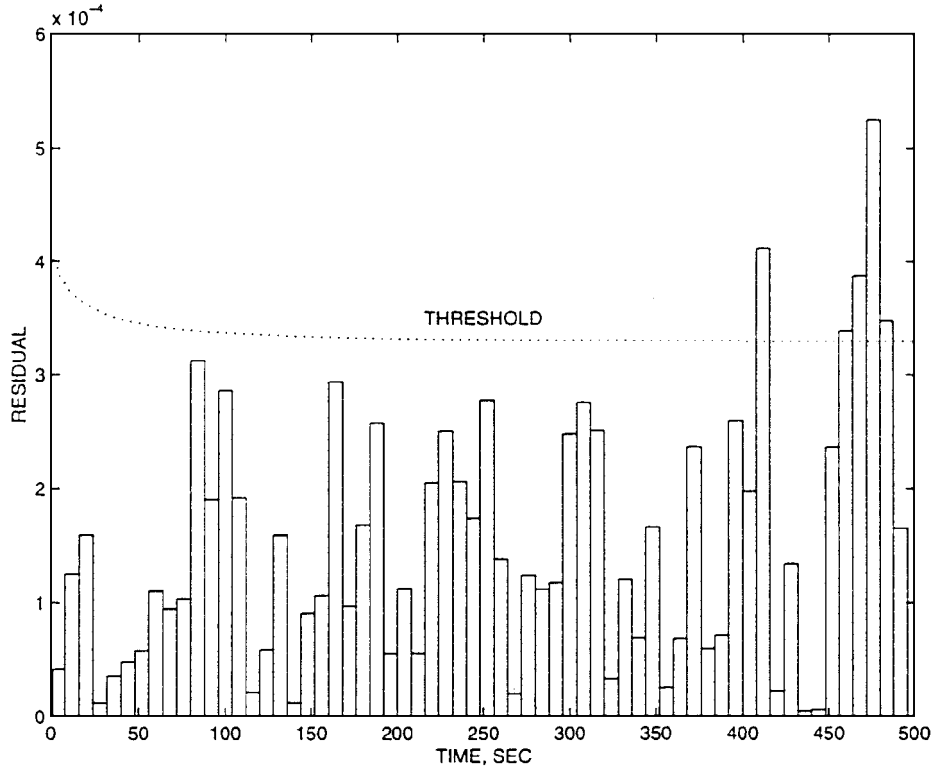


Fig. 7 Response of the residual r^{231}_{m+1} to an unknown contamination source: point 231.

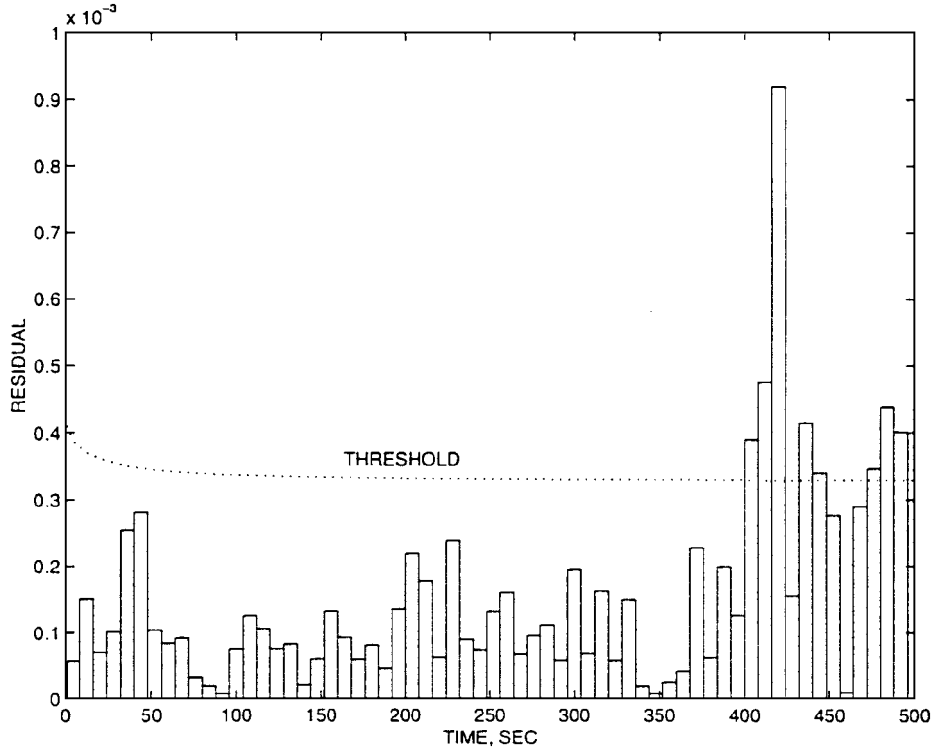
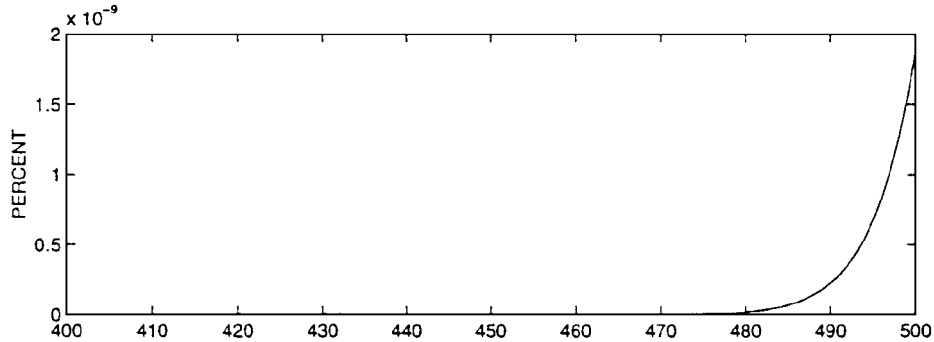
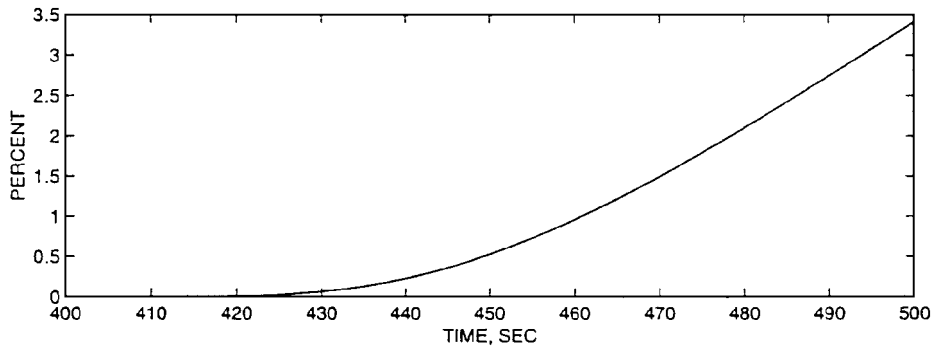


Fig. 8 Response of the residual r^{287}_{m+1} to an unknown contamination source: point 287.



Point 231



Point 287

Fig. 9 Relative effect of unknown source on the contaminant concentration in spatial.

Assume that the unmodeled point contamination source with the constant rate of emission (Table 2) has been applied to the system at $t = 401$ s. The response of the detection system on the appearance of a new and unknown contamination source (NUCS) is depicted in Figs. 7 and 8 for $N^{231} = N^{287} = 8$ and $\kappa^{231} = \kappa^{287} = 0.8$. Both residuals exceed their threshold limits after 8 s since the application of NUCS. However, residual r^{231} provides less conclusive results. This can be explained by observing that the effect of NUCS on the concentration measurement of sensor 231 is very small. In fact, this effect is also quite small for sensor 287, as can be seen from Fig. 9, where we plot the model predicted relative effect of NUCS, calculated according to the following expression:

$$\text{relative effect of NUCS} = \left(1 - \frac{\text{concentration with NUCS}}{\text{concentration without NUCS}}\right) \times 100\% \tag{57}$$

Applied NUCS can be interpreted as a small leak. For the first 100 s, its effect on the measurements of sensor 287 does not exceed a few

Table 2 Capacity and location of unmodeled source: unknown source

Location	Spatial point 43
Capacity	0.02 mg/(m^3 · s)

percentage points. Nevertheless, the residual r^{287} shows a clear trend of deviation from a zero mean.

As discussed earlier, the detection and isolation of a faulty sensor require multiple Kalman filters that use measurements from different subsets of available sensors to generate the estimates of the contaminant concentration.

Conclusions

The perspective of long-term and remote space missions requires the development of a sophisticated environmental control system, which uses a real-time distributed parameter contaminant transport model for air-quality monitoring and decision making. Existing and anticipated control engineering, gas measurements, and computer technologies create the basis for developing an advanced air-quality control system for the enclosed and remote environments with such features as on-line estimation of the contaminant concentration distributions, early detection of the contamination events, the identification of unknown sources of contamination, and health-risk management and mitigation. In this paper we propose an integrated approach to the air-quality monitoring and emission detection that is based on the distributed parameter model of the contaminant transport and the IKF of on-line contaminant concentration measurements. Results of the computer experiments, described in this paper, show the feasibility of the real-time estimation of the contaminant concentration profile and early emission detection.

Acknowledgments

The authors acknowledge the support of the NASA Specialized Center for Research and Training on Environmental Health under Grant 89988-G004 and NASA support under Grant NRA93-OLMSA-07.

References

[1]Berglund, B., Grimsrud, D. T., and Seifert, B., "Special Issue on Indoor Air Quality," *Environment International*, Vol. 15, Nos. 1–6, 1989.

[2]Berglund, B., Berglund, U., Lindvall, T., Seifert, B., and Sundall, J., "Special Issue on Indoor Air Quality," *Environment International*, Vol. 12, Nos. 1–4, 1986.

[3]Mølhave, L., Bach, B., and Pedersen, O. F., "Human Reactions to Low Concentrations of Volatile Organic Compounds," *Environment International*, Vol. 12, Nos. 1–4, 1986, pp. 167–175.

[4]James, J. T., and Coleman, M. E., "Toxicology of Airborne Gaseous and Particulate Contaminants in Space Habitat," *Life Support and Habitability*, AIAA, Washington, DC, 1994, pp. 37–60.

[5]James, J. T., Limero, T. F., Leano, H. J., Boyd, J. F., and Covington, P. A., "Volatile Organic Contaminants Found in the Habitable Environment of the Space Shuttle: STS–26 to STS–55," *Aviation, Space and Environmental Medicine*, Vol. 65, No. 9, 1994, pp. 851–857.

[6]Todd, P., Skliar, M., Ramirez, W. F., Smith, G., Morgenthaler, G., Mckinon, J., Oberdöerster, G., and Schulz, J., "Inhalation Risk in Low-Gravity Spacecraft," *Acta Astronautica*, Vol. 33, 1994, pp. 304–315.

[7]Todd, P., Skliar, M., Smith, G., Ramirez, W. F., Oberdöerster, G., and Morgenthaler, G., "Physics, Chemistry and Risk of Thermodegradation in Long-Mission Space Flight," *Proceedings of the Twenty-Third SAE International Conference on Environmental Systems* (Colorado Springs, CO), Society of Automotive Engineers, Warrendale, PA, 1993, pp. 1–12 (SAE TP 932144).

[8]Coleman, M. E., and James, J. T., "Airborne Toxic Hazards," *Space Biology and Medicine*, Lea and Febiger, Philadelphia, PA, 1994, pp. 141–156.

[9]Covault, C., "Explosion and Leak Cripples Salyut-7 Effort," *Aviation Week and Space Technology*, Oct. 10, 1983, pp. 23–26.

[10]Peto, P. G., "Results of Soviet-Hungarian Space Research," *East European Report*, No. 699, April 1981, pp. 4–12.

[11]National Research Council Committee on Toxicology, *Spacecraft Maximum Allowable Concentrations*, National Academy Press, Washington, DC, 1994.

[12]National Research Council Committee on Toxicology, *Guidelines for Developing Spacecraft Maximum Allowable Concentrations for Space Station Contaminants*, National Academy Press, Washington, DC, 1992.

[13]White, R. J., and Lujan, B. F., "Current Status and Future Direction of NASA's Space Life Sciences Program," *Working in Orbit and Beyond: The Challenges for Space Medicine*, Vol. 72, Science and Technology Series, American Astronautical Society, San Diego, CA, 1989, pp. 1–7.

[14]Logan, J. S., "Health Maintenance on Space Station," *Working in Orbit and Beyond: The Challenges for Space Medicine*, Vol. 72, Science and Technology Series, American Astronautical Society, San Diego, CA, 1989, pp. 87–99.

[15]Godish, T., *Sick Buildings. Definition, Diagnosis and Mitigation*, CRC Press, Boca Raton, FL, 1995.

[16]Mølhave, L., "The Sick Building and Other Buildings with Indoor Climate Problems," *Environment International*, Vol. 15, 1989, pp. 65–74.

[17]Kocache, R., "Gas Sensors," *Sensor Review*, Vol. 14, No. 1, 1994, pp. 8–12.

[18]Berry, C., and Brackbenbury, A., "On-Line Gas Analysis," *Measurement and Control*, Vol. 24, Oct. 1991, pp. 231–236.

[19]*Airborne Particle Measurement in the Space Shuttle*, NASA, Washington, DC, 1989.

[20]*Indoor Pollutants*, National Academy of Sciences, National Academy Press, Washington, DC, 1981.

[21]Skliar, M., and Ramirez, W. F., "Source Identification in the Distributed Parameter Systems," *Applied Mathematics and Computer Science* (to be published).

[22]Skliar, M., and Ramirez, W. F., "Source Identification of the Distributed Parameter Transport Processes," *Proceedings of the American Control Conference* (Albuquerque, NM), American Automatic Control Council, Evanston, IL, 1997, pp. 2228–2232.

[23]Bird, R. B., Steward, W. E., and Lightfoot, E. N., *Transport Phenomena*, Wiley, New York, 1960.

[24]Son, C., and Barker, R. S., "U.S. Lab—A Module Cabin Air Distribution in Space Station," *Proceedings of the Twenty-Third SAE International Conference on Environmental Systems* (Colorado Springs, CO), Society of Automotive Engineers, Warrendale, PA, 1993, pp. 1–15 (SAE TP 932192).

[25]Douglas, J., "Alternating Direction Methods for Three Space Variables," *Numerische Mathematik*, Vol. 4, 1962, pp. 41–63.

[26]Skliar, M., and Ramirez, W. F., "Implicit Kalman Filtering," *International Journal of Control*, Vol. 66, No. 3, 1997, pp. 393–412.

[27]Skliar, M., and Ramirez, W. F., "Square Root Implicit Kalman Filtering," *Proceedings of the 13th IFAC World Congress* (San Francisco, CA), 1996, pp. 251–256.

[28]Basseville, M., and Nikiforov, I. V., *Detection of Abrupt Changes: Theory and Applications*, Prentice–Hall, Englewood Cliffs, NJ, 1993.

[29]Halme, A., and Selkainahe, J., "Advanced Fault Detection for Sensors and Actuators in Process Control," *Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS'91)*, edited by R. Isermann and B. Freyermuth, Pergamon, Oxford, England, UK, 1992, pp. 91–97.

[30]Emami-Naeini, A., Akhter, M. M., and Rock, S. M., "Robust Detection Isolation and Accommodation for Sensor Failures," NASA CR-174825, July 1986.

[31]Clark, R. N., "Instrument Fault Detection," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 14, No. 3, 1978, pp. 456–465.

[32]Frank, P. M., "Fault Diagnosis in Dynamic Systems via State Estimation—A Survey," *System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches*, edited by S. Tzafestas, M. Singh, and G. Schmidt, Vol. 1, Reidel, Dordrecht, The Netherlands, 1987, pp. 35–98.

[33]Willsky, A. S., "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, Vol. 12, No. 6, 1976, pp. 601–611.

[34]Deckert, J. C., Desai, M. N., Deyst, J. J., and Willsky, A. S., "F-8 DFBW Sensor Failure Identification Using Analytic Redundancy," *IEEE Transactions on Automatic Control*, Vol. 22, No. 5, 1977, pp. 795–803.

[35]Labarrère, M., "Aircraft Sensor Failure Detection by Analytic Redundancy," *Systems and Control Encyclopedia*, edited by M. G. Singh, Vol. 1, Pergamon, Oxford, England, UK, 1987, pp. 246–251.

[36]Skliar, M., "State Estimation and Fault Diagnosis for Distributed Parameter Transport Processes with Application to Air Contamination Control," Ph.D. Thesis, Dept. of Chemical Engineering, Univ. of Colorado, Boulder, CO, 1996.

[37]Skliar, M., and Ramirez, W. F., "Monitoring and Detection of Indoor Air Contamination," *Proceedings of the 1997 American Control Conference* (Albuquerque, NM), American Automatic Control Council, Evanston, IL, 1997, pp. 319–323.

I. D. Boyd
Associate Editor