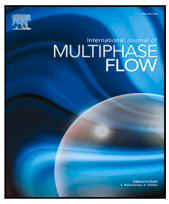
International Journal of Multiphase Flow 188 (2025) 105201

Contents lists available at ScienceDirect

International Journal of Multiphase Flow

journal homepage: www.elsevier.com/locate/ijmulflow



Research Paper

Twin experiments for data assimilation of cavitating flow around a hydrofoil

Shungo Okamura, Kie Okabayashi *

Department of Mechanical Engineering, Osaka University, 2-1 Yamadaoka, Suita, 565-0871, Osaka, Japan

ARTICLE INFO

Keywords:
Cavitation
Hydrofoil
Twin experiment
Local ensemble transform Kalman filter
Data assimilation

ABSTRACT

Twin experiments are conducted to clarify whether assimilation effects can be achieved by data assimilation (DA) with measurement data obtained by existing cavitation flow measurement techniques. The analysis object is the cavitation flow around a Clark-Y11.7% hydrofoil. The pseudo-measurement data are velocity fields from tomographic particle image velocimetry (TPIV) or two-dimensional PIV, both containing missing data in cavity, along with indirectly obtained cavity interface shapes. A large-eddy simulation is used for unsteady simulation of cavitating turbulent flow, and a local ensemble transform Kalman filter is used as the DA method. Visualized flow fields of the ensemble mean show that characteristic phenomena of the pseudo-measurement data are qualitatively reproduced. In addition, the time-series data at an observation point located at the position where the pseudo-measurement data exists converged following the pseudo-measurement data. The velocity inside the cavity, where no pseudo-measurement data exists, is also complemented by CFD that incorporates information from outside of the cavity. However, the complementation performance depends on the accuracy of the cavitation model. This DA program is applied to the real PIV data of single-phase flow and qualitatively reproduces the flow field. Furthermore, the observation noise is reduced and the data outside the measurement domain are complemented.

1. Introduction

Since the 1990s, methods for analysis of cavitation flow as a gas–liquid two-phase flow based on Navier–Stokes equation have been actively developed. Currently, these methods account for the mainstream of practical cavitation flow analysis methods for fluid machinery. These cavitation models are broadly classified into homogeneous model and two-fluid model.

In the homogeneous fluid model, cavitation phenomena are expressed by the density change of the homogeneous mixture fluid composed of liquid and vapor phases. The mixture density ρ_m is defined by

$$\rho_m = \alpha\rho_V + (1-\alpha)\rho_L \tag{1}$$

where α denotes vapor volume fraction and ρ_L and ρ_V denote the density of liquid and vapor, respectively. Furthermore, depending on the method used to model the change in ρ_m, the homogeneous fluid models are classified into the following categories: equation of state models in which ρ_m is expressed as a function of pressure only (Delannoy and Kueny, 1990; Coutier-Delgosha et al., 2003); transport equation models that include source terms representing the evaporation rate and condensation rate (Merkle et al., 1998; Kunz et al., 2000; Singhal et al., 2002; Schnerr and Sauer, 2001; Zwart et al., 2004); bubble dynamics models that compute the vapor volume fraction from the representative bubble diameter, which are calculated from a simplified Rayleigh–Plesset (R–P) equation, and the fixed bubble number density (Kubota et al., 1992; Ito et al., 2005); and simple models that include source terms based on the difference between pressure and saturated vapor pressure to the transport equations for density or void fraction (Chen and Heister, 1995; Okita and Kajishima, 2002). Homogeneous fluid models are implemented in commercial software and widely used in industry owing to its simplicity and low computational cost. However, in two-fluid models, the basic equations for both vapor and liquid phases are solved. The two-fluid model is classified into two types, the Eulerian type and the Lagrangian type, depending on the method used to predict the spatial distribution of bubbles. The bubble dynamics process in both types is based on the R–P equation. Tamura et al. (2001) and Tamura and Matsumoto (2009) proposed a Eulerian-type two-fluid model, in which the transport equation for the number density of bubbles is solved. They computed vapor and liquid velocities u_V and u_L, respectively, and considered the slip velocity, $u_V - u_L$, in R–P equation. The Lagrangian-type two-fluid model tracks bubbles based on Newton's equations of motion in the Lagrangian framework, while the liquid phase is solved in the Eulerian framework. In the early stages, a method using one-way coupling, which assumes that the motion of discrete bubbles does not affect the Eulerian liquid phase, was proposed (Hsiao

* Corresponding author.
E-mail address: okabayashi@mech.eng.osaka-u.ac.jp (K. Okabayashi).

https://doi.org/10.1016/j.ijmultiphaseflow.2025.105201
Received 16 April 2024; Received in revised form 17 January 2025; Accepted 26 February 2025
Available online 8 March 2025

et al., 2003; Hsiao and Chahine, 2005). However, it was difficult to convert between structures of different scales, such as the coalescence and breakup of bubbles. Vallier (2013) realized two-way coupling by combining a method that treats macroscopic structures using a Eulerian frame and discrete bubbles using a Lagrangian frame, and a method that transitions between structural scales along with coalescence and breakup. Subsequently, the model has been improved from various perspectives (Wang et al., 2021, 2023). Although the computational cost of the two-way coupling is high, the inception and collapse of subgrid bubbles can be considered. Therefore, it has an advantage in analyzing cavitation erosion and acoustics.

Therefore, cavitation models have been developed from various perspectives. However, as cavitation is a phenomenon with a variety of morphologies and multi-scale spatio-temporal characteristics, no model has yet been developed that can represent the various cavitation phenomena and related phenomena in a unified manner. Even for the most basic flow, such as those around a hydrofoil, current models have been unable to quantitatively reproduce cavitation phenomena. It has been pointed out that this is due to the inability of a single model to deal with locally different cavitation behaviors and their related physics (Kato, 2011).

As a breakthrough in the development of cavitation models, our research group has been working on the development of a "data-driven cavitation model" (Noda and Okabayashi, 2023) that can represent cavitation phenomenon accurately and generally by using a machine learning model trained on measurement data, rather than a conventional mathematical model. This model is inspired by the fact that the measured flow field data, which are obtained by the recent optical measurement techniques, can be regarded as pixels in image data (Fenjan et al., 2016; Ling et al., 2016). The flow field data are input as training data for neural network (NN) of image recognition, and a certain relation between input and output is obtained as a form of NN. The NN, i.e., data-driven cavitation model, is coupled with the governing equation as an alternative to the conventional mathematical cavitation model. Noda and Okabayashi (2023) established a framework of data-driven model with CFD data as the training dataset, but if measurement data could be used for training, the data-driven model would be even more reflective of real physical phenomena.

The measurement data used to train the NN are assumed to be the velocity field data of cavitating flow obtained by particle image velocimetry (PIV), high-speed camera images, and wall-mounted pressure sensor data etc. However, raw PIV data are not suitable as training data because it lacks information inside the bubbles and near the boundary layer. Moreover, it contains noise and measurement errors, and generates numerical conservation problems when combined with computational fluid dynamics (CFD). Furthermore, there is currently no method for measuring the spatio-temporal distribution of bubbles, which is an important physical quantity in cavitating flow, at high resolution. Pervunin et al. (2021) proposed a related method for measuring the time-averaged bubble population distribution from the spatio-temporal particle abundance of PIV. Nevertheless, this method does not provide the time-series data of well-defined gas–liquid interfaces and the bubble population distribution. As mentioned above, experimental measurement of cavitating flow is challenging; however, in the future, it would be ideal to be able to construct machine learning models based on readily available measurement data.

Thus, in the analysis of cavitating flow, where both numerical simulations and experimental measurements are not yet matured, the data assimilation (DA) approach is expected to be effective. DA is a method for statistically modifying uncertainties in numerical simulations with measurement data and has been widely used in the field of numerical weather prediction (Dee et al., 2011; Honda et al., 2018). The advantage of this method is that it can both improve the results of numerical simulations with measurement data and supplement unmeasured data with CFD. The introduction of DA to cavitating flow analysis may lead to a more detailed understanding of the phenomenon and improve the accuracy of cavitation models. Furthermore, if this method can reproduce the cavitating flow field with the missing data complemented, it could be used to create a training dataset for the construction of the machine learning model described above. For example, Brajard et al. (2021) used a dataset created by DA as the training data for constructing a data-driven climate model using machine learning. The two-scale Lorenz model and the Modular Arbitrary-Order Ocean–Atmosphere Model (MAOOAM), a type of coupled atmosphere–ocean model that is utilized to predict changes in temperature and other parameters due to air and water flow, were used for validation. DA was used to supplement missing data in a realistic scenario with noisy and spatially sparse observations. They constructed machine learning models using the training datasets created in this way and reported that the constructed models showed high predictive performance for both validation cases. The machine learning models in the example above adopt a concept that is similar to that of data-driven cavitation models and demonstrate the suitability of DA for creating training data sets.

Recently, the four-dimensional variational method (Rabier et al., 2000) based on optimal control theory and the ensemble Kalman filter (EnKF) (Evensen, 1994), a sequential method based on Bayesian estimation theory, have been identified as representative DA methods. These methods are expected to be utilized in the field of fluid engineering because of their high affinity with CFD. The applications of DA in fluid engineering include estimation of initial and boundary conditions (Mons et al., 2016; Sousa and Gorlé, 2019), parameter optimization (Kato et al., 2015; Deng et al., 2018), improvement of measurement systems (Misaka and Obayashi, 2014), and creation of data sets that integrate measurement data with numerical simulations. We focus on the creation of datasets, but there are currently no studies that focus on the creation of datasets using this method in fluid engineering field. However, with the recent trend toward complementary approaches between experimental fluid dynamics (EFD) and CFD, represented by the Virtual Diagnostics Interface System (ViDI) (Schwartz and Fleming, 2007) at the NASA Langley Research Center and the Digital/Analog-Hybrid Wind Tunnel (DAHWIN) (Watanabe et al., 2014) at the Japan Aerospace Exploration Agency (JAXA), the need for analytical data sets is expected to increase in the future as DA techniques become more sophisticated.

In this study, we perform a basic validation of assimilation algorithm performance to establish a dataset of cavitating flow. Several examples of such basic validations exist in the field of fluid engineering, although their main objective is not to only create datasets. Labahn et al. (2020) used the EnKF to validate the performance of an assimilation algorithm for turbulence state estimation. They assimilated a three-dimensional velocity field obtained from tomographic PIV (TPIV) into a turbulent jet with a Reynolds number of 13500 and verified the reproducibility of the statistical behavior of the turbulent field. They systematically determined the effects of parameters and measurement conditions on the assimilation performance by changing the EnKF parameters such as the number of ensemble members and localization radius, as well as parameters related to measurement conditions, such as measurement data uncertainty, assimilation frequency, and data sparsity. They also confirmed that the EnKF is highly robust to measurement data. As a basis for the integration of wind tunnel experiments and CFD using DA, Kato and Obayashi (2011) verified the effectiveness of DA for the correction of the uncertainties in numerical simulations of flows around a square cylinder: They assimilated measured surface pressure data at only three points on the cylinder using the EnKF. Although the direct correction of the flow field by assimilation is only performed near the square cylinder, it also improves the reproducibility of the wake of the cylinder. These examples suggest that DA using the EnKF is effective even for cavitation flows, which are highly unsteady, turbulent, and are expected to have a large spatial data deficit. However, there are few examples of actual applications of DA to such flow fields, and it is not clear whether sufficient assimilation can be achieved

Table 1
Problem settings.

Section	Problem setting	Reference data of DA	Assimilation simulation
6	Twin experiment to assimilate pseudo-measurement data assuming TPIV data of cavitating flow.	3D pseudo-measurement data created from the database computed with the OK model.	Cavitating flow simulation with the CH model.
7	Twin experiment to assimilate pseudo-measurement data assuming PIV data of cavitating flow.	2D pseudo-measurement data created from the database computed with the OK model.	Cavitating flow simulation with the CH model.
8	Data assimilation of single-phase airflow.	2D-PIV data (Nonomura et al., 2021).	Single-phase flow simulation.

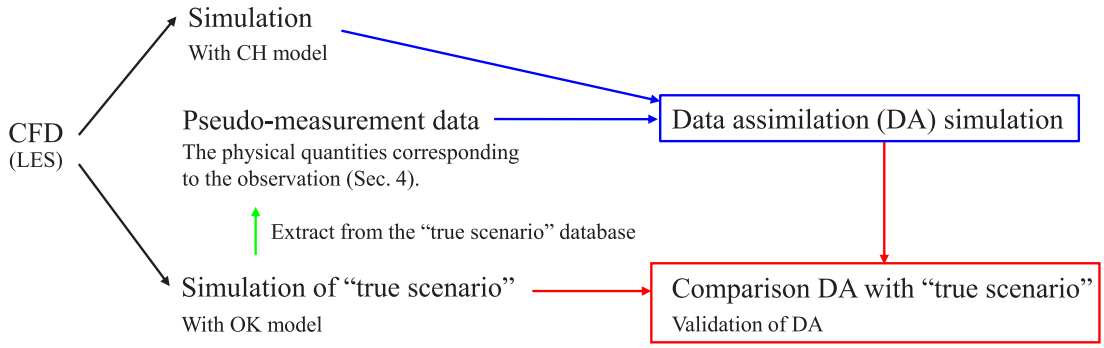


Fig. 1. Schematic diagram of the twin experiment.

using data from cavitating flow fields that can be obtained with existing measurement techniques.

Therefore, in this study, we verify the effectiveness of DA by using cavitating flow data that can be obtained with existing measurement techniques. As one of the most basic examples, cavitating flow around a hydrofoil is employed. Here, we introduce and validate DA through the twin experiment. We demonstrate that the flow field based on pseudo-measurement data can be reproduced using existing cavitation models by introducing DA. We also explore how the spatial dimension of the measurement data influence on the assimilation performance. Furthermore, as we aim to establish a framework for performing assimilation calculations as soon as real measurement data of the cavitating flow field can be obtained, verification using available real measurement data of single-phase flow around an airfoil is conducted.

2. Problem setting

This study aims to reproduce a flow field in which real physical phenomena are incorporated by assimilating measurement data into a numerical simulation of cavitating flow using existing cavitation models. In a problem setting similar to this situation, a twin experiment is conducted. The term "twin experiment" is often used in the field of DA. The schematic diagram of twin experiment is shown in Fig. 1. It is referred to as a "twin" because it computes both the "true scenario" and the predicted value from a single simulation model. The pseudo-measurement data is extracted from the true scenario simulation database. The greatest advantage of the twin experiment is that the true scenario, which is difficult to determine (because measurements can only obtain partial or sparse physical quantities of the flow field), is obtained through the computation, thus, it is possible to determine whether the flow field estimated by DA is correct by comparing the results of DA with the true scenario. Therefore, the twin experiment is a validation of DA.

In this study, three different problems are set as listed in Table 1. In Section 6, 3D-TPIV data of a cavitating flow are assumed as the pseudo-measurement data for the twin experiment. In Section 7, conversely, 2D-PIV data of a cavitating flow are assumed as the pseudo-measurement data to investigate the influence of the spatial dimension of pseudo-measurement data. These pseudo-measurement data are created from the numerical results of cavitating flow simulation using the source-term homogeneous fluid model developed by Okita and Kajishima (2002) (OK model) as the cavitation model. OK model simulation is regarded as the true scenario in this twin experiment. Another source-term homogeneous fluid model developed by Chen and Heister (1995) (CH model), which is simpler than the OK model, is used in the assimilation simulation of both sections. If the CH model, a relatively simple cavitation model, can be used to reproduce the flow field of the OK model, which can represent more complex phenomena, we can demonstrate that DA can be effectively performed for measurement data using an existing cavitation model. Finally, in Section 8, we conduct an assimilation validation on real measurement data of a single-phase airflow. Note that this is not a twin experiment: in this problem, a true scenario does not exist and the pseudo-measurement data is not extracted, either.

3. Overview of CFD

In this section, the overview of the simulation that outputs the database for creating pseudo-measurement data of the cavitating flow and the data assimilation simulation are given. The two different cavitation models, the OK model and CH model, are used for those simulations, respectively; however, the governing equations, turbulence model, numerical methods, and computational conditions are common.

3.1. Governing equations and models

The governing equations of the filtered flow field ($\bar{u}_i$, $\bar{p}$) for a large-eddy simulation (LES) are employed. Hereafter, all variables are non-dimensionalized by the chord length C, the mainstream velocity U_∞, and the liquid density at a sufficiently far position $\rho_{L\infty}$. The density of the homogeneous fluid ρ is approximated as $\rho \approx \rho_L f_L$ by the local liquid density ρ_L and the liquid volumetric fraction f_L. Here, the control volume used to define f_L is the volume of computational cells, and thus $\bar{f}_L = f_L$.

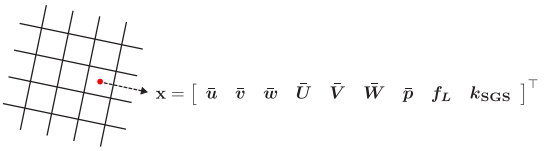


Fig. 2. Illustration of the state variables on each computational grid point.

3.1.1. Governing equations

In cavitating flow, strong pressure fluctuations are generated by the expansion and contraction of cavitation. Therefore, a low-Mach number formulation (Inagaki et al., 2000) is used to account for the compressibility of the liquid phase. The conservation laws for the mass of the liquid phase using the low-Mach number formulation is

$$\frac{Df_L}{Dt} + f_L\left(\mathrm{M}^2\frac{D\bar{p}}{Dt} + \frac{\partial \bar{u}_i}{\partial x_i}\right) = 0. \tag{2}$$

The Mach number $\mathrm{M} = U_\infty/c$ (c: sound speed of a pure liquid) in Eq. (2) is a constant that is uniformly given over the entire computational domain. The filtered Navier–Stokes equation is

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j\frac{\partial \bar{u}_i}{\partial x_j} = -\frac{1}{f_L}\frac{\partial}{\partial x_i}\left(\bar{p} + \frac{2}{3}f_L k_{\mathrm{SGS}}\right) + \frac{\partial}{\partial x_j}\left[2\left(\nu_{\mathrm{SGS}} + \frac{1}{\mathrm{Re}}\right)\bar{S}_{ij}\right]. \tag{3}$$

In Eq. (3), $\bar{S}_{ij}$ represents the rate-of-strain tensor. The subgrid-scale (SGS) turbulent kinetic energy k_{SGS} and SGS kinetic eddy viscosity coefficient ν_{SGS} are described in Section 3.1.3. The Reynolds number is defined as $\mathrm{Re} = U_\infty C/\nu$. From Eqs. (2) and (3), the state vector of the DA system is defined as

$$\mathbf{x} = \left[\ \bar{\boldsymbol{u}}\quad \bar{\boldsymbol{v}}\quad \bar{\boldsymbol{w}}\quad \bar{\boldsymbol{U}}\quad \bar{\boldsymbol{V}}\quad \bar{\boldsymbol{W}}\quad \bar{\boldsymbol{p}}\quad \boldsymbol{f_L}\quad \boldsymbol{k_{\mathrm{SGS}}}\ \right]^\top. \tag{4}$$

Eq. (4) is composed of state variables that progress in time at each grid point (Fig. 2). Here, $\bar{\boldsymbol{u}}$, $\bar{\boldsymbol{v}}$, and $\bar{\boldsymbol{w}}$ are the physical components of the velocity in the x, y and z directions, respectively, and $\bar{\boldsymbol{U}}$, $\bar{\boldsymbol{V}}$, and $\bar{\boldsymbol{W}}$ are the contravariant velocities which appear in Eqs. (2) and (3). They are computed on a collocated grid and defined in cell face generally; however, they are treated as physical quantities similar to other state variables by interpolating to the center of the grid in this computation.

3.1.2. Cavitation model

To reproduce unsteady cavitation phenomena, changes in liquid volumetric fraction f_L are represented by the source-term homogeneous fluid model. In this study, the OK model (Okita and Kajishima, 2002),

$$\frac{Df_L}{Dt} = \left[C_g\left(1-f_L\right) + C_l f_L\right]\left(p - p_v\right), \tag{5}$$

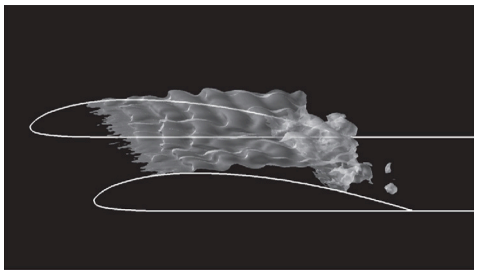
is used for generating the pseudo-measurement data as described in Section 4, and the CH model (Chen and Heister, 1995),

$$\frac{Df_L}{Dt} = C_{\mathrm{CH}}\left(p - p_v\right), \tag{6}$$

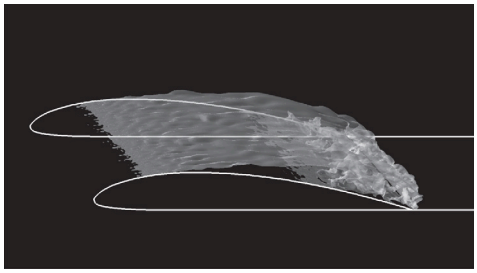
is used for the time evolution of f_L in the assimilation computation. The OK model is a numerical model based on the linearized Rayleigh–Plesset equation, which is an improved version of the CH model, and can reproduce complex phenomena such as three-dimensionality at the sheet cavity interface (Fig. 3(a)). In contrast, the CH model hardly reproduces such cavitation shapes (Fig. 3(b)). Other differences between the flow fields of the OK model and CH model are described in Sections 3.3 and 6.2. The saturated vapor pressure p_v is given by the cavitation number

$$\sigma = \frac{\tilde{p}_\infty - \tilde{p}_v}{\frac{1}{2}\rho_{L\infty}U_\infty^2} \tag{7}$$

where $\tilde{}$ denotes a dimensional quantity and $\tilde{p}_\infty$ represents the pressure at a sufficiently far distance. The constants C_g and C_l of Eq. (5) are optimized for cavitating flow around a square cylinder, with $C_g = 1000$ and $C_l = 1$ for $p < p_v$, and $C_g = 100$ and $C_l = 1$ for $p > p_v$. To attenuate the shock at cavity collapse, C_g is set to be smaller when $p > p_v$. The constant C_{CH} of Eq. (6) is defined as $C_{\mathrm{CH}} = C_0 C U_\infty$ in Chen and Heister (1995), which is effectively proportional to the Reynolds number. It is reported that the larger C_{CH} is, the more the pressure on the object surface converges to a certain value, and C_0 is set to 5000, the value of C_0 in case $\mathrm{Re} = 1.36\times10^5$ in the headform solution. Therefore, the optimal value is estimated as approximately 25000 as the Reynolds number in this study is 6.41×10^5 (described below in Section 3.3). However, in this study, C_{CH} is set to 100 as a reference for C_g, considering the stability of the computation at the cavity collapse described above. The insufficient quantitative reproducibility of the CH model is not a problem when examining the assimilation effect; rather, it is convenient for examining the effect of model uncertainty.



(a) OK model



(b) CH model

Fig. 3. Example of a computed instantaneous cavity interface (isosurfaces of $f_L = 0.75$). (Clark-Y 11.7%, AoA = 2°, $\sigma = 0.5$).

3.1.3. Subgrid-scale model

The one-equation dynamic SGS model (Kajishima and Nomachi, 2003) is employed. The kinetic eddy viscosity ν_{SGS} in Eq. (3) is obtained by

$$\nu_{\mathrm{SGS}} = C_\nu \Delta_\nu \sqrt{k_{\mathrm{SGS}}} \tag{8}$$

from the view point of dimensional analysis. The turbulence kinetic energy k_{SGS} is obtained from the transport equation (Okamoto and Shima, 1999)

$$\frac{\partial k_{\mathrm{SGS}}}{\partial t} + \bar{u}_j\frac{\partial k_{\mathrm{SGS}}}{\partial x_j} = -\tau_{ij}^a\bar{S}_{ij} - C_\epsilon\frac{k_{\mathrm{SGS}}^{\frac{3}{2}}}{\bar{\Delta}} - \frac{2}{\mathrm{Re}}\frac{\partial\sqrt{k_{\mathrm{SGS}}}}{\partial x_j}\frac{\partial\sqrt{k_{\mathrm{SGS}}}}{\partial x_j} + \frac{\partial}{\partial x_j}\left[\left(C_d\Delta_\nu\sqrt{k_{\mathrm{SGS}}} + \frac{1}{\mathrm{Re}}\right)\frac{\partial k_{\mathrm{SGS}}}{\partial x_j}\right], \tag{9}$$

which models the production, dissipation, and diffusion terms on the right-hand side. Here, $\tau_{ij}^a = \tau_{ij} - \frac{1}{3}\delta_{ij}\tau_{kk}$ denotes the anisotropic part of the SGS stress, and $\bar{\Delta}$ denotes the length of the grid filter. The length scale of the turbulent diffusion term Δ_ν is defined as the characteristic length with the modification to represent the behavior near the solid

wall:

$$\Delta_v = \frac{\bar{\Delta}}{1 + C_k \bar{\Delta}^2 |\bar{S}|^2 / k_{\mathrm{SGS}}}, \tag{10}$$

where $|\bar{S}| = \sqrt{2\bar{S}_{nm}\bar{S}_{nm}}$ is the magnitude of the rate-of-strain tensor $\bar{S}_{ij}$.

One characteristic feature of the one-equation dynamic SGS model is that the dynamic procedure proposed by Germano et al. (1991) and Lilly (1992) is used for the production term $-\tau^a_{ij}\bar{S}_{ij}$, the first term of the right-hand side of Eq. (9), to deal with the complexity of the flow field and the boundary condition at the solid wall: The production term is calculated by

$$-\tau^a_{ij}\bar{S}_{ij} = 2C_s\bar{\Delta}^2|\bar{S}|^3 \tag{11}$$

rather than by using the eddy viscosity model given in Eq. (8), and C_s is given by the dynamic Smagorinsky model. The model constants $C_v = 0.005$, $C_\varepsilon = 0.835$, $C_d = 0.10$ and $C_k = 0.08$ are the same as those of the LES performed by Okabayashi et al. (2023).

3.2. Numerical method

The numerical simulation of unsteady flow is based on the fractional step method with the collocated arrangement of the variables. The convection term in Eq. (3) is discretized by the Quadratic Upstream Interpolation for Convective Kinematics (QUICK) scheme, and the other terms in Eq. (3) are discretized by the central finite difference of second-order accuracy. For the time-marching method, the Adams–Bashforth method of second-order accuracy is used for both the convective and viscous terms. In the pressure equation, the three-step method is used to determine the time-derivative of the pressure, and the central difference method of second-order accuracy is used for the spatial derivatives of the other terms. The successive over-relaxation (SOR) method is used to perform the iterative computation. Furthermore, the Adams–Bashforth method of second-order accuracy is used for the time marching of the transport equation for turbulence kinetic energy k_{SGS}, the donor cell method is used for the spatial difference of the convection term for stability, and the other terms are discretized by the central difference method of second-order accuracy. The details of the time marching of the liquid volumetric fraction in the OK model are provided in Okita and Kajishima (2002), and the same time-marching method is used for the CH model.

3.3. Computational condition

The object of the analysis is the flow around a Clark-Y11.7% hydrofoil at an angle of attack of 2°, where reliable experimental data (Numachi et al., 1949; Watanabe et al., 2013) has been obtained. The chord length and the mainstream velocity are 70.0 mm and 11.0 m/s respectively. The sound speed c and kinetic viscosity ν are those of pure water at 13 °C. Accordingly, the Reynolds number and Mach number are 6.41×10^5 and 7.60×10^{-3}, respectively. The computational domain and boundary conditions are summarized in Fig. 4. For the pressure boundary conditions, the non-reflective conditions of Okita and Kajishima (2002) are used for the inflow boundary, outflow boundary, and upper and lower far regions, considering the pressure waves caused by the fluctuation in the liquid volume. The number of grid points in each direction is $N_\xi \times N_\eta \times N_\zeta = 360 \times 100 \times 100$. The grid resolution is based on Suzuki et al. (2011): In their LES of cavitation flow, although the turbulence model is different, the C-type grid and OK model as in this study were used, and their grid resolution for single-phase conditions was $(\Delta^+_\xi, \Delta^+_\eta, \Delta^+_\zeta) = (200, 25, 250)$ (minimum size for Δ^+_ξ and Δ^+_η). In addition, the reproducibility of the flow for single-phase flow was verified. In this study, based on Suzuki et al. (2011), the grid resolution is set to $(\Delta^+_\xi, \Delta^+_\eta, \Delta^+_\zeta) = (140, 3, 250)$. The resolution in the wall-normal direction, η, is considerably finer: the first grid point

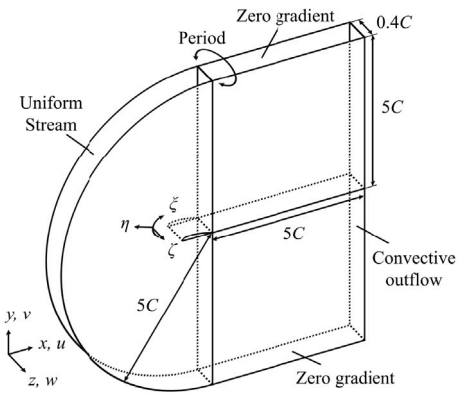


Fig. 4. Computational domain and boundary conditions (Clark-Y11.7% at AoA = 2°).

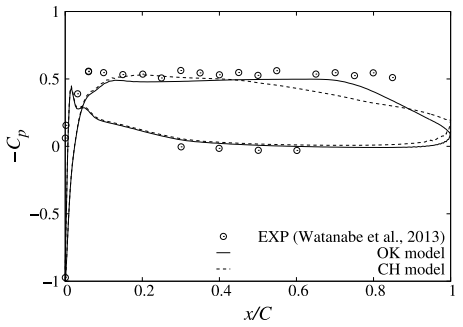


Fig. 5. Static pressure distribution.

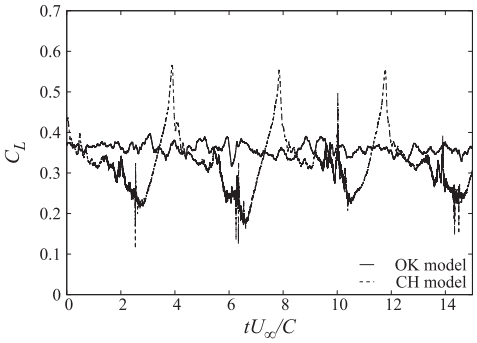


Fig. 6. Time evolution of the lift coefficient.

on the wall is in the viscous sublayer under single-phase conditions; thus, the grid resolution is within an acceptable range for imposing no-slip conditions. Nevertheless, our single-phase flow results are almost identical with Suzuki et al. (2011). Therefore, the grid resolution in this study is sufficiently fine. As the boundary layer becomes even thicker under cavitation conditions, this resolution is sufficient for cavitation conditions.

In this study, we validate the assimilation performance under highly unsteady cavitation conditions. Therefore, the cavitation number σ is set to 0.5, where the transition cavity oscillations associated with the re-entrant jet occur (Watanabe et al., 2013) and the reproducibility of cavitation phenomena differs significantly between the OK model and CH model. Figs. 5 and 6 show the static pressure distribution and the time evolution of lift coefficient at $\sigma = 0.5$, respectively.

In Fig. 5, the region where $-C_p = \sigma$ corresponds to the region where the hydrofoil surface is always covered by the cavity, and the pressure is kept at saturated vapor pressure. Thus, the position where the re-entrant jet reaches the most upstream can be estimated from x/C where

$-C_p < \sigma$. By comparison, this position is $x/C \approx 0.8$ in the experiment, whereas it is $x/C \approx 0.5$ in the CH model, revealing a large difference in quantitative accuracy. The cavitation behavior is also qualitatively very different. In the CH model, the excessive progression of the re-entrant jet results in the large cloud cavity shedding compared to the OK model. These differences also cause the lift coefficient to fluctuate significantly in the CH model (Fig. 6): Okamura and Okabayashi (2023) reported that the Strouhal number of the C_L fluctuation computed using the OK model shows good agreement with the experimental results (Watanabe et al., 2013). In Fig. 6, conversely, the period of the C_L fluctuation of the CH model is clearly different from that of the OK model. We attempted to utilize these different characteristics of the OK model and CH model: in Sections 6 and 7, we attempted to reproduce the phenomenon more accurately by using the results of the OK model as reference data for the assimilation simulation using the CH model.

4. Pseudo-measurement data setup

For the pseudo-measurement data, we assume the TPIV measurement data of a three-dimensional velocity field around the hydrofoil. Pseudo-measurement data is extracted from the database of the cavitating flow computed using the OK model. In this twin experiment, for simplicity, the pseudo-measurement positions correspond to grid points, and pseudo-measurement data are given every two grid points (at $z = 0.008C$ intervals) in the spanwise direction, i.e., the number of the pseudo-measurement cross-section is 50. The computational grid near the hydrofoil and an example of the instantaneous pseudo-measurement area and its boundary (red and green points) are shown in Fig. 7. The green points correspond to the "cavity interface" described below. The data inside the cavity is assumed to be missing because no particles enter the cavity, and no measurement data of velocity are given. Therefore, the pseudo-measurement area of each spanwise cross-section varies according to the distribution of the cavity at each moment when the pseudo-measurement data are obtained. The pseudo-cavity interface is set to isosurface of $f_L^{\text{interface}} = 0.75$ by comparison of the time-averaged liquid volumetric fraction distribution around the hydrofoil and the raw PIV data (Pervunin et al., 2021; Ilyushin et al., 2023). Although it is difficult to measure the liquid volumetric fraction f_L, the cavity interface and the liquid phase region can be estimated indirectly, given that the region where particles exist in TPIV is in the liquid phase. Therefore, as the pseudo-measurements of liquid volumetric fraction, $f_L = f_L^{\text{interface}}$, is given at the boundary of the pseudo-TPIV measurement area (green points in Fig. 7) and the liquid phase condition $f_L = 1$ is given for the entire pseudo-measurement area (red points in Fig. 7). In Section 7, as the pseudo-PIV measurement data, the cross-section at $z = 0.2C$ of pseudo-TPIV data is used. The data for physical quantities other than the $f_L^{\text{interface}}$ and 3D or 2D velocity field are not usually available despite TPIV or PIV being conducted on a cavitating flow. To represent this situation virtually, no other physical quantities are given as pseudo-measurement data in the twin experiment. The physical quantities other than the $f_L^{\text{interface}}$ and velocity field are modified dependently based on the given pseudo-measurement data. All of these pseudo-measurements are non-dimensionalized. The sampling frequency is 5000 Hz, which corresponds to 4000 time steps at time increments of 5.0×10^{-6}.

5. Overview of data assimilation

5.1. Local ensemble transform Kalman filter (LETKF)

Numerical simulation and measurement are integrated through the local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). This method is based on the local ensemble Kalman filter (LEKF) (Ott et al., 2004) that incorporates the localization algorithm (described in Section 5.2) into the EnKF and an ensemble update method for the ensemble transform Kalman filter (ETKF) (Bishop et al., 2001),

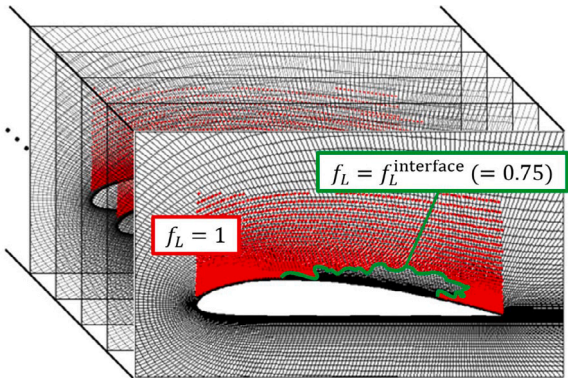


Fig. 7. Computational grid near the hydrofoil and instantaneous pseudo-TPIV measurement area (red and green points). The cross-section at $z = 0.2C$ is used as the pseudo-PIV measurement data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reducing the large error covariance matrix $\mathbf{P}$ to a realistically computationally feasible size and improving the computational efficiency. As the cavitating flow around a hydrofoil is a highly unsteady and turbulent phenomenon, the sampling error increase considerably if the number of ensemble members is not sufficient in the EnKF or ETKF. Moreover, localizing the error covariance matrix $\mathbf{P}$ is impractical because doing so requires direct computation of the error covariance matrix, which has a matrix size of approximately $O(10^{15})$ under the present computational conditions. Therefore, the LETKF is suitable for this simulation. In a conventional LETKF, localization is implemented using a local patch based on the computational grid. However, in this study, the grids are densely packed near the wall, so that the localization scale becomes smaller near the hydrofoil surface, and measurement data cannot be sufficiently assimilated. Thus, we use the LETKF without a local patch (Miyoshi et al., 2007) and implement localization based on the physical length.

In the LETKF, the analysis is performed in an m-dimensional space spanned by m ensemble members in the "local region". This local region is defined as a sphere of locarization radius R (described in Section 5.2) centered at each grid point. The analytical ensemble matrix $\mathbf{X}^a$ is represented by the sum of its ensemble mean $\hat{\mathbf{X}}^a$ and the perturbation matrix $\mathbf{E}^a$, and the ensemble update is obtained by a linear transformation $\mathbf{E}^a = \mathbf{E}^f \mathbf{W}^a$. Here, superscripts a and f denote forecast and analysis, respectively, and $\hat{}$ denotes the ensemble mean. The LETKF analysis equation is given by

$$
\begin{aligned}
\mathbf{X}^a =& \hat{\mathbf{X}}^a + \mathbf{E}^a \\
=& \left[\hat{\mathbf{X}}^f + \mathbf{K}\left(\mathbf{y}^o - H\left(\hat{\mathbf{X}}^f\right)\right)\right] + \left[\mathbf{E}^f \mathbf{W}^a\right] \\
=& \left[\hat{\mathbf{X}}^f + \mathbf{E}^f \tilde{\mathbf{P}}^a \left(\mathbf{H}\mathbf{E}^f\right)^{\top} \mathbf{R}^{-1}\left(\mathbf{y}^o - H\left(\hat{\mathbf{X}}^f\right)\right)\right] \\
&+ \left[\mathbf{E}^f \sqrt{m-1}\left(\tilde{\mathbf{P}}^a\right)^{\frac{1}{2}}\right] \\
=& \hat{\mathbf{X}}^f + \mathbf{E}^f \left[\tilde{\mathbf{P}}^a \left(\mathbf{H}\mathbf{E}^f\right)^{\top} \mathbf{R}^{-1}\left(\mathbf{y}^o - H\left(\hat{\mathbf{X}}^f\right)\right) + \sqrt{m-1}\left(\tilde{\mathbf{P}}^a\right)^{\frac{1}{2}}\right]. \qquad (12)
\end{aligned}
$$

The details of each symbol are shown in Table 2. Here, N_v and N_c are the number of variables (i.e., $N_v = 9$ as shown in Eq. (4)) and the number of grid points in a local region, respectively, and $N(= N_v N_c)$ is the dimension of the state vector. Furthermore, N_o is the number of observation points in a local region (number of dimensions of the observation space in a local region). Assuming that observations are uncorrelated, which is a common assumption, $\mathbf{R}$ is diagonal. In addition, assuming that all measurements have an uncertainty of approximately 3%, $\mathbf{R}$ is set to $9 \times 10^{-4}\mathbf{I}$. The value of $\mathbf{H}\mathbf{E}^f$ is obtained by replacing the observation operator $\mathbf{H}$ with the nonlinear operator H resulting in $\mathbf{H}\mathbf{E}^f \approx H\left(\mathbf{X}^f\right) - H\left(\hat{\mathbf{X}}^f\right)$. Considering the covariance

Table 2
List of symbols.

Symbol	Size	Meaning
$\mathbf{X}$	$N \times m$	Ensemble matrix composed of N-dimensional state vectors $\mathbf{x}^{(k)}$ $(k = 1, \dots, m)$
$\hat{\mathbf{X}}$	$N \times m$	Matrix composed of m ensemble mean of the state vectors $\hat{\mathbf{x}}$
$\mathbf{E}$	$N \times m$	Perturbation matrix constructed by subtracting $\hat{\mathbf{x}}$ from each column of $\mathbf{X}$
$\mathbf{y}^o$	N_o	Observation vector
$\mathbf{H}$	$N_o \times N$	Observation operator to transform from model space to observation space
H	$N_o \times N$	Nonlinear observation operator
$\tilde{\mathbf{P}}^a$	$m \times m$	Analysis error covariance matrix in m-dimensional space
$\mathbf{R}$	$N_o \times N_o$	Observation error covariance matrix
$\mathbf{K}$	$N \times N_o$	Kalman gain

Algorithm 1 Non-local patch Local Ensemble Transform Kalman Filter

Generate initial ensemble.
while $t < t_{\mathrm{end}}$ **do**
  Advance each ensemble independently until measurement data is obtained.
  Compute $\mathbf{X}_g^f$ and $H_g\left(\mathbf{X}_g^f\right)$.
  Compute $\hat{\mathbf{X}}_g^f$ and $H_g\left(\hat{\mathbf{X}}_g^f\right)$, and subtract these from the components of $\mathbf{X}_g^f$ and $H_g\left(\mathbf{X}_g^f\right)$ to obtain $\mathbf{E}_g^f$ and $\mathbf{H}_g\mathbf{E}_g^f$.
  for $l = 1$ to $N_{cg}(= N_\xi N_\eta N_\zeta)$ **do**
    if $N_o \neq 0$ **then**
      Select the rows of $H_g\left(\hat{\mathbf{X}}_g^f\right)$ and $\mathbf{H}_g\mathbf{E}_g^f$ corresponding to a local region around the l-th grid point, and form the N_o-dimensional vector $H\left(\hat{\mathbf{X}}^f\right)$ and the $N_o \times m$ matrix $\mathbf{HE}^f$.
      Likewise, select the corresponding rows of $\mathbf{y}_g^o$ and rows and columns of $\mathbf{R}_g$ to form the N_o-dimensional vector $\mathbf{y}^o$ and the $N_o \times N_o$ matrix $\mathbf{R}$.
      Select the rows of $\mathbf{E}_g^f$ and $\hat{\mathbf{X}}_g^f$ corresponding to the l-th grid point, and form the $N_v \times m$ matrices $\mathbf{E}^f$ and N_v-dimensional vector $\hat{\mathbf{X}}^f$.
      Compute the $m \times N_o$ matrix $\mathbf{C} = \left(\mathbf{HE}^f\right)^\top \mathbf{R}^{-1}$.
      Compute the $m \times m$ matrix $\tilde{\mathbf{P}}^a = \left[\frac{m-1}{1+\delta}\mathbf{I} + \mathbf{C}\left(\mathbf{HE}^f\right)\right]^{-1}$ with the eigenvalue decomposition in Eq. (13).
      Compute the $m \times m$ matrix $\mathbf{W}^a = \sqrt{m-1}\left(\tilde{\mathbf{P}}^a\right)^{\frac{1}{2}}$ with the eigenvalue decomposition in Eq. (13).
      Compute the m-dimensional vector $\hat{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\mathbf{C}\left(\mathbf{y}^o - H\left(\hat{\mathbf{X}}^f\right)\right)$ and add it to each column of $\mathbf{W}^a$ to form the $m \times m$ matrix $\mathbf{w}^a$.
      Multiply $\mathbf{E}^f$ by $\mathbf{w}^a$ and add $\hat{\mathbf{X}}^f$ to get the $N_v \times m$ matrix $\mathbf{X}^a$ at the l-th grid point.
    else
      $\mathbf{X}^a = \mathbf{X}^f$
    end if
  end for
  Form the $N_g(= N_v N_{cg}) \times m$ matrix $\mathbf{X}_g^a$ from $\mathbf{X}^a$ of all grid points.
end while

inflation (described in Section 6.1.1), $\tilde{\mathbf{P}}^a$ is computed by the following eigenvalue decomposition.

$$\begin{aligned}\tilde{\mathbf{P}}^a &= \left(\frac{m-1}{1+\delta}\mathbf{I} + \left(\mathbf{HE}^f\right)^\top \mathbf{R}^{-1}\left(\mathbf{HE}^f\right)\right)^{-1} \\ &= \left(\mathbf{UDU}^\top\right)^{-1} \\ &= \mathbf{UD}^{-1}\mathbf{U}^\top \end{aligned} \tag{13}$$

The parameters δ, $\mathbf{U}$ and $\mathbf{D}$ are the inflation parameter, eigenvector and matrix with eigenvalues as diagonal components, respectively. In this way, we can obtain the square root of $\tilde{\mathbf{P}}^a$ by sharing the same eigenvalue decomposition. From Eq. (12),

$$\tilde{\mathbf{P}}^a\left(\mathbf{HE}^f\right)^\top \mathbf{R}^{-1}\left(\mathbf{y}^o - H\left(\hat{\mathbf{X}}^f\right)\right) + \sqrt{m-1}\left(\tilde{\mathbf{P}}^a\right)^{\frac{1}{2}} \tag{14}$$

is computed only for variables in the observation space and does not require N-dimensional vectors or matrices. Therefore, as soon as $\mathbf{HE}^f$ and $H\left(\hat{\mathbf{X}}^f\right)$ are obtained, $\mathbf{X}^a$, $\mathbf{X}^f$, and $\mathbf{E}^f$ are redefined as $N_v \times m$ matrices using only the variables at a grid point. The detailed procedure is shown in Algorithm 1. In Algorithm 1, the subscript g denotes a vector or matrix composed of variables in the global computational domain. The variables without subscript g are defined in a local region.

5.2. Localization

Localization is a method through which the influence of observations distant from a grid point is eliminated, assuming that there is no error correlation between the distant points. When assimilation is performed with a limited number of ensemble members m, the computed correlation with distant observations is mostly due to the sampling error. Therefore, in the LETKF, the influence of distant observations is eliminated by assimilating only observations within the local region for each grid point.

Furthermore, in this study, the observation error covariance matrix $\mathbf{R}$ is localized to weight observations near a grid point and to completely suppress the sampling error when the observation is farther from the grid point. Multiplying the inverse of the localization function $L(r)$ by each element of the observation error covariance matrix leads to an infinite measurement error far away from the grid point, where r denotes the distance from the grid center normalized by the localization radius R. The localization function $L(r)$ is a fifth-order function that approximates the Gaussian function proposed by Gaspari and Cohn (1999):

$$L(r) = \begin{cases} 1 - \frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 & (r \le 1) \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1} & (1 < r \le 2) \\ 0 & (r > 2) \end{cases}. \tag{15}$$

This fifth-order function is set to 0 when $r = 2$. The sphere range represented by $L(2) = 0$ corresponds to the local region based on the physical distance of the LETKF.

By introducing this localization, assimilation can be performed accurately even with a limited number of ensemble members m.

6. Results and discussion of the twin experiment

In this section, TPIV is assumed as the pseudo-measurement data.

6.1. Parameter tuning

The above-mentioned assimilation method includes three parameters: the number of ensemble members m, the localization radius R, and the inflation parameter δ. Among them, the number of ensemble members m is the parameter related to the assimilation accuracy and computational cost, where the higher its value, the lower the sampling error. The other two parameters depend on m. As the main objective of this study is to verify the effectiveness of DA, the number of members is set to 10 to prioritize computational efficiency. Below, we determine the other parameters based on $m = 10$. Furthermore, we consider the number of ensemble members needed to achieve highly accurate DA.

6.1.1. Covariance inflation

Covariance inflation is a technique of intentionally increasing the forecast error covariance matrix $\mathbf{P}^f$ by multiplying it by $(1 + \delta)$, a number larger than 1. This technique is introduced to prevent $\mathbf{P}^f$ from becoming excessively small through repeated assimilation, making it difficult to reflect the observed information. In the LETKF, covariance inflation is introduced as in Eq. (13). In general, the inflation parameter δ is set to minimize the root mean square error (RMSE), a measure of the magnitude of forecast error. In recent years, adaptive covariance inflation has been proposed and improved to reduce the cost and improve the assimilation accuracy in the tuning process. In this study, we estimate the inflation parameter δ dynamically by using adaptive inflation based on the Gaussian approach developed by Miyoshi (2011). The only parameter in this method is the prior inflation variance v^b, which has been tuned to an optimal value of $v^b = 0.04^2$ in the Simplified Reparameterization, Primitive-Equation Dynamics model (SPEEDY), a low-resolution Atmospheric General Circulation Model (AGCM) (Miyoshi, 2011). Therefore, by assigning several values of v^b around this optimum value, we can adjust this value so that the RMSE is minimized in this problem setup.

Fig. 8 shows the time series of the ensemble mean RMSE and Spread, a measure of the spread of ensemble members, at $v^b = 0.04^2, 0.08^2$, and 0.16^2. Here, the number of ensemble members and the localization radius are set to $m = 10$ and $R = 0.015C$, respectively. The localization radius is set to a value simply estimated by the method described in Section 6.1.2. As the pseudo-measurements are located in non-uniform grids, the RSME and Spread are computed as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_c} \left(\mathrm{y}_i^o - H\left(\hat{\mathbf{x}}^f\right)_i\right)^2 S_{\mathrm{cell}\,i}}{\sum_{i=1}^{N_c} S_{\mathrm{cell}\,i}}} \tag{16}$$

and

$$\mathrm{Spread} = \sqrt{\frac{\sum_{i=1}^{N_c} \left(\frac{1}{m-1}\sum_{k=1}^{m}\left(H\left(\mathbf{x}^{f(k)}\right)_i - H\left(\hat{\mathbf{x}}^f\right)_i\right)^2\right) S_{\mathrm{cell}\,i}}{\sum_{i=1}^{N_c} S_{\mathrm{cell}\,i}}}, \tag{17}$$

taking into account the weights due to the non-uniform grids, where S_{cell} is the area of the cell where the pseudo-measurement data is located. As shown in Fig. 8, the RMSE decreases to approximately 0.02 for all the prior inflation variances. This is smaller than the set standard deviation of the observation error, which is 0.03. This may be due to a widely contained uniform flow area with almost no error from the pseudo-measurements in the measurement area. The RMSE and Spread

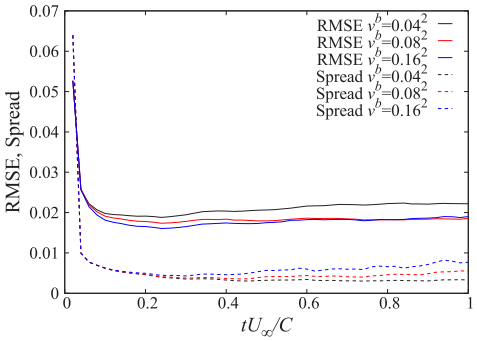


Fig. 8. Time series of the RMSE and Spread at each prior inflation variance.

should be of similar magnitude, and the RMSE being larger than Spread means that the Spread is underestimated, and the observation is not fully assimilated. Under this condition, the RMSE does not decrease to the same degree as the Spread for any of the prior inflation variances. The details are described below in Section 6.1.3.

From Fig. 8, the RMSE at $v^b = 0.04^2$ is larger than the other two and thus unsuitable. The RMSEs at 0.08^2 and 0.16^2 are roughly equal in the end, but the Spread is larger at 0.16^2, overestimating the variance of each ensemble member. Considering that a prior inflation variance that is too large can cause the inflation parameter to fluctuate noisily (Miyoshi, 2011), a prior inflation variance of approximately 0.08^2 is considered optimal in this problem setup.

6.1.2. Localization radius

The localization radius R is an important parameter that considerably affects the reproducibility of the vortical structure of the flow field. Therefore, it must be tuned to a value that reflects the physical information of the flow field. In this study, we tune the localization radius by focusing on the characteristic turbulent vortical structure simulated by the OK model.

Fig. 9 shows the distribution of v (velocity component of y-direction) simulated by the OK model. In the cavitating flow around a hydrofoil simulated by the OK model, the spanwise vorticies are generated near the inception point attributed to the baroclinic torque term, $\nabla p \times \nabla f_L$, of the vorticity equation and advect along the sheet cavity surface (Okabayashi et al., 2019, 2023). These spanwise vorticies cause rippling on the sheet cavity surface as shown in Fig. 9, and the spatial gradient of f_L generates additional spanwise vortices. The rippling interface and the spanwise vortices are maintained through this interaction, and this interaction is an essential phenomenon to be reproduced. In contrast, such spanwise vortices and interface ripples are not observed in the simulation with the CH model. Therefore, we focus on this vortical structure characteristic of the OK model and determine the localization radius based on the scale of the spanwise vortex. In this study, R is determined with reference to the spanwise vortex radius R_s in Fig. 9, as we consider that the local features of spanwise vortices can be captured as measurement information if the localization radius is smaller than the vortex radius. As the size of R_s almost does not change in time, three cases of assimilation computations with $R = 0.01C, 0.015C, 0.02C$ are attempted.

The v distributions of $R = 0.01C, 0.015C, 0.02C$ at time $tU_\infty/C = 0.2$ are shown in Fig. 10 as a representative instantaneous result. The number of ensemble members and prior inflation variance are fixed to $m = 10$ and $v^b = 0.08^2$, respectively. Fig. 10(b) shows that the reproducibility of the vortex in the region A is low when the localization radius is small. This is because the measurement information near the boundary of the measurement area cannot be sufficiently incorporated, and the unsteadiness of this region is particularly strong.

In contrast, the reproducibility of the spanwise vortices at the cavity interface (region B) does not deteriorate even when the localization

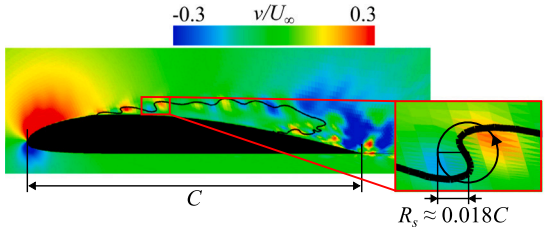


Fig. 9. Sideview of spanwise vortices represented by instantaneous contour of v/U_∞ simulated by the OK model. Black line : isoline of $f_L = f_L^{\text{interface}}$, representing the boundary of the measurement area.

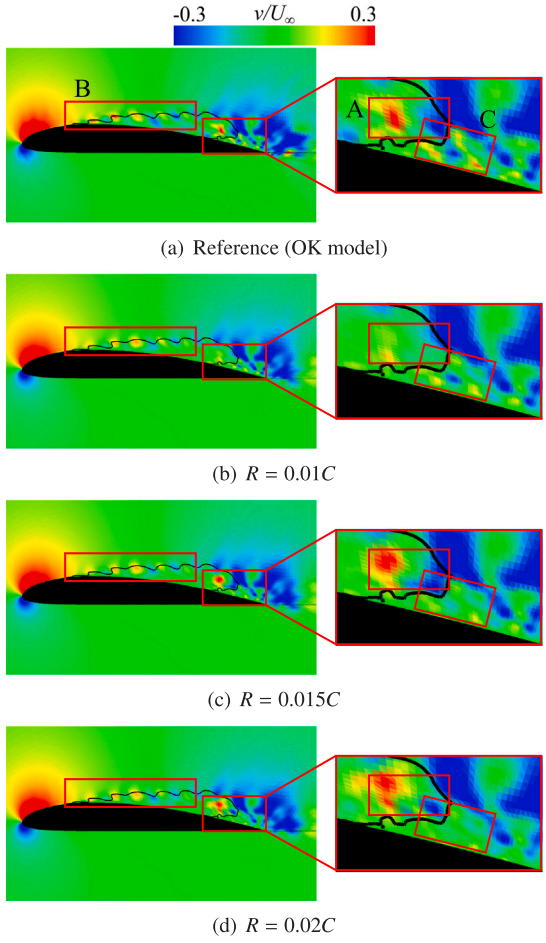


Fig. 10. Sideviews ($z = 0.2C$) of instantaneous velocity contours simulated by the CH model + DA (ensemble mean) for different localization radii. Black lines : isoline of $f_L = f_L^{\text{interface}}$ of OK model.

radius is small. As emphasized above, the essential phenomenon to be reproduced in region B is the interaction between the rippling interface and the spanwise vortices. In the present DA, this interaction is reproduced owing to three factors: the ripping interface given indirectly by the pseudo-interface, baroclinic torque term, $\nabla p \times \nabla f_L$, and pseudo-velocity field corresponding to the spanwise vortices. The pseudo-interface is considered to be the most dominant among these, because without the pseudo-measurement information of the interface ($f_L = f_L^{\text{interface}}$), the rippling interface is not reproduced well (discussed in Section 6.2). Because the pseudo-interface is thin, its effect does not depend on the size of the localization radius that incorporates it. Thus, the reproducibility of the spanwise vortices in region B in Fig. 10(b) is not affected by the localization radius.

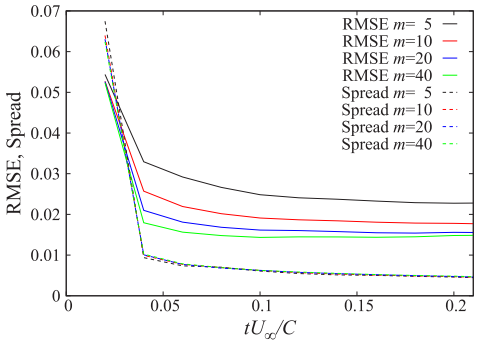


Fig. 11. Time series of the RMSE and Spread at each number of ensemble members.

In region C, it is observed that the fine-scale vortical structure becomes blurred as the localization radius increases. This is because measurement information is excessively captured in the local region at the rear of the sheet cavity, where the vortices are finer, and the flow field is smeared out.

Thus, the localization radius should be determined so that phenomena of various scales in the flow field are reproduced evenly. In this study, the localization radius is set to $R = 0.015C$ based on the above observation. However, it is not always possible to apply a constant localization radius for cavitating flow around a hydrofoil, where the vortex scale is spatially very different. Therefore, it is desirable to introduce a dynamic local radius estimation method based on the spatial scale of the vortex in the future.

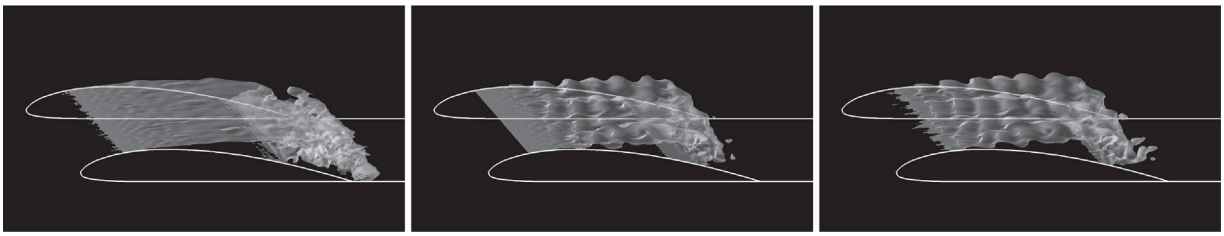
6.1.3. Number of ensemble members

In general, the larger the number of ensemble members m, the easier it is to suppress the sampling error and the more accurate the assimilation. However, in large-scale numerical simulations, increasing the number of members is a trade-off for computational resources and time, and a realistically feasible number of members must be set. In this section, we investigate the effect of the number of ensemble members on assimilation accuracy under the above tuned parameters.

The RMSE and Spread at $m = 5, 10, 20, 40$ until $tU_\infty/C = 0.2$ when the assimilation is approximately stable are shown in Fig. 11, where the localization radius and prior inflation variance are fixed as $R = 0.015C$ and $v^b = 0.08^2$, respectively. Fig. 11 shows that the RMSE is reduced by increasing the number of members, and at $tU_\infty/C = 0.2$, the RMSEs for $m = 20, 40$ are almost the same. However, even at $m = 40$, the RMSE exceeds the Spread, and the assimilation accuracy is insufficient. Therefore, other factors besides the number of members are considered as responsible for the fundamental lack of assimilation accuracy. One candidate is the localization radius. The localization radius is determined to be $R = 0.015C$ in Section 6.1.2 so that phenomena of the various scales in the flow field are reproduced evenly. However, better reproducibility of the fine-scale turbulent vortices behind the sheet cavity (Fig. 10(b) region C) is observed when the localization radius is set to R = 0.01C. Thus, the method in which the localization radius is given as a constant is not capable of reproducing multiscale vortices. The current lack of assimilation accuracy is likely to be improved by introducing a dynamic localization radius estimation method.

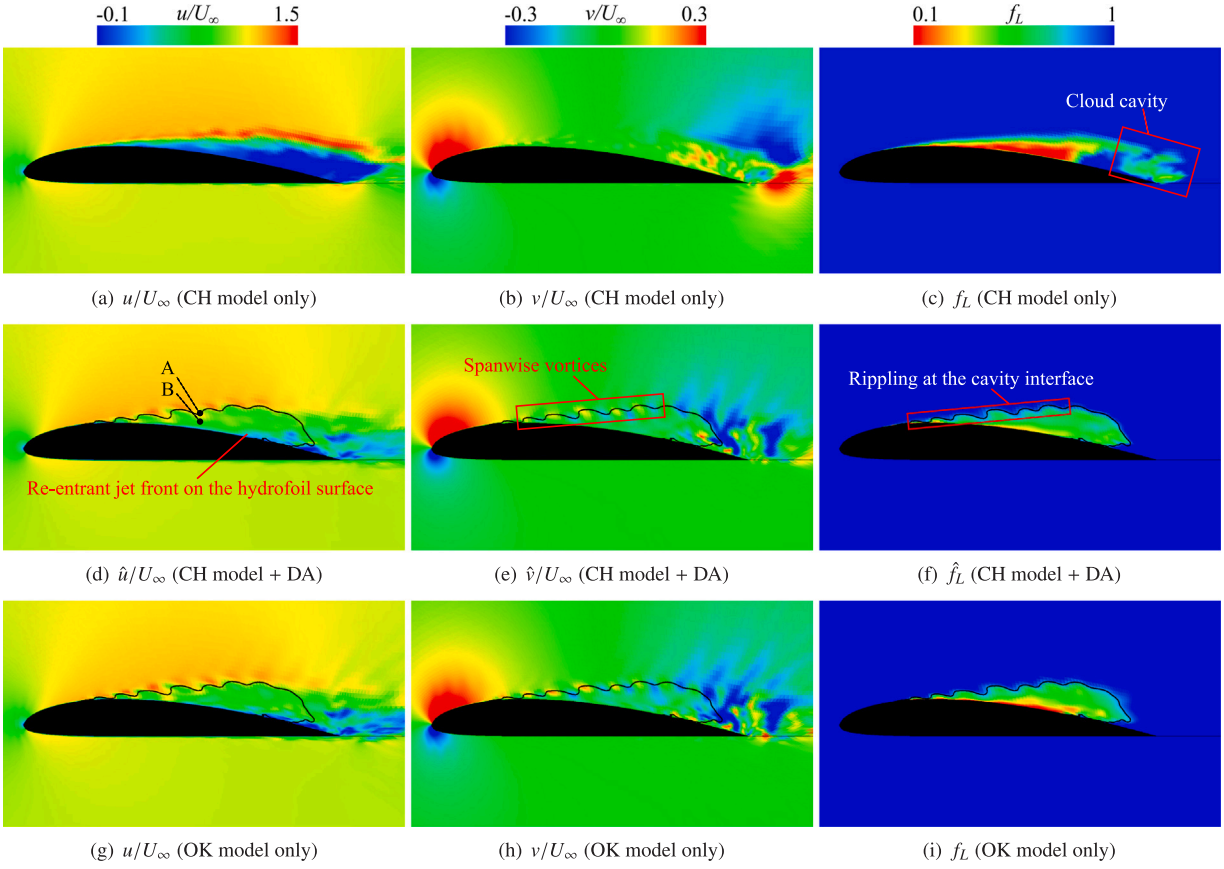
6.2. Assimilation result and discussion

Fig. 12 shows a comparison of the overview of the cavity interface (isosurface of $f_L = f_L^{\text{interface}}$) without and with assimilation, Fig. 12(a) and (b), respectively, and the pseudo-measurement data, Fig. 12(c). By applying DA to the CH model, the rippling cavity interface based on the OK model is qualitatively reproduced. This result suggests that existing cavitation models can be used to reconstruct the flow field based on



(a) CH model only (b) CH model + DA (ensemble mean) (c) OK model only (pseudo-measurement data)

Fig. 12. Overviews of the instantaneous isosurface of $f_L = f_L^{\text{interface}}$ at $tU_\infty/C = 1$. The TPIV data is assumed for the pseudo-measurement data.



(a) u/U_∞ (CH model only) (b) v/U_∞ (CH model only) (c) f_L (CH model only)

(d) $\hat{u}/U_\infty$ (CH model + DA) (e) $\hat{v}/U_\infty$ (CH model + DA) (f) $\hat{f}_L$ (CH model + DA)

(g) u/U_∞ (OK model only) (h) v/U_∞ (OK model only) (i) f_L (OK model only)

Fig. 13. Sideviews ($z = 0.2C$) of the instantaneous contour of each variable at $tU_\infty/C = 1$. (a), (b), (c) CH model only, (d), (e), (f) CH model + DA (: ensemble mean) and (g), (h), (i) OK model only (the database for creating the pseudo-measurement data). The TPIV data is assumed for the pseudo-measurement data. Black lines : isoline of $f_L = f_L^{\text{interface}}$ of OK model.

actual phenomena, provided that measurement data of the cavitation flow that can serve as reference data exist.

The ensemble mean of the u, v, and f_L distributions at $z = 0.2C$ with $tU_\infty/C = 1$ is shown in Fig. 13. The distribution of each variable is also qualitatively reproduced based on the pseudo-measurement data; namely CFD result by OK model only. In particular, important features of cavitating flow, such as the front-line of the re-entrant jet (reverse flow) above the hydrofoil surface (Figs. 13(a), (d), (g)), the spanwise vortices at the cavity interface (Figs. 13(b), (e), (h)), the rippling of the cavity interface, and the presence of cloud cavities (Figs. 13(c), (f), (i)), are well reproduced. The interface rippling is replicated despite the absence of measurements within the cavity because of the pseudo-measurement of $f_L = f_L^{\text{interface}}$ at the measurement region boundary and the interaction between the spanwise vortices and the interface rippling induced by the vortices. The large cloud cavity, which is generated in the CH model (Fig. 13(c)), no longer appears owing to assimilation, because the pseudo-measurement $f_L = 1$ is given for the entire TPIV measurement area. Thus, f_L, which is difficult to directly observe, can be used as the measurement information. The assimilation computation without the pseudo-measurement of the liquid volumetric fraction f_L is also performed in the same way. Without the pseudo-measurement information of the interface ($f_L = f_L^{\text{interface}}$), the rippling cavitation shape is not reproduced well. Furthermore, without the pseudo-measurement information of the liquid phase region ($f_L = 1$), the reproducibility of cloud cavities behind the sheet cavity is significantly impaired. There is room for discussion regarding the complementarity of the missing data when the measurement data is limited. For example, if the pseudo-interface $f_L^{\text{interface}}$ is set to more larger value, such as 0.9, the spanwise vortices will be inside the "cavity" determined by $f_L^{\text{interface}}$, and outside the measurement region. Therefore, the spanwise vortices will become difficult to reproduce by DA. However, we expect that it will be possible to obtain the characteristic velocity field near the cavity interface, even if it is only partly, using PIV. Based on this expectation, it is not the author's intention to recognize all the characteristic flows near the cavity interface as missing data.

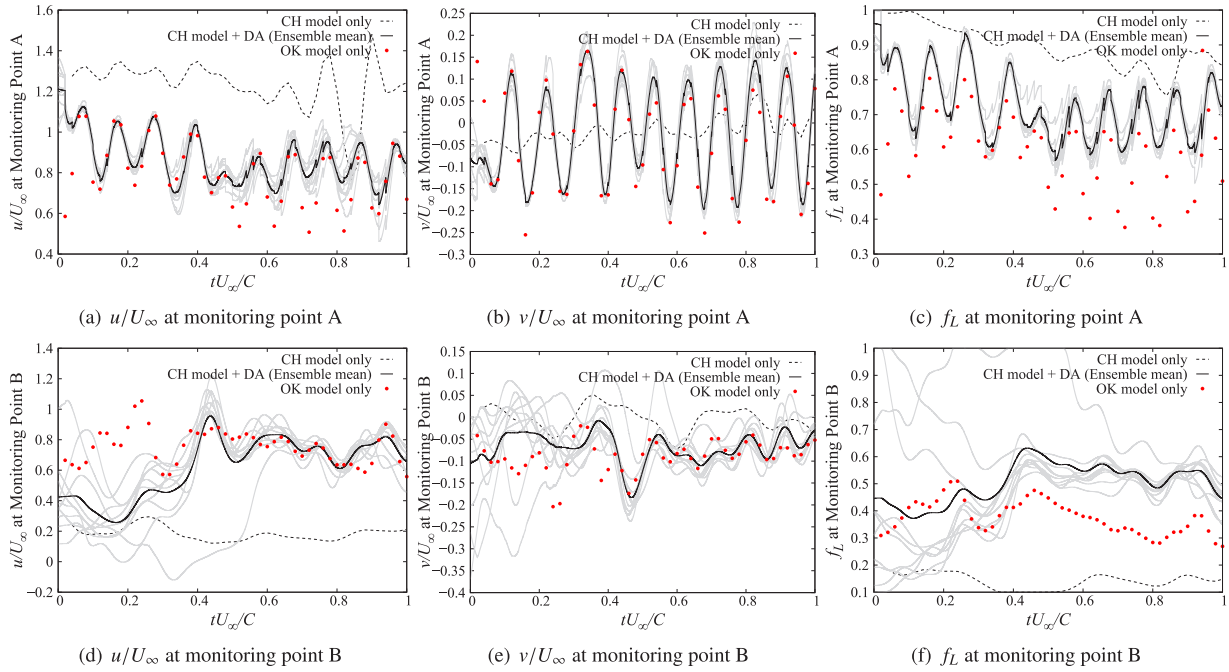


Fig. 14. Time evolutions of each variable at monitoring point A and B. TPIV data is assumed for the pseudo-measurement data. "OK model only" indicates the database for creating the pseudo-measurement data. Gray lines : time evolution for each ensemble member.

Next, we investigate the accuracy of the quantitative reproduction of the flow field. Fig. 14 shows the time evolution of u, v, and f_L at observation points A and B in Fig. 13(d). Observation points A and B are placed at the sheet cavity interface and inside the cavity, respectively, to investigate the reproducibility of the characteristic behaviors of the cavitation simulated by the OK model (i.e., rippling at the interface and low velocity region inside the cavity). For observation point A, u and v follow the OK model relatively well. In Fig. 14(a), there is a deviation from the OK model at $tU_\infty/C = 0.4 - 0.9$, this is because the observation point A is outside the pseudo-measurement area at $tU_\infty/C = 0.4 - 0.9$. This can be confirmed from the result that f_L is lower than $0.75(= f_L^{\text{interface}})$ with the OK model at $tU_\infty/C = 0.4 - 0.9$ in Fig. 14(c). For the liquid volumetric fraction f_L, the period and phase of the fluctuation agree with those of the OK model, but the assimilated values are generally toward the liquid phase side, i.e., $f_L = 1$. This is because many pseudo-measurements with $f_L = 1$ are included in the local region at grid points near the interface. However, it is very difficult to obtain a dense spatial distribution of f_L near the interface in actual measurements. Therefore, it is important that the period and the phase of fluctuation and the spatial distribution of f_L can be only qualitatively represented using only existing measurement methods.

For observation point B, the region inside the sheet cavity where the pseudo-measurement data are not given is not assimilated sequentially, and each ensemble member evolves independently (Figs. 14(d), (e), (f)). This result indicates that there is almost no discontinuity. Discontinuities are observed in Figs. 14(a), (b), (c), in the time evolution of each ensemble member: the discontinuity occurs when the correction of the flow field by assimilation is performed. At observation point B, the correction of the flow field is limited to the region around the grid points where pseudo-measurements are obtained, owing to the localization effect of the LETKF. However, as the computation progresses, the values of each ensemble member at observation point B converge, especially for u and v, showing a trend close to that of the database for creating the pseudo-measurement data (OK model only). This is because the flow around the cavity is assimilated and affects the flow inside the cavity located downstream of the assimilated flow. This suggests the existence of a complementary effect, as described in Kato and Obayashi (2011), where the assimilation of the upstream flow field improves the reproducibility of the downstream flow field. For the liquid volume fraction f_L, the values of each ensemble member gradually converge as in the case of u and v, but the time evolution after convergence is different from the trend for the OK model. This is because f_L inside the cavity, which is not directly assimilated, is simulated according to the cavitation model used. As the CH model is used as the cavitation model, the behavior of the CH model is strongly represented. Thus, the complementary effect described above depends on the accuracy of the cavitation model used with the assimilation computations.

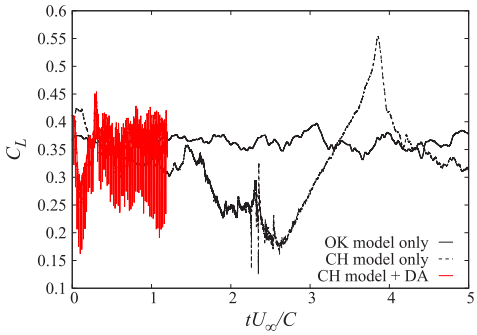
Fig. 15 shows the time evolution of the lift coefficient. After assimilation, the region covered by the cavity on the hydrofoil surface is comparable to that of the OK model, and the lift coefficient tends to be roughly similar with that of the OK model. A spike in the lift occurs immediately after assimilation, indicating that the pressure field is disturbed. However, as the iteration progresses, the assimilated flow field develops and the spike disappears.

Finally, we investigate numerical conservation in the flow field after assimilation. We define

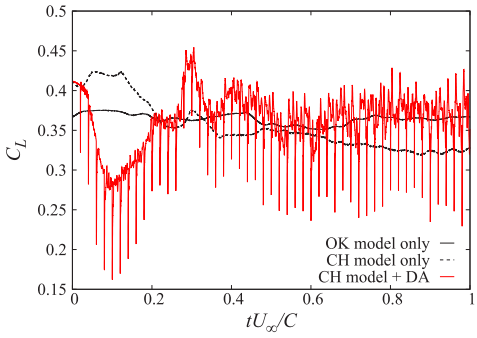
$$\text{Con.} = \frac{Df_L}{Dt} + f_L\left(\mathrm{M}^2\frac{D\bar{p}}{Dt} + \frac{\partial \bar{u}_i}{\partial x_i}\right) \tag{18}$$

using Eq. (2). The pressure equation based on Con. = 0 is computed iteratively by CFD. Therefore, the numerical conservation in this assimilation computation can be evaluated by Con. after the iterative computation.

Fig. 16 shows the time series of the spatial mean and maximum values of Con. for an ensemble member. The value of Con. after assimilation is within the order of 1 for both the mean and maximum values, compared to the flow field before assimilation (red circle in Fig. 16). The reason for the increase in Con. is the above-mentioned pressure field. Immediately after assimilation, errors in the conservation of the velocity field accumulate, and the pressure convergence calculation is performed based on that velocity field, which gradually disturbs the pressure field. Therefore, to reduce Con. further, it is desirable to introduce a method that properly corresponds the measurement data to the numerical simulation, for example, by post-processing the velocity field immediately after assimilation.



(a) Approximately one period of cavity oscillation



(b) Enlarged figure of $tU_\infty/C = 0 - 1$

Fig. 15. Time evolution of the lift coefficient. "CH model + DA" is the result of an ensemble member. The TPIV data is assumed for the pseudo-measurement data.

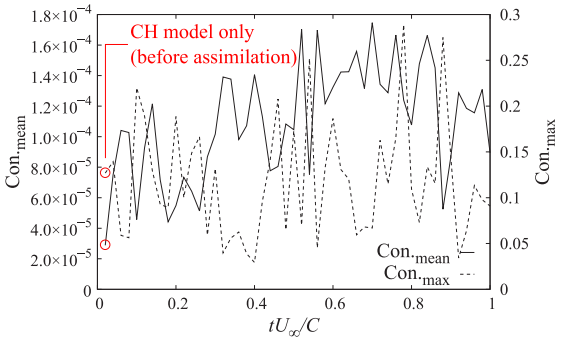


Fig. 16. Time series of the spatial mean and maximum values of Con. for an ensemble member. The TPIV data is assumed for the pseudo-measurement data.

7. Influence of spatial dimension of pseudo-measurement data

In Section 6.2, TPIV data is assumed for the pseudo-measurement data, but in most situations, the available data is PIV data, which is relatively easier to measure than TPIV data. Therefore, in this section, we investigate the assimilation performance of the pseudo-measurement data, assuming a two-dimensional velocity field, obtained by PIV under the same problem setup. In this case, the irradiated cross section of the laser sheet is assumed to be set to $z = 0.2C$: the cross-section at $z = 0.2C$ of Fig. 7 is used as the pseudo-PIV data.

Fig. 17 shows an overview of the cavity interface (isosurface of $f_L = f_L^{\mathrm{isosurface}}$) assimilated by the pseudo-PIV data. The assimilated region is limited to the width of the local region centered at $z = 0.2C$, which is assumed to be the cross-section where the laser sheet is irradiated, owing to the localization effect of the LETKF. The ensemble means of the u, v, and f_L distributions at $z = 0.2C$ are shown in Fig. 18. The result assimilating the pseudo-PIV data reproduce the flow field of the OK model qualitatively, and do not differ significantly from the ensemble mean shown in Figs. 13(d), (e), (f).

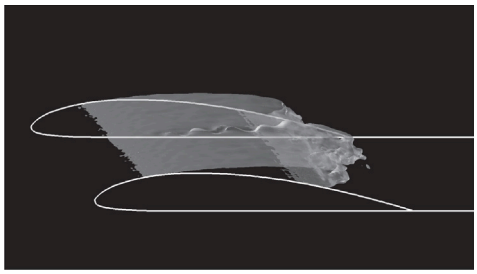


Fig. 17. Overview of the instantaneous isosurface of $f_L = f_L^{\mathrm{interface}}$ at $tU_\infty/C = 1$ (CH model + DA (ensemble mean)). The PIV data is assumed for the pseudo-measurement data.

Next, we investigate the accuracy of the quantitative reproducibility of the flow field. The time evolution of u, v, and f_L at observation points A and B (same as in Fig. 13(d)) are shown in Fig. 19. For observation point A, as in the case of the pseudo-TPIV measurements, the variables generally follow the results of the OK model, and the liquid volumetric fraction f_L also generally shifts toward the liquid-phase side ($f_L = 1$), which is the same trend shown in Fig. 14(c). The spread of each ensemble member is slightly higher compared to those in Figs. 14(a), (b), (c). In contrast, for observation point B, each ensemble member evolves individually depending on the cavitation model, i.e., CH model, as shown in Figs. 14(d), (e), (f), and the respective values do not converge, which means that the complementary effect generated by the assimilation around the cavities located upstream is not obtained inside the cavities. Thus, the distribution of each variable inside the cavities is significantly different for each ensemble member. This is because the effect of the initial flow field, which is independent for each ensemble member, persists in the region beyond $z = 0.2C$ to be assimilated. The region inside the cavity at $z = 0.2C$ is strongly affected by changes in the flow in the mainstream direction, such as the development of the sheet cavity and the progression of the re-entrant jet (Fig. 20). The increased spread of each ensemble member at observation point A is presumably due to the same event.

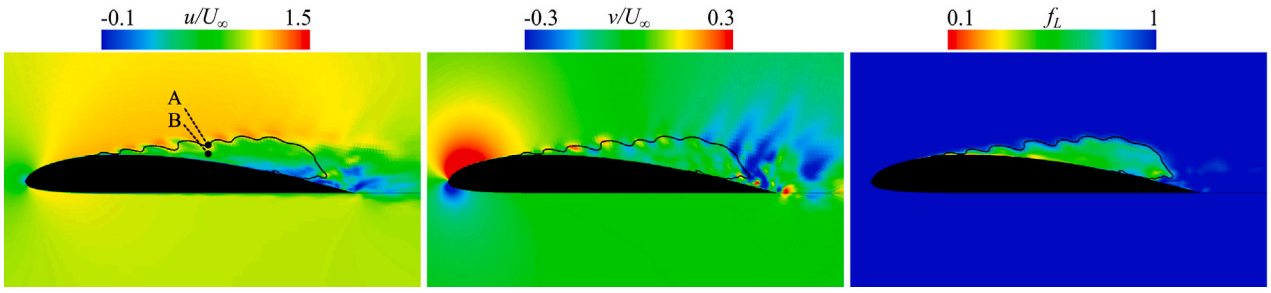
In conclusion, it is found that the flow field in the unmeasured region is not assimilated by the 2D-PIV data in three-dimensional numerical simulations, and the influence of the remaining initial flow field in each ensemble member has a negative impact on the assimilation performance of the measured cross section. The effect is particularly large for missing data completion inside cavities, indicating the need to align the dimensions of the numerical simulation and measurement when completing missing data.

8. Assimilation verification for real measurement data

We have validated DA for cavitating flow around a hydrofoil through the twin experiment. In this section, we validate the applicability of DA to real measurement data. As we do not have measurement data for cavitating flow around a hydrofoil yet, we use the publicly available 2D-PIV data set (Nonomura et al., 2021) for this validation.

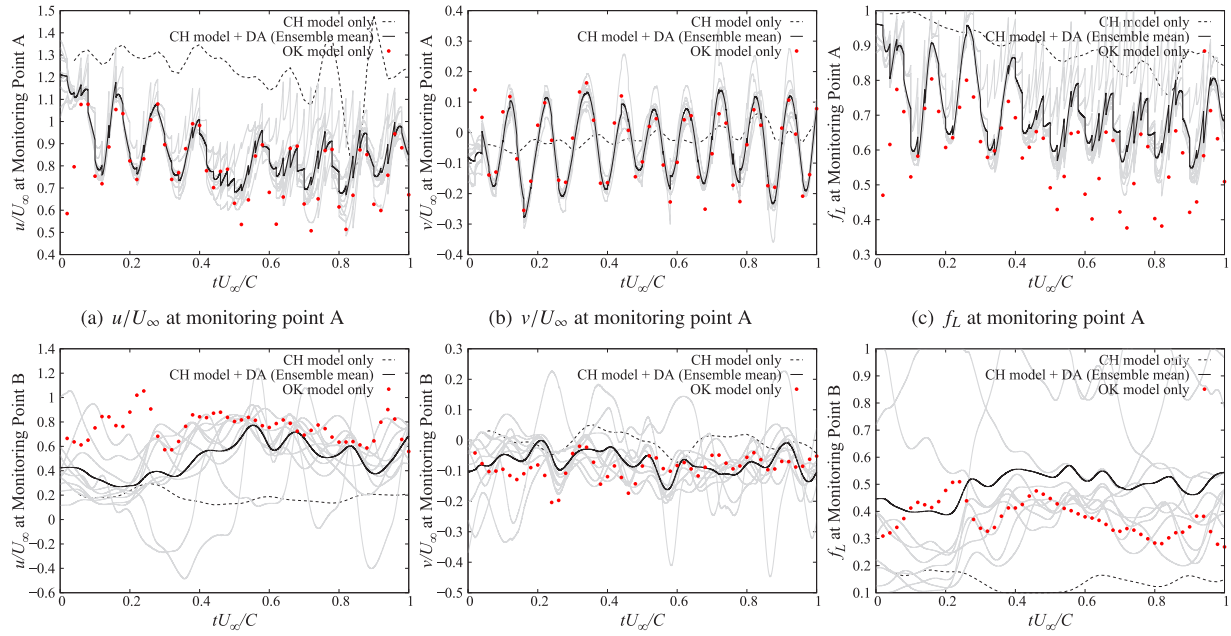
8.1. Overview of CFD

This simulation is a validation for the assimilation calculation for the real measurement data of the cavitating flow. Therefore, the governing equation, numerical model, and numerical method are the same as those in Sections 3.1 and 3.2, and the single-phase flow simulation is achieved by setting the cavitation number to 100 so that $f_L = 1$ at all times.



(a) $\hat{u}/U_\infty$ (CH model + DA) (b) $\hat{v}/U_\infty$ (CH model + DA) (c) $\hat{f}_L$ (CH model + DA)

Fig. 18. Sideviews ($z = 0.2C$) of the instantaneous contour of each variable at $tU_\infty/C = 1$ (CH model + DA (ˆ: ensemble mean)). The PIV data is assumed for the pseudo-measurement data. Black lines : isoline of $f_L = f_L^{\text{interface}}$ of OK model.



(a) u/U_∞ at monitoring point A (b) v/U_∞ at monitoring point A (c) f_L at monitoring point A

(d) u/U_∞ at monitoring point B (e) v/U_∞ at monitoring point B (f) f_L at monitoring point B

Fig. 19. Time evolutions of each variable at monitoring points A and B. PIV data is assumed for the pseudo-measurement data. "OK model only" indicates the database for creating the pseudo-measurement data. Gray lines : time evolution for each ensemble member.

8.2. Experimental setup and computational condition

The object of the analysis is the single-phase flow around a NACA0 015 airfoil. The chord length, the spanwise length, the angle of attack, the mainstream velocity, and the Reynolds number are set to 100.0 mm, 300.0 mm, 18°, 10 m/s and 6.4×10^4, respectively. The sampling frequency is 5000 Hz, which is the same as the problem setup for the twin experiment in Section 4. These setups are the same as those of the experiments conducted by Nonomura et al. (2021). Based on the experimental conditions, the numerical conditions are determined. As there is no description of the fluid, the fluid is assumed to be air at a temperature of 20°C, and each physical property is also set accordingly. Therefore, the Mach number is set to 2.91×10^{-2}. The computational domain and boundary conditions are summarized in Fig. 21. The number of grid points in each direction is $N_\xi \times N_\eta \times N_\zeta = 512 \times 200 \times 100$. The spanwise length of the computational domain is the same as in Fig. 4, and the irradiated section of the laser sheet is set to $z = 0.2C$.

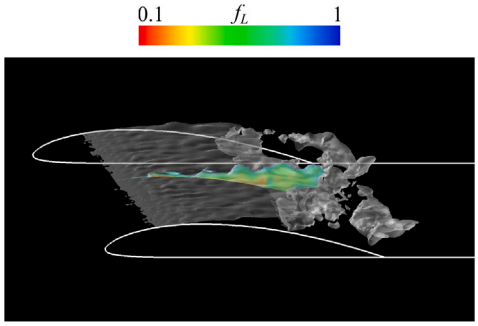
8.3. Overview of data assimilation

We validate this data assimilation program for the real measurement data using the assimilation method described in Section 5. The difference between the assimilation method and the twin experiment is that the measurement data are not arranged in grid coordinates (Fig. 22). Therefore, the calculation of the nonlinear operator H when using real measurement data corresponds to interpolating variables placed on the grid coordinates to the measurement coordinates. In this study, the inverse distance weighting method is used to perform this interpolation calculation. The observation error covariance matrix is set to $\mathbf{R} = 9 \times 10^{-4}\mathbf{I}$ as in the twin experiment.

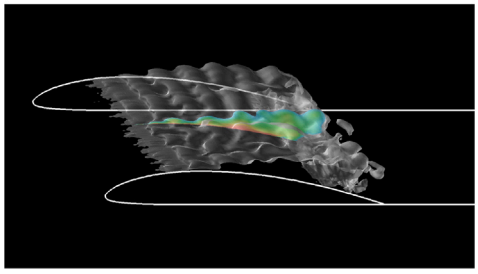
The number of members and the prior inflation variance are set to $m = 10$ and $v^b = 0.08^2$, respectively. The localization radius is determined based on the spatial scale of the vortex and the reproducibility of the flow field, as is the case for the localization radius in cavitating flow around a hydrofoil. Under this experimental condition, $R = 0.02C$ is set based on the fact that the minimum radius of the observed separated vortex is approximately $0.025C$ (figure omitted). However, no detailed tuning of these values is conducted as described in Section 6.1.

8.4. Assimilation result

The assimilation computations are conducted until $tU_\infty/C = 0.2$, where the assimilation is stable in the twin experiment, and the reproducibility of the qualitative flow field is evaluated. Because of the short computation time, no observation points are set in this validation.



(a) CH model + DA (an ensemble member)



(b) OK model only

Fig. 20. Overview of the instantaneous isosurface of $f_L = f_L^{\text{interface}}$ and contour of f_L. PIV data is assumed for the pseudo-measurement data. "OK model only" indicates the database for creating the pseudo-measurement data. The region inside the cavity at $z = 0.2C$ is strongly affected by the progression of the re-entrant jet in the region away from the $z = 0.2C$.

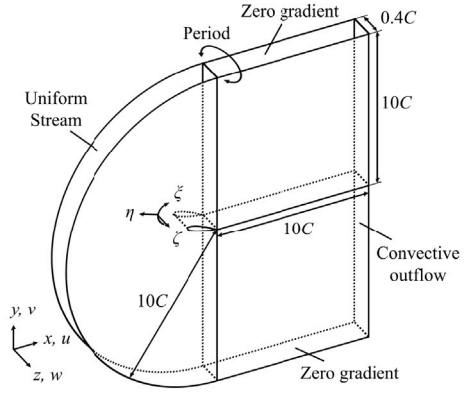


Fig. 21. Computational domain and boundary conditions (NACA0015 at AoA = 18°).

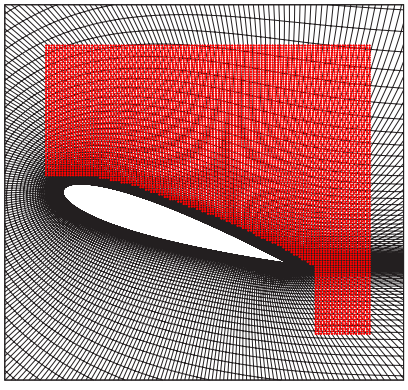


Fig. 22. Computational grid near the airfoil and measurement region at $z = 0.2C$.

Fig. 23 shows the ensemble mean of the u and v distribution at $tU_\infty/C = 0.2$. Through assimilation, the area of the reverse flow region, which is overestimated by CFD, reduces (Fig. 23(c)), and the reproducibility of the large-scale separation vortex in the flow behind the airfoil is greatly improved (Fig. 23(d)). The PIV data contains noise in the entire flow field, but the time evolution of the flow field associated with the assimilation computation reduces the noise and reproduces a natural flow field. Furthermore, even in areas where reliable measurement data could not be obtained, such as the vicinity of the airfoil surface and behind the measurement area, CFD can be used to generate additional data. Thus, the complementarity between measurement data and numerical simulation, which is expected when DA methods are employed, is confirmed in this assimilation program.

In conclusion, we have established a framework that is capable of conducting assimilation calculations when PIV or TPIV data in a cavitating flow field are available.

9. Conclusion

A twin experiment of data assimilation (DA) using the local ensemble transform Kalman filter (LETKF) is conducted for cavitating flow around a Clark-Y 11.7% hydrofoil. As pseudo-measurement data which can be obtained by existing measurement techniques, particle image velocimetry (PIV) and tomographic PIV (TPIV), which are intensively missing data inside the cavity, are assumed. The pseudo-measurement velocity field and the liquid volumetric fraction are obtained from the simulation by the source-term homogeneous fluid model developed by Okita and Kajishima (2002) (OK model).

As parameters related to LETKF, the best results are obtained for a localization radius of $R = 0.015$ and a prior inflation variance of $v^b = 0.08^2$ for ensemble members of 10. However, the localization radius should be given dynamically rather than as a constant to improve the quantitative assimilation accuracy, as the vortex scale varies widely and spatially in cavitating flow around a hydrofoil.

With the optimized parameters above, the flow field based on the reference data is qualitatively and quantitatively reproduced by assimilating the measurement data assuming TPIV. The missing data inside the cavity are complemented by computational fluid dynamics (CFD) and gradually converge to the pseudo-measurement data. This is because the flow around the cavity is assimilated and influences the flow inside the downstream cavity. However, the degree to which the missing data are complemented depends on the accuracy of the numerical simulation and model used with the assimilation. In addition, when assimilating the measurement data assuming 2D-PIV, the flow field around the cavity is assimilated in the laser sheet cross-section, but no assimilation effect is obtained inside the cavity. This is due to the inconsistency between the dimensions of the numerical simulation and the measurement data, and these dimensions need to be aligned if missing data is to be complemented.

The DA program is applied using real measurement data (2D-PIV) of single-phase flow around a NACA0015 airfoil. The flow field based on the real measurement data is qualitatively reproduced. Furthermore, the assimilated flow field reduces the noise in the real measurement data and fills in data gaps in the measurement data. Therefore, this DA program demonstrates the complementarity between the measurement data and the numerical simulation.

Through this study, we show that DA for cavitating flow is effective using data that can be obtained with existing measurement techniques. Furthermore, we construct a framework for conducting assimilation simulations as soon as PIV or TPIV data are available for the cavitating flow field. In the future, it will be necessary to further investigate methods for estimating the localization radius dynamically and improving the pressure oscillation immediately after assimilation. This is crucial for achieving precise reproduction of the flow field by using the actual measurement data of cavitating flows as training data for machine learning models.

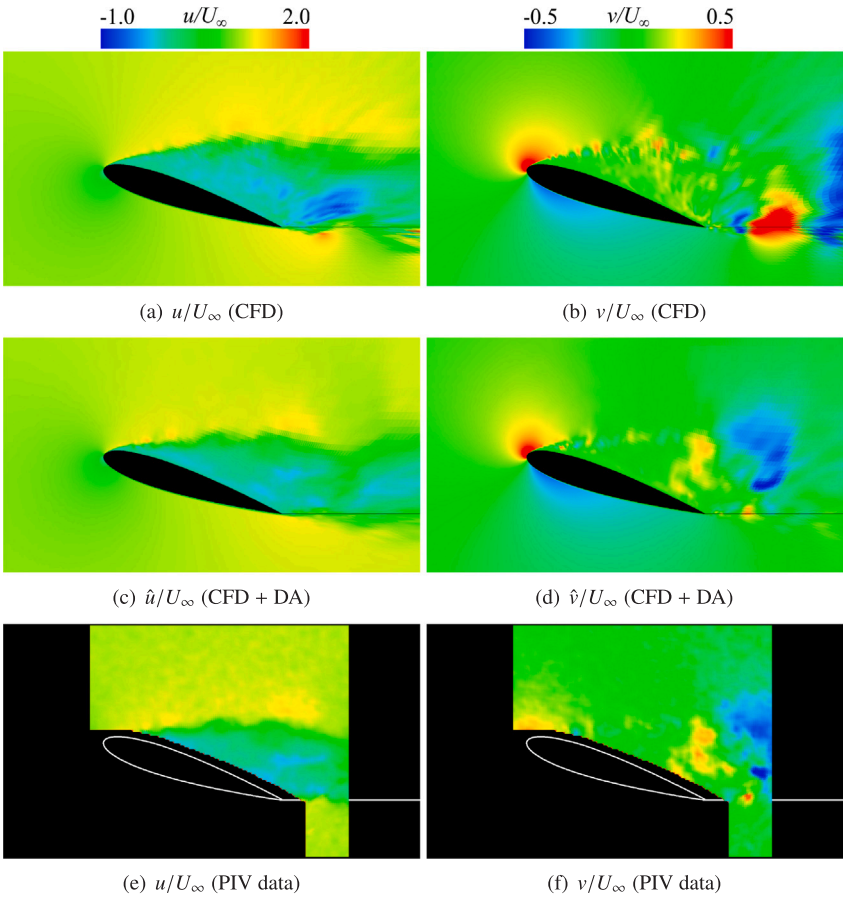


Fig. 23. Sideviews ($z = 0.2C$) of the instantaneous contour of each variable at $tU_\infty/C = 0.2$. (a), (b) CFD, (c), (d) CFD + DA ($\hat{}$: ensemble mean) and (e), (f) PIV data.

CRediT authorship contribution statement

Shungo Okamura: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Kie Okabayashi:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kie Okabayashi reports financial support was provided by Japan Society for the Promotion of Science. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by JSPS KAKENHI Grant No. JP22K03925. This work was partly achieved through the Research Proposal-based Use of the Large-scale Computer System at the Cybermedia Center, Osaka University.

Data availability

Data will be made available on request.

References

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Weather Rev. 129 (3), 420–436. http://dx.doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2021. Combining data assimilation and machine learning to infer unresolved scale parametrization. Phil. Trans. R. Soc. A 379 (2194), 20200086. http://dx.doi.org/10.1098/rsta.2020.0086.

Chen, Y., Heister, S.D., 1995. Two-phase modeling of cavitated flows. Comput. & Fluids 24 (7), 799–809. http://dx.doi.org/10.1016/0045-7930(95)00017-7.

Coutier-Delgosha, O., Reboud, J.L., Delannoy, Y., 2003. Numerical simulation of the unsteady behaviour of cavitating flows. International Journal for Numerical Methods in Fluids 42 (5), 527–548. http://dx.doi.org/10.1002/fld.530.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. http://dx.doi.org/10.1002/qj.828.

Delannoy, Y., Kueny, J.L., 1990. Two phase flow approach in unsteady cavitation modelling. Proc. Cavitation Multiph. Flow Forum, ASME- FED 98, 153–158.

Deng, Z., He, C., Wen, X., Liu, Y., 2018. Recovering turbulent flow field from local quantity measurement: turbulence modeling using ensemble-Kalman-filter-based data assimilation. J. Vis. 21, 1043–1063. http://dx.doi.org/10.1007/s12650-018-0508-0.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Research: Ocean. 99 (C5), 10143–10162. http://dx.doi.org/10.1029/94JC00572.

Fenjan, S.A., Bonakdari, H., Gholami, A., Akhtari, A.A., 2016. Flow variables prediction using experimental, computational fluid dynamic and artificial neural network models in a sharp bend. Int. J. Eng. 29 (1), 14–22.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. 125 (554), 723–757. http://dx.doi.org/10.1002/qj.49712555417.

Germano, M., Piomelli, U., Moin, P., Cabot, W.H., 1991. A dynamic subgrid-scale eddy viscosity model. Phys. Fluids A 3 (7), 1760–1765. http://dx.doi.org/10.1063/1.857955.

Honda, T., Miyoshi, T., Lien, G.-Y., Nishizawa, S., Yoshida, R., Adachi, S.A., Terasaki, K., Okamoto, K., Tomita, H., Bessho, K., 2018. Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). Mon. Weather Rev. 146 (1), 213–229. http://dx.doi.org/10.1175/MWR-D-16-0357.1.

Hsiao, C.T., Chahine, G.L., 2005. Scaling of tip vortex cavitation inception noise with a bubble dynamics model accounting for nuclei size distribution. J. Fluids Eng. 127 (1), 55–65. http://dx.doi.org/10.1115/1.1852476.

Hsiao, C.T., Chahine, G.L., Liu, H.L., 2003. Scaling effect on prediction of cavitation inception in a line vortex flow. J. Fluids Eng. 125 (1), 53–60. http://dx.doi.org/10.1115/1.1521956.

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Phys. D: Nonlinear Phenom. 230 (1–2), 112–126. http://dx.doi.org/10.1016/j.physd.2006.11.008.

Ilyushin, B.B., Timoshevskiy, M.V., Pervunin, K.S., 2023. Vapor concentration and bimodal distributions of turbulent fluctuations in cavitating flow around a hydrofoil. Int. J. Heat Fluid Flow 103 (109197), http://dx.doi.org/10.1016/j.ijheatfluidflow.2023.109197.

Inagaki, M., Murata, O., Abe, K., Kondoh, T., 2000. Numerical prediction of fluid-resonant oscillations at low mach number. Trans. Jpn. Soc. Mech. Eng. Ser. B 66 (649), 2274–2281. http://dx.doi.org/10.1299/kikaib.66.649_2274.

Ito, Y., Sawasaki, K., Tani, N., Nagasaki, T., Nagashima, T., 2005. A blowdown cryogenic cavitation tunnel and CFD treatment for flow visualization around a foil. J. Therm. Sci. 14, 346–351. http://dx.doi.org/10.1007/s11630-005-0056-5.

Kajishima, T., Nomachi, T., 2003. One-equation subgrid-scale model with dynamically determined energy production term. Trans. Jpn. Soc. Mech. Eng. Ser. B 69 (685), 1996–2001. http://dx.doi.org/10.1299/kikaib.69.1996.

Kato, C., 2011. Industry-university collaborative project on numerical predictions of cavitating flows in hydraulic machinery: Part 1–Benchmark test on cavitating hydrofoils. In: Fluids Engineering Division Summer Meeting, vol. 44403, pp. 445–453. http://dx.doi.org/10.1115/AJK2011-06084.

Kato, H., Obayashi, S., 2011. Integration of CFD and wind tunnel by data assimilation. J. Fluid Sci. Technol. 6 (5), 717–728. http://dx.doi.org/10.1299/jfst.6.717.

Kato, H., Yoshizawa, A., Ueno, G., Obayashi, S., 2015. A data assimilation methodology for reconstructing turbulent flows around aircraft. J. Comput. Phys. 283, 559–581. http://dx.doi.org/10.1016/j.jcp.2014.12.013.

Kubota, A., Kato, H., Yamaguchi, H., 1992. A new modelling of cavitating flows: a numerical study of unsteady cavitation on a hydrofoil section. J. Fluid Mech. 240, 59–96. http://dx.doi.org/10.1017/S002211209200003X.

Kunz, R.F., Boger, D.A., Stinebring, D.R., Chyczewski, T.S., Lindau, J.W., Gibeling, H.J., Venkateswaran, S., Govindan, T.R., 2000. A preconditioned Navier–Stokes method for two-phase flows with application to cavitation prediction. Comput. & Fluids 29 (8), 849–875. http://dx.doi.org/10.1016/S0045-7930(99)00039-0.

Labahn, J.W., Wu, H., Harris, S.R., Coriton, B., Frank, J.H., Ihme, M., 2020. Ensemble Kalman filter for assimilating experimental data into large-eddy simulations of turbulent flows. Flow, Turbul. Combust. 104, 861–893. http://dx.doi.org/10.1007/s10494-019-00093-1.

Lilly, D.K., 1992. A proposed modification of the Germano subgrid-scale closure method. Phys. Fluids A 4 (3), 633–635. http://dx.doi.org/10.1063/1.858280.

Ling, J., Kurzawski, A., Templeton, J., 2016. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. J. Fluid Mech. 807 (25), 155–166. http://dx.doi.org/10.1017/jfm.2016.615.

Merkle, C.L., Feng, J., Buelow, P.E.O., 1998. Computational modeling of dynamics of sheet cavitation. In: Proc. 3rd International Symposium on Cavitation. CAV1998, pp. 307–311.

Misaka, T., Obayashi, S., 2014. Sensitivity analysis of unsteady flow fields and impact of measurement strategy. Math. Probl. Eng. 2014, 1–12. http://dx.doi.org/10.1155/2014/359606.

Miyoshi, T., 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Mon. Weather Rev. 139 (5), 1519–1535. http://dx.doi.org/10.1175/2010MWR3570.1.

Miyoshi, T., Yamane, S., Enomoto, T., 2007. Localizing the error covariance by physical distances within a local ensemble transform Kalman filter (LETKF). Sola 3, 89–92. http://dx.doi.org/10.2151/sola.2007-023.

Mons, V., Chassaing, J.-C., Gomez, T., Sagaut, P., 2016. Reconstruction of unsteady viscous flows using data assimilation schemes. J. Comput. Phys. 316, 255–280. http://dx.doi.org/10.1016/j.jcp.2016.04.022.

Noda, T., Okabayashi, K., 2023. Development of framework of data-driven cavitation model using CFD data. Jpn. J. Multiph. Flow 37 (1), 94–102. http://dx.doi.org/10.3811/jjmf.2023.010.

Nonomura, T., Nankai, K., Iwasaki, Y., Komuro, A., Asai, K., 2021. Quantitative evaluation of predictability of linear reduced-order model based on particle-image-velocimetry data of separated flow field around airfoil. Exp. Fluids 62 (5), 112. http://dx.doi.org/10.1007/s00348-021-03205-8.

Numachi, F., Tsunoda, K., Senda, I., 1949. Cavitation tests on six profiles for blade elements. Rep. Inst. High Speed Mech. 1 (1), 1–16.

Okabayashi, K., Furukawa, A., Otsu, T., Kajishima, T., 2023. Large eddy simulation study on the relationship between large-scale turbulence vortices and unsteady cavitation around a hydrofoil. J. Fluids Eng. 145 (4), 041204. http://dx.doi.org/10.1115/1.4056525.

Okabayashi, K., Inaoka, T., Luo, W., Kajishima, T., 2019. The unsteady simulation for turbulent cavitating flow around a hydrofoil at a low angle of attack. Trans. Jpn. Soc. Mech. Eng. 85 (876), 19–00124. http://dx.doi.org/10.1299/transjsme.19-00124.

Okamoto, M., Shima, N., 1999. Investigation for the one-eaquation-type subgrid model with eddy-viscosity expression including the shear-damping effect. JSME Int. J. Ser. B Fluids Therm. Eng. 42 (2), 154–161. http://dx.doi.org/10.1299/jsmeb.42.154.

Okamura, S., Okabayashi, K., 2023. Development of a cavitation model tracking the sheet cavity interface and investigation of its impact on the quantitative accuracy of hydrofoil performance. Jpn. J. Multiph. Flow 37 (2), 234–242. http://dx.doi.org/10.3811/jjmf.2023.018.

Okita, K., Kajishima, T., 2002. Numerical simulation of unsteady cavitating flows around a hydrofoil. Trans. Jpn. Soc. Mech. Eng. Ser. B 68 (667), 637–644. http://dx.doi.org/10.1299/kikaib.68.637.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D., Yorke, J.A., 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus A: Dyn. Meteorol. Ocean. 56 (5), 415–428. http://dx.doi.org/10.3402/tellusa.v56i5.14462.

Pervunin, K.S., Timoshevskiy, M.V., Ilyushin, B.B., 2021. Distribution of probability of the vapor phase occurrence in a cavitating flow based on the concentration of PIV tracers in liquid. Exp. Fluids 62 (247), 1–12. http://dx.doi.org/10.1007/s00348-021-03344-y.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., Simmons, A., 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Q. J. R. Meteorol. Soc. 126 (564), 1143–1170. http://dx.doi.org/10.1002/qj.49712656415.

Schnerr, G.H., Sauer, J., 2001. Physical and numerical modeling of unsteady cavitation dynamics. In: Proc. 4th International Conference on Multiphase Flow, vol. 1, pp. 1–12.

Schwartz, R.J., Fleming, G.A., 2007. Virtual diagnostics interface: Real time comparison of experimental data and CFD predictions for a NASA Ares I-like vehicle. In: 2007 22nd International Congress on Instrumentation in Aerospace Simulation Facilities. IEEE, pp. 1–12. http://dx.doi.org/10.1109/ICIASF.2007.4380901.

Singhal, A.K., Athavale, M.M., Li, H., Jiang, Y., 2002. Mathematical basis and validation of the full cavitation model. J. Fluids Eng. 124 (3), 617–624. http://dx.doi.org/10.1115/1.1486223.

Sousa, J., Gorlé, C., 2019. Computational urban flow predictions with Bayesian inference: Validation with field data. Build. Environ. 154, 13–22. http://dx.doi.org/10.1016/j.buildenv.2019.02.028.

Suzuki, T., Yamade, Y., Kato, C., 2011. Accuracy validation of cavitation performance prediction for single hydrofoil by LES. SEISAN KENKYU 63 (1), 55–60. http://dx.doi.org/10.11188/seisankenkyu.63.55.

Tamura, Y., Matsumoto, Y., 2009. Improvement of bubble model for cavitating flow simulations. J. Hydrodyn. 21 (1), 41–46. http://dx.doi.org/10.1016/S1001-6058(08)60117-1.

Tamura, Y., Sugiyama, K., Matsumoto, Y., 2001. Cavitating flow simulations based on the bubble dynamics. In: Proc. 4th International Symposium on Cavitation. CAV2001, pp. 1–8.

Vallier, A., 2013. Simulations of Cavitation-From the Large Vapour Structures to the Small Bubble Dynamics (Doctoral Thesis). Lund University.

Wang, Z., Cheng, H., Ji, B., 2021. Euler-Lagrange study of cavitating turbulent flow around a hydrofoil. Phys. Fluids 33, 112108. http://dx.doi.org/10.1063/5.0070312.

Wang, Z., Cheng, H., Ji, B., Peng, X., 2023. Numerical investigation of inner structure and its formation mechanism of cloud cavitating flow. Int. J. Multiph. Flow 165, 104484. http://dx.doi.org/10.1016/j.ijmultiphaseflow.2023.104484.

Watanabe, S., Kuchi-Ishi, S., Murakami, K., Hashimoto, A., Kato, H., Yamashita, T., Yasue, K., Imagawa, K., Saiki, H., Ogino, J., 2014. Towards EFD/CFD integration: development of DAHWIN-digital/analog-hybrid wind tunnel. In: 52nd Aerospace Sciences Meeting. p. 0982. http://dx.doi.org/10.2514/6.2014-0982.

Watanabe, S., Suefuji, N., Yamaoka, W., Furukawa, A., 2013. Lift and drag characteristics of a cavitating Clark-Y 11.7% hydrofoil. Turbomachinery 41 (7), 440–446. http://dx.doi.org/10.11458/tsj.41.440.

Zwart, P.J., Gerber, A.G., Belamri, T., 2004. A two-phase flow model for predicting cavitation dynamics. In: Proc. 5th International Conference on Multiphase Flow, vol. 152, pp. 1–11.