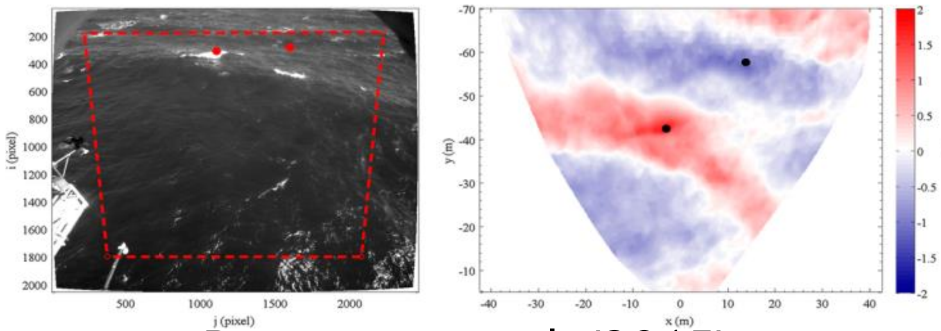
Ensemble-based data assimilation for predictable zones and application for non-linear deep-water waves

Wataru Fujimoto and Kinya Ishibashi

Nippon Kaiji Kyokai (ClassNK), Research Institute

ClassNK

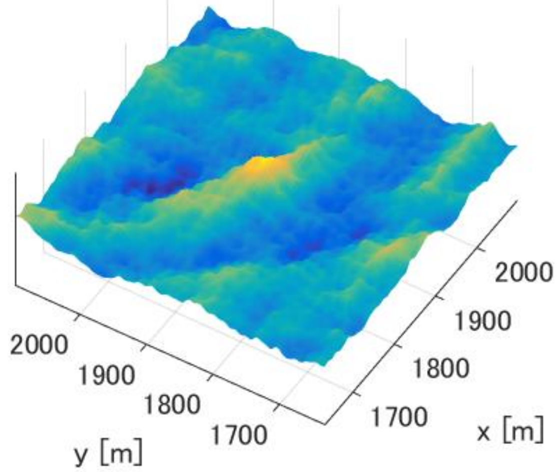
Observational data
Radar, Buoy, Stereo Camera…



Benetazzo et al. (2015)

Data assimilation

Spatio-temporal data of wave

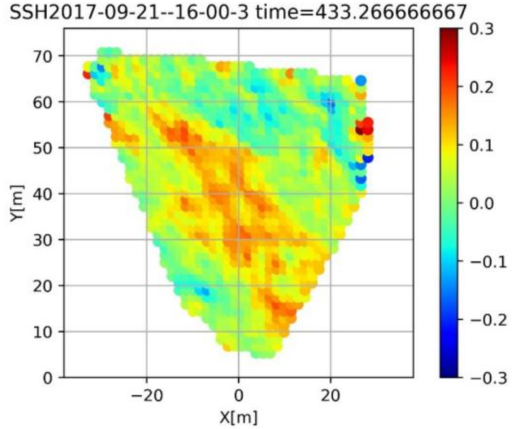


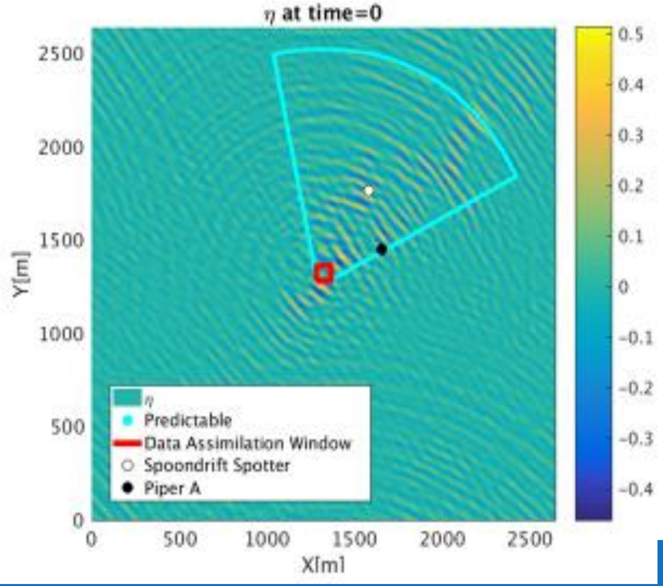
1. Surveillance of sea waves
    - Deterministic Sea Wave Prediction (DSWP)
    - For marine operations and wave energy converters (WEC)
    - Moderate sea
2. Investigation on freak wave occurrence
    - For maritime accident survey
    - Rough sea ← This study's objective

Previous study on ocean wave reconstruction

- Fujimoto & Waseda (2020) proposed a method to construct wave field from a limited amount of observational data.
- Surface Wave reconstruction using the Ensemble Adjoint-free Data assimilation method (**SWEAD**). It was applied to a field measurement.
- Waseda et al. (2021) estimated the wave field outside of observational area by SWEAD.

Stereo camera mounted on an observational tower in Japan coast measured wave propagation



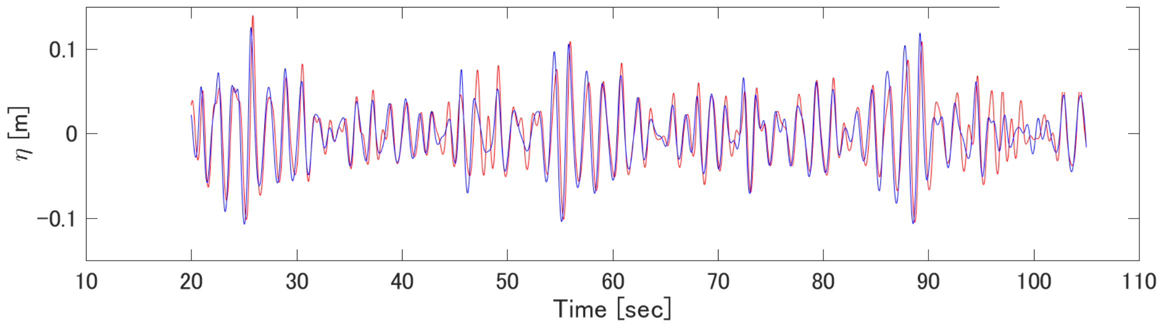


Two key elements of SWEAD

ClassNK

- Higher Order Spectral Method (**HOSM***)
  - Phase-resolved method which solves the Euler equations by spectral method
  - Widely applied for freak wave studies
- **Ensemble-based variational method**
  - The cost function can be defined as the squared-error between observation and model prediction.
  - Variational Method optimizes it to estimate the physical state like initial condition

y : Observational data



$\mathbf{A}(\mathbf{x})$: Model (HOSM) estimation

$$L(\mathbf{x}) = \frac{1}{2}||\mathbf{A}(\mathbf{x}) - \mathbf{y}||^2$$

$\mathbf{x}$: Fourier coefficient of initial surface elevation

* Dommermuth & Yue (1987), West et al (1987)

Variational method

ClassNK

- If the $\mathbf{x}$ is optimal, the gradient of cost function should be zero.
- To calculate the gradient, adjoint method, a major variational method, differentiates all formulation of physical models.
- Implementation cost is large.
- Ensemble-based method uses numerical differentiation by perturbed model simulation.
  - Easy to implement and parallelize

Minimum Condition

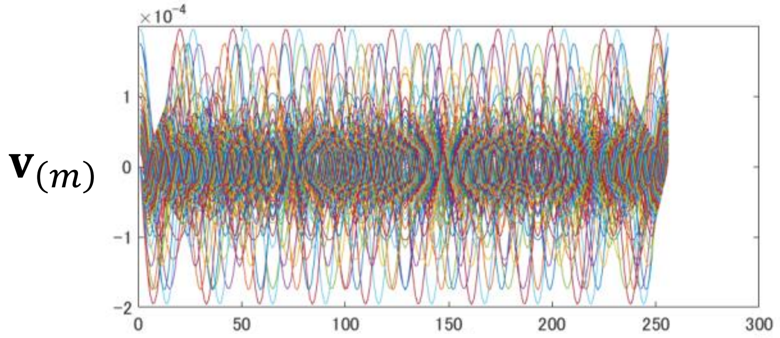
$$\nabla L(\boldsymbol{x}) = \mathbf{A}^{*}(\mathbf{A}(\mathbf{x}) - \mathbf{y}) = \mathbf{0}$$

Adjoint code calculate the adjoint matrix

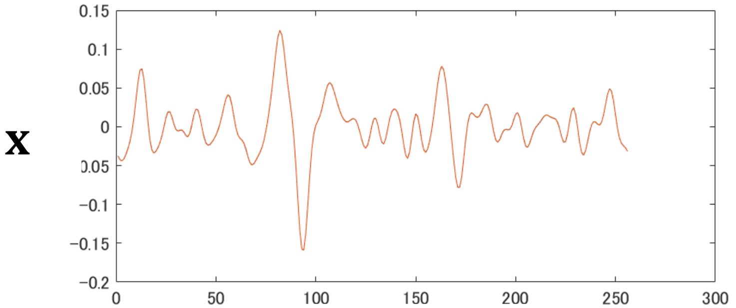
Perturbed HOSM simulation

ClassNK

The initial value perturbation $\mathbf{v}_{(m)}$
Cosine or sine wave



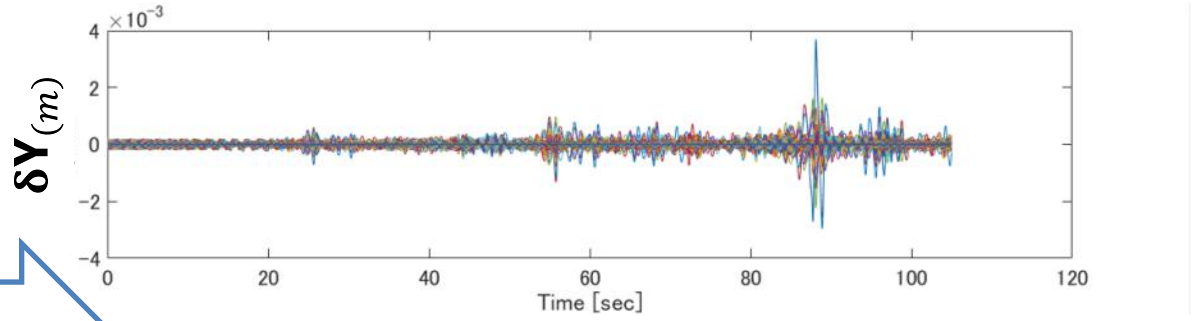
$\mathbf{x}_n$: Initial surface elevation



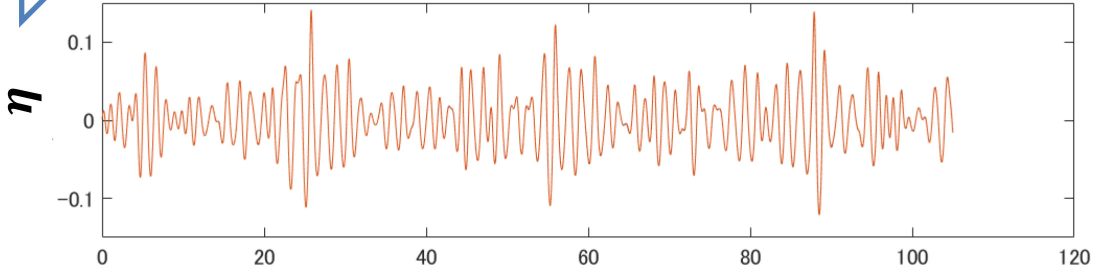
Position

HOSM

Corresponding the model perturbation $\boldsymbol{\delta}\mathbf{Y}^{(m)}$



$\mathbf{A}(\mathbf{x}_n + \varepsilon \mathbf{v}_{(m)})$: Model estimation



Time

$$\boldsymbol{\delta}\mathbf{Y}_{(m)} = \frac{1}{\varepsilon}\{\mathbf{A}(\mathbf{x}_n + \varepsilon \mathbf{v}_{(m)}) - \mathbf{A}(\mathbf{x}_n)\} \approx \mathbf{A}\mathbf{v}_{(m)}$$

Matrix form $\boldsymbol{\delta}\mathbf{Y} = \mathbf{A}\mathbf{V}$

A: Jacobian matrix of the cost function

Ensemble-based Variational method

ClassNK

- The minimum condition of the cost function can be written with **V**.
- Increment for next iteration is expressed by linear superposition of **V**.

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{V}\mathbf{w}_n$$

- Weighting coefficient $\mathbf{w}_n$ can be obtained by solving the linear equation on the right side.
- How to generate perturbations **V** is important for optimization efficiency.
- SWEAD used Fourier mode.

Minimum condition

$$\mathbf{V}^*\nabla L(\mathbf{w}_n) = \mathbf{V}^*\mathbf{A}^*(\mathbf{A}(\mathbf{x}+\mathbf{w}_n)-\mathbf{y})$$

$$= \boldsymbol{\delta}\mathbf{Y}^*(\mathbf{A}(\mathbf{x})+\boldsymbol{\delta}\mathbf{Y}\mathbf{w}_n-\mathbf{y}) = \mathbf{0}$$

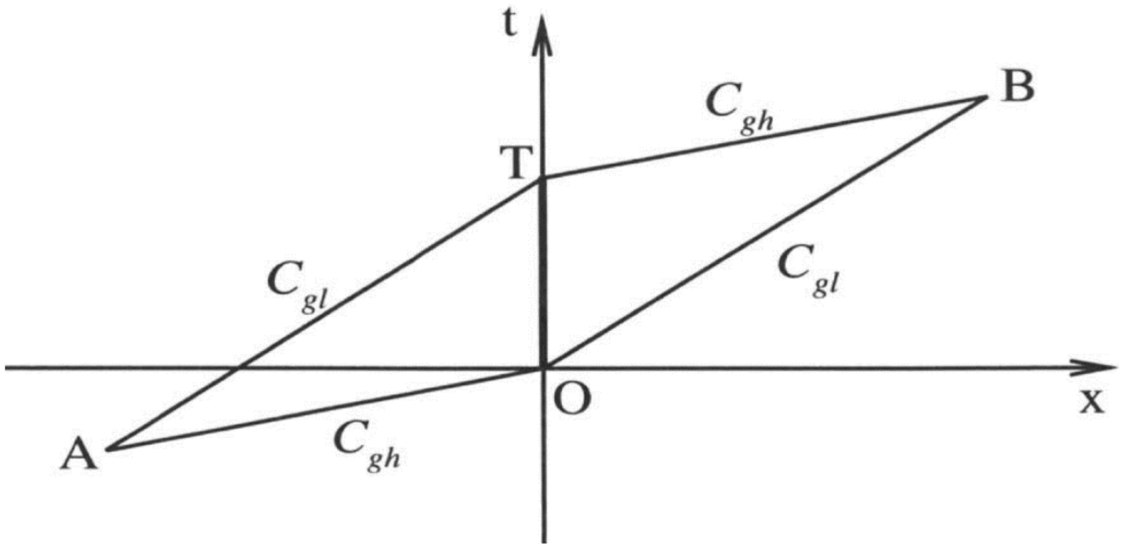
$$\therefore \boldsymbol{\delta}\mathbf{Y}^*\boldsymbol{\delta}\mathbf{Y}\mathbf{w}_n = -\boldsymbol{\delta}\mathbf{Y}^*(\mathbf{A}(\mathbf{x}_n)-\mathbf{y})$$

Predictable Zone Theory

ClassNK

- By the way, in dispersive waves like deep-water waves, the dispersion relationship confines the **predictable zone** for a limited amount of observational data.
- The observational data is time series, the predictable zone is a parallelogram surrounded by the slowest and the fastest wave components.
- **Fourier mode is global basis and might be redundant for a limited predictable zone.**

Slowest group velocity C_{gl}



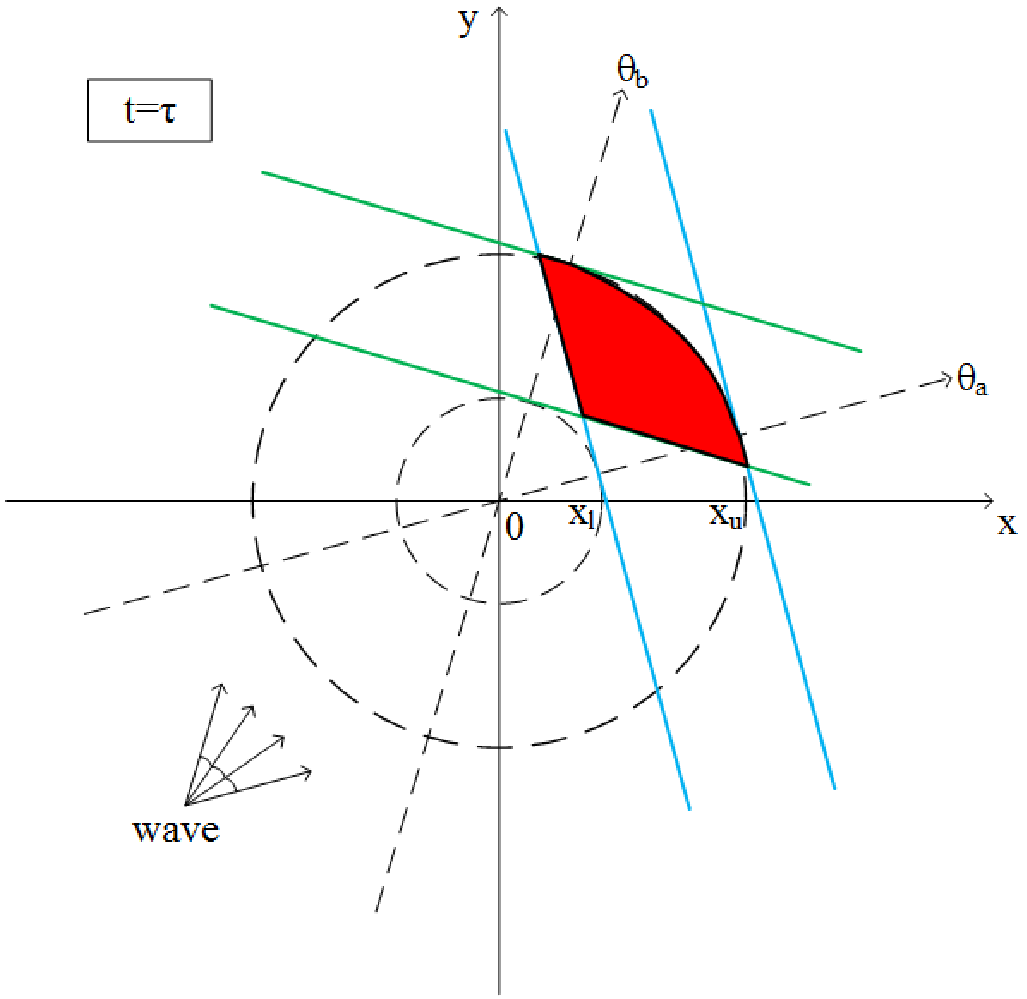
Fastest group velocity C_{gh}

Guangyu Wu, MIT Ph.D. Thesis (2004)

Predictable Zone in Directional Wave

ClassNK

- The directionality also confines the predictable zone.
- If the observational data is limited in spatial extent, predictable zone is also limited.



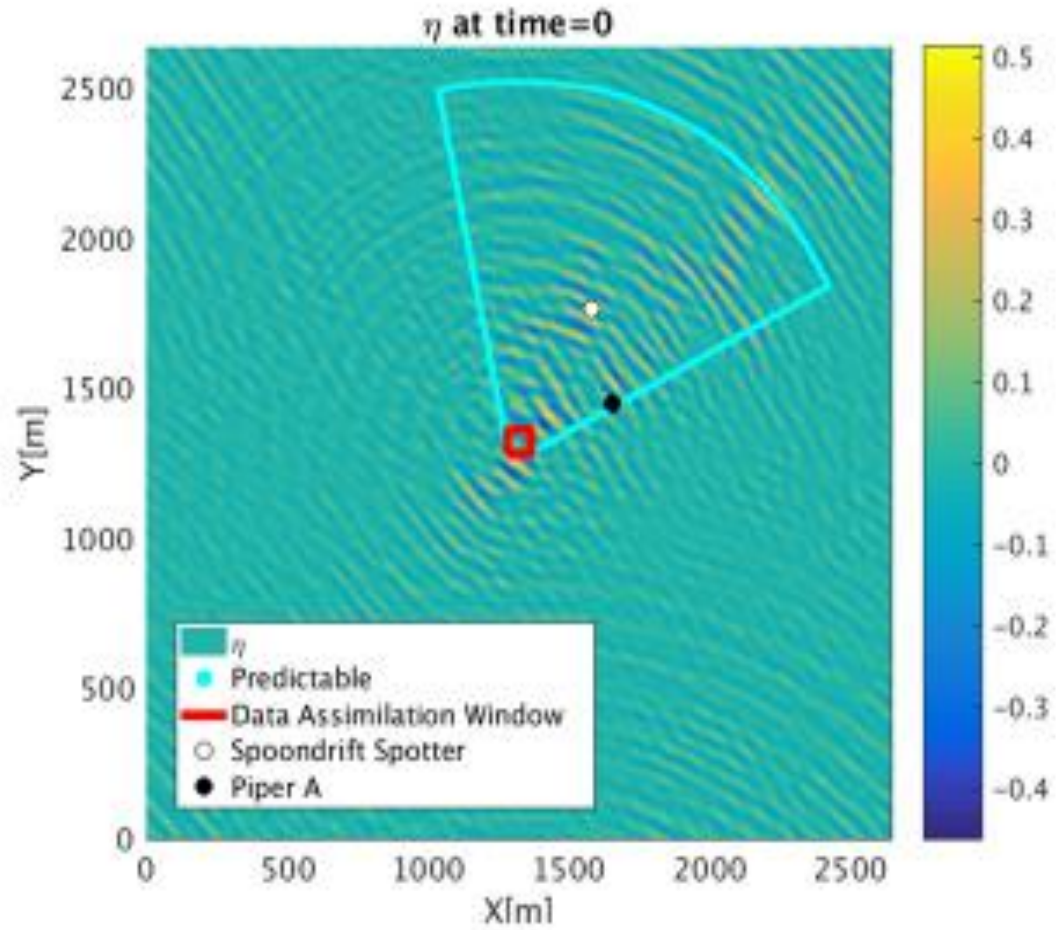
Qi et al. (2018)
[Predictable zone for phase-resolved reconstruction and forecast of irregular waves - ScienceDirect]

Objective: Dimension reduction of data assimilation

ClassNK

- **In reality, predictable zone is narrower than spatial domain of HOSM simulation.**
- **Idea of this study**
  - **Identify the predictable zone to find most effective initial value perturbation for computational efficiency.**
- **Predictable zone should be obtained analytically (next section)**

Waseda et al. (2021)



PREDICTABLE ZONE THEORY
IN A VIEWPOINT OF INVERSION PROBLEM

Predictable Zone Theory as Inversion Problem

ClassNK

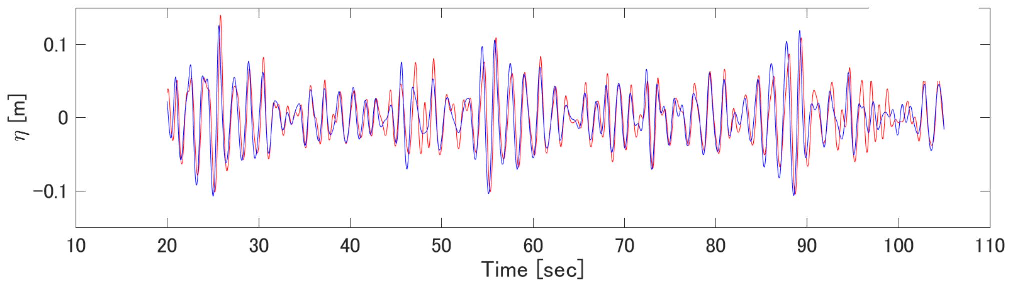
To estimate a physical state from an observational data, $\mathbf{A}(\mathbf{x}) = \mathbf{y}$ should be solved. If the model is linear (e.g., linear water wave), the equation can be written in a matrix and vector.

$$\mathbf{Ax} = \mathbf{y}$$

In general, the matrix A is not rectangular, and an ordinary inverse matrix cannot be used to obtain the solution in general.

$$\times \ \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$$

$\boldsymbol{y}$: Observational data

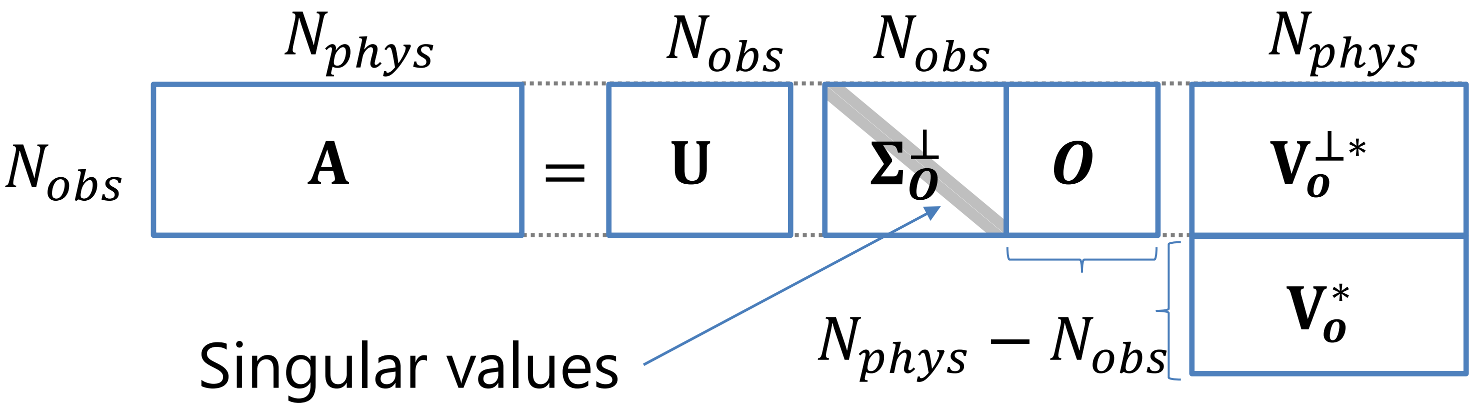


$\mathbf{A}(\mathbf{x})$: Model estimation

$\mathbf{x}$: Physical state (e.g., Initial condition)

Singular Value Decomposition (SVD)

- SVD $\mathbf{A} = \mathbf{U\Sigma V}^*$ is a generalization of eigenvalue decomposition for general matrix $\mathbf{A}$ $(m \neq n)$. $\mathbf{\Sigma}$ cosists of a diagonal matrix and a zero matrix.
- $\mathbf{U}$,$\mathbf{V}$ are orthogonal matrices. "Singular Vectors"
- There are singular vectors corresponding to zero singular values and non-zero singular values $\mathbf{V}_o$, $\mathbf{V}_o^\perp$. They are orthogonal to each other.



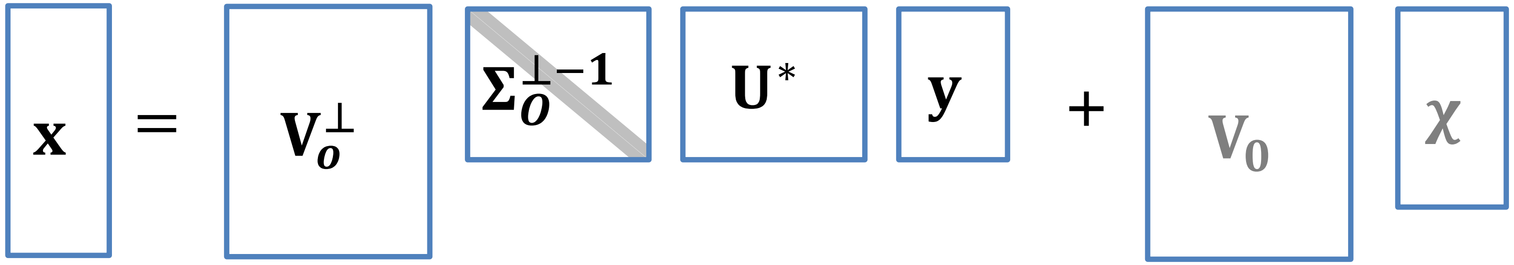
Predictable zone in viewpoint of inversion problem

ClassNK

- By using SVD, the solution of $\mathbf{Ax} = \mathbf{y}$ can be written as

$$\mathbf{x} = \mathbf{V}_o^{\perp}\mathbf{\Sigma}_o^{\perp-1}\mathbf{U}^{*}\mathbf{y} + \mathbf{V}_o\chi.$$

- χ is an arbitrary vector, and $\mathbf{V}_o\chi$ corresponds to an indefinite part of the solution.
  - $\mathbf{A}\,\mathbf{V}_o\chi = \mathbf{U\Sigma V}^{*}\mathbf{V}_o\chi = \mathbf{O}\chi = \mathbf{0}$ (vanishes!)
- From the observed data $\mathbf{y}$, only the first term of the above equation can be calculated; the second term is unknown owing to the arbitrary vector χ.
- **$\mathbf{V}_o^{\perp}$, which corresponds to non-zero singular values, spans the predictable zone.**

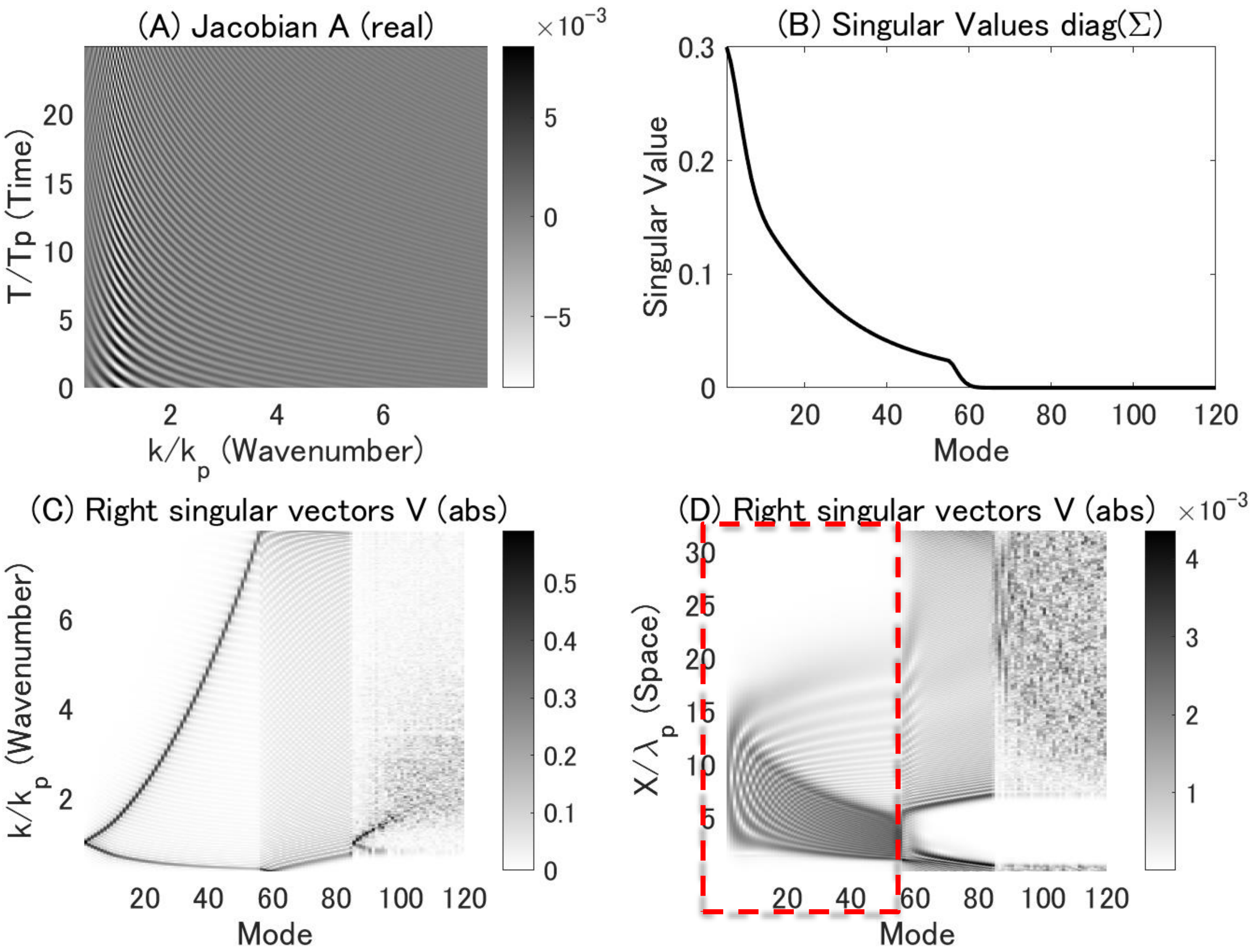


Example of SVD for linear wave $\omega^2 = gk$

- If the observed data is from a water level gauge and $\mathbf{x}$ is the Fourier coefficient of the initial surface elevation, then

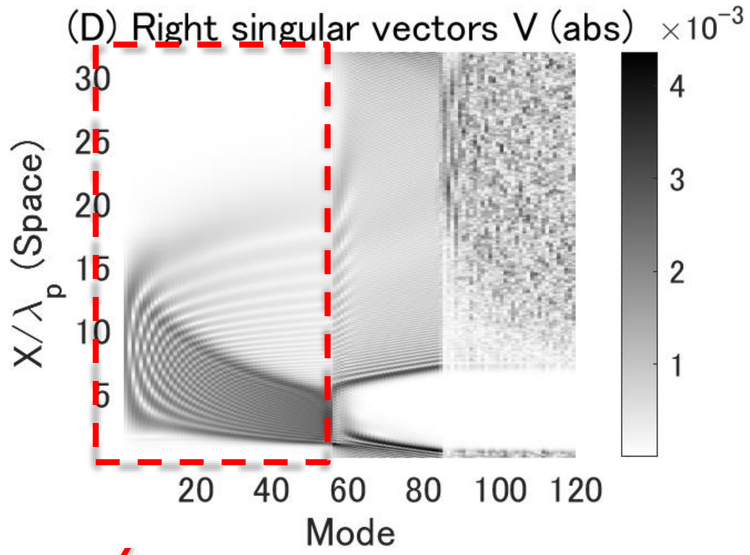
$$(\mathbf{A})_{qr} = \exp\left[i\left(\omega_r t_q\right)\right]$$

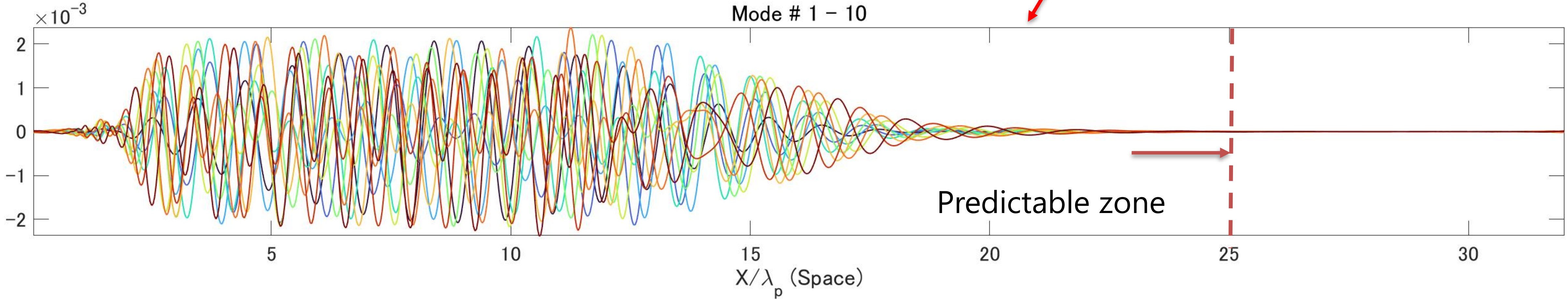
- **Right singular vectors V in red dashed line correspond to the predictable zone.**



Singular Vectors corresponding to the predictable zone

- If length of time series data is N wave period, the size of predictable zone is $N/2$ wave length.
- For example, if time series length is 50 T_p, the predictable zone size is $\mathbf{25}\ \boldsymbol{\lambda_p}$
- This figure shows the most dominant 10 right singular vectors, which are confined in the predictable zone.





UTILIZING PREDICTABLE ZONE THEORY FOR DATA ASSIMILATION

Ensemble-based data assimilation
using **Fourier modes**

vs.

Ensemble-based data assimilation
using **SVD**

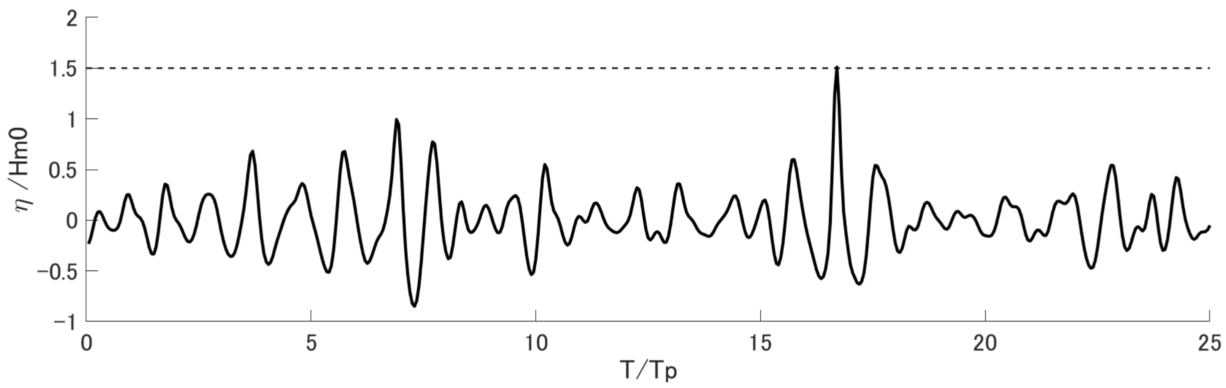
Problem setting

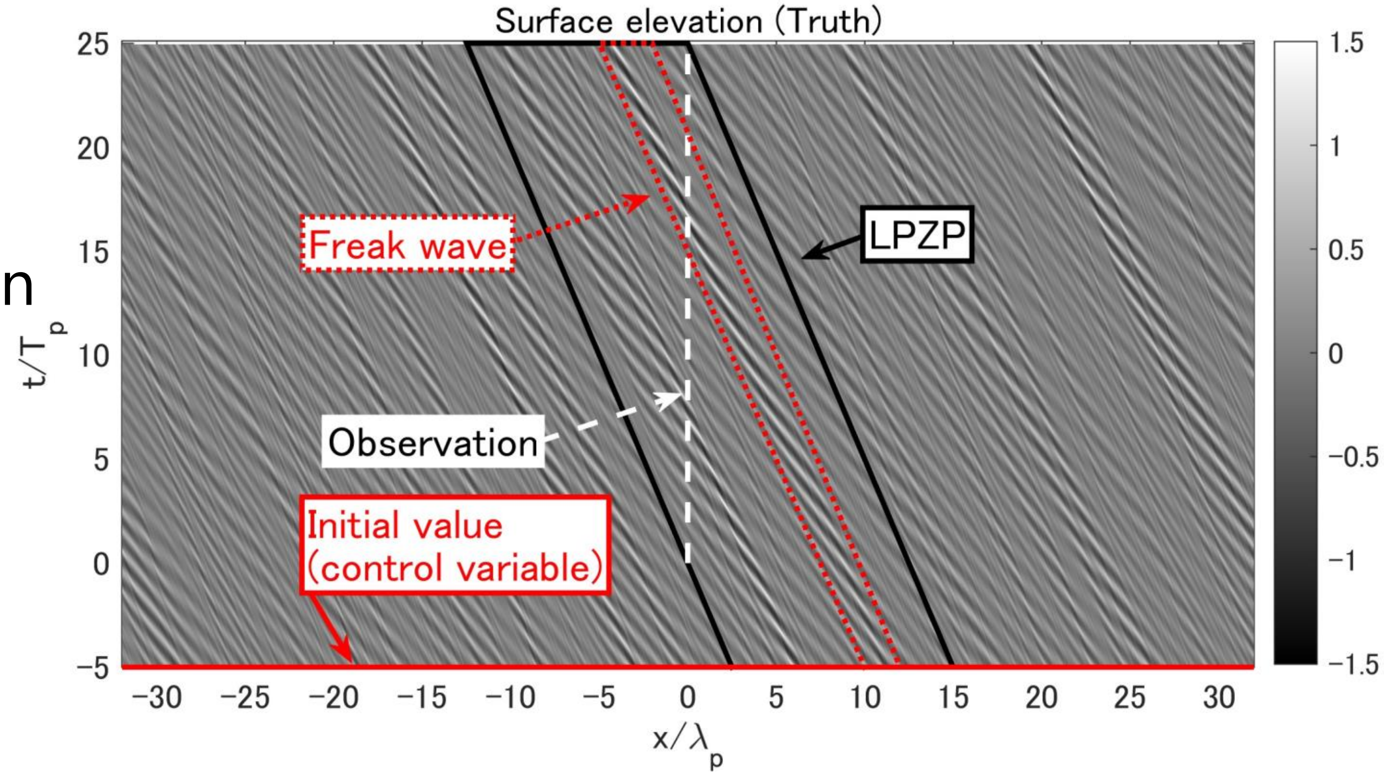
ClassNK

At the first, a freak wave is generated

- 3^{rd} order NL HOSM
- Duration: 50 T_p
- JOSNWAP γ=3.3
- Steepness: $H_s k_p/2$=0.11

- Reconstruct wave field from the observational data by data assimilation
- Compare the true value and the reconstructed value
- Linearly Predictable Zone corresponding Peak wavenumber **(LPZP)**





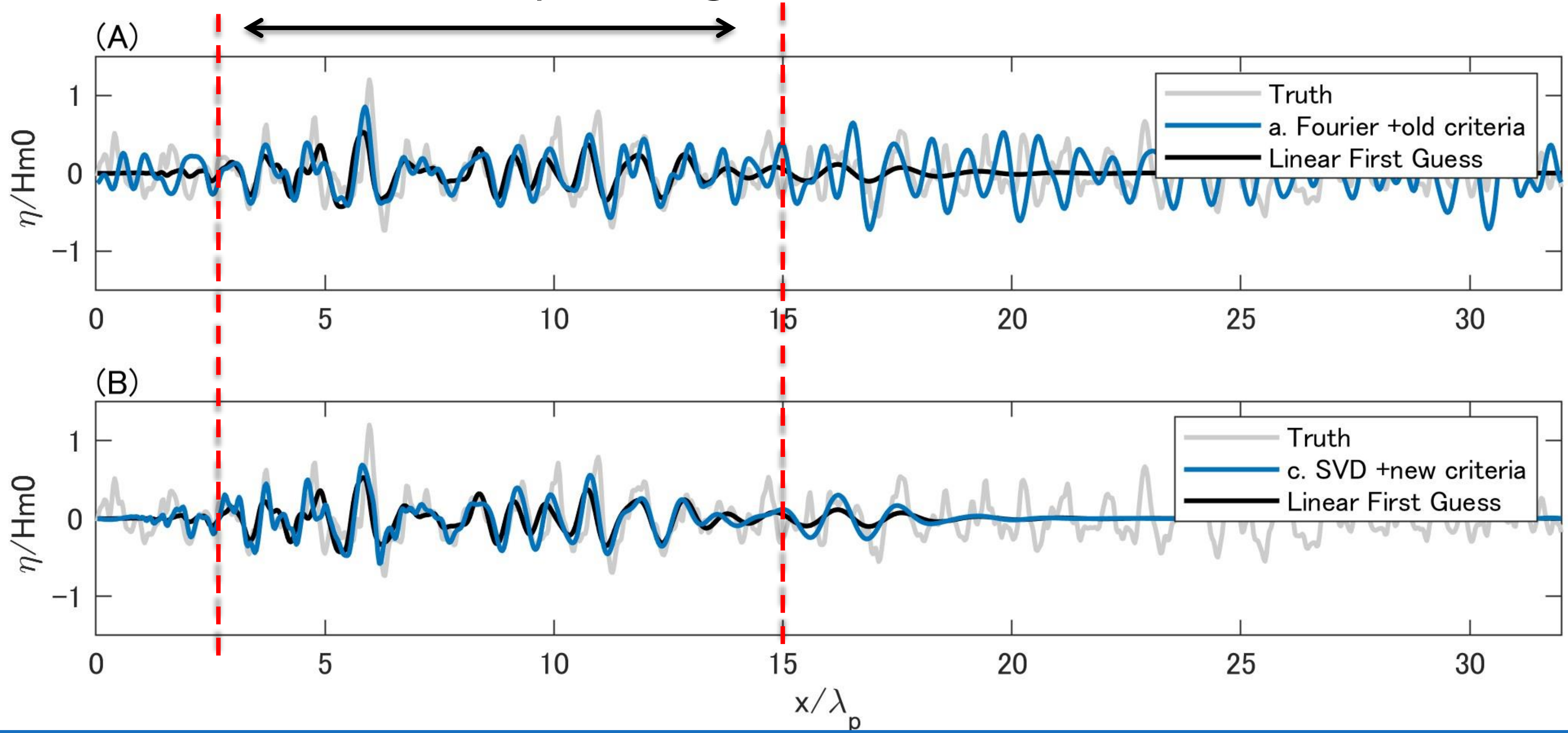
Analyses in the 20th iteration

- When Fourier was used, update from the linear first guess outside of LPZP is visible.
- When SVD was used, update from the linear first guess outside of LPZP is small.

Linearly Predictable Zone corresponding Peak wavenumber **(LPZP)**

Data assimilation using Fourier modes

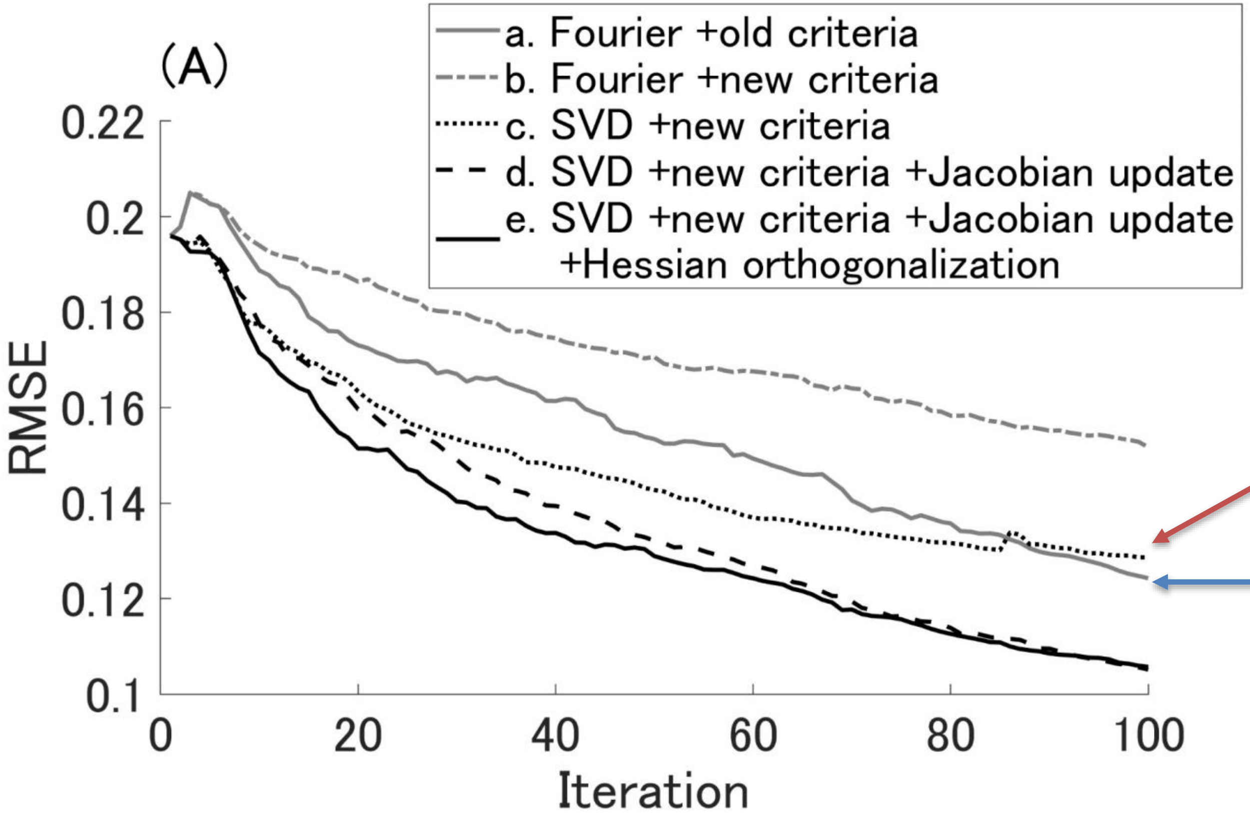
Data assimilation using SVD



Accuracy in LPZP for each iteration

ClassNK

- Root Mean Squared-Error (RMSE) decreased with iteration by the optimization.
- The method using SVD outperformed the method using Fourier mode until 80-th iter.
- After that, the method slowed down, and the method using Fourier mode was better.

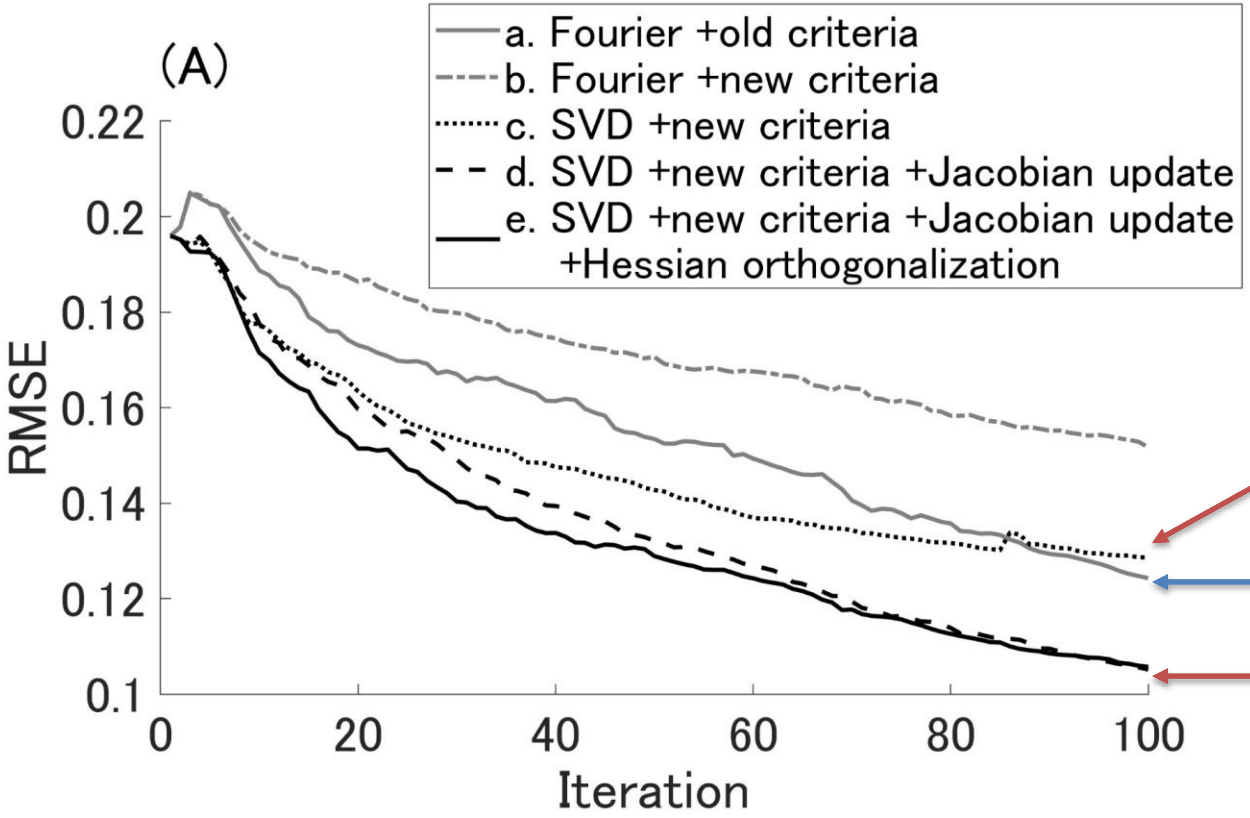


SVD and <u>neglecting</u> nonlinearity

Fourier modes

Accuracy in LPZP for each iteration

- The reason why the method using SVD slows down is that it neglects **nonlinearity**.
- We devised a method including nonlinearity for SVD, and the method was more efficient than the method using Fourier modes.



Propagation speed of Stokes wave increases with the steepness

$$\frac{\omega}{k} = 1 + \frac{1}{2}(ak)^2$$

ak: Steepness

SVD and neglecting nonlinearity

Fourier modes

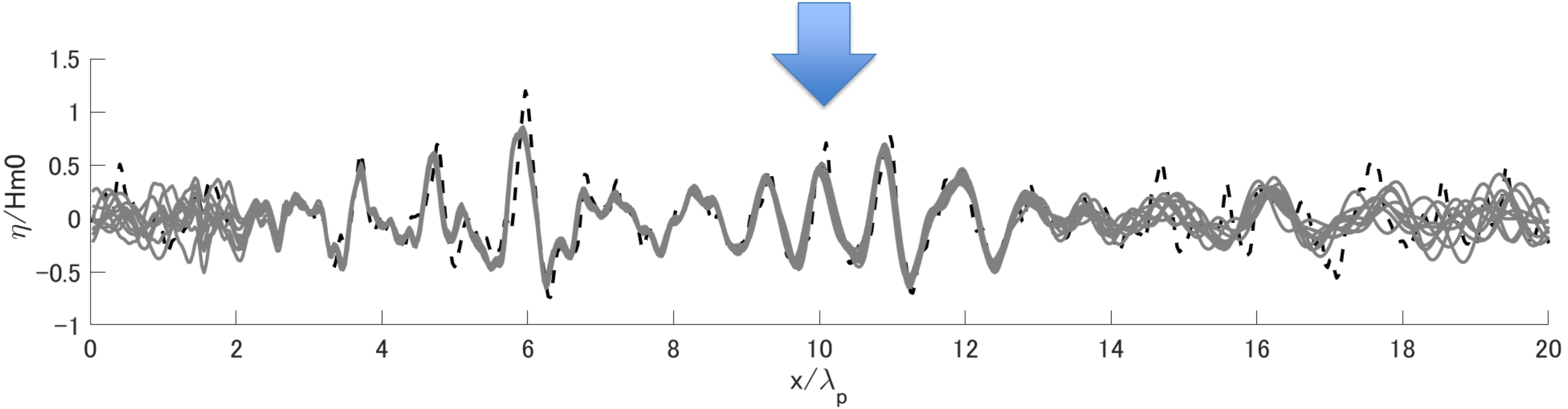
SVD and including nonlinearity

Analyses in the 100th iteration

ClassNK

Black dotted line -- : Truth
Gray solid lines: Analysis (different realization of noise added to the observation)

Original position of wave group leading to the freak wave at the initial time



- By the new method, the freak wave was reproduced well.

Conclusion

- This study proposed a new method using SVD to generate perturbations only in the predictable zone.

- We also devised a method including nonlinear dispersion, and the method is more efficient than the method using Fourier modes.
  - For the detail of the method, please check the following paper
    - **Fujimoto, W., and K. Ishibashi. 2023, Ensemble-based data assimilation for predictable zones and application for non-linear deep-water waves, Front Mar Sci, 10, .**

- Future issues: validation by tank tests and field measurements.

Thank you!