Influence of Spatial Heterogeneity in Error Characterization Using Triple Collocation

EGU General Assembly 2025

Diksha Gupta[1] & C.T. Dhanya[2]

Department of Civil Engineering, Indian Institute of Technology Delhi, India

[1] dikshagupta922@gmail.com, [2] dhanya@civil.iitd.ac.in

This presentation participates in OSPP – Outstanding Student & PhD candidate Presentation contest

1. Background and Objective

- Accurate error characterization is essential for validating satellite-based geophysical products.
- Triple Collocation (TC) estimates random error variances of three mutually independent datasets but assumes a common spatial scale—a condition rarely met in practice [1].
- Spatial heterogeneity in the ground truth and mismatches in spatial resolution introduces "spatial representativeness errors", whose influence on error variance estimates remains unexamined.

$$i = \alpha_i + \beta_i \theta + \epsilon_i$$

$i \in$ [X,Y, Z] represents three mutually independent spatially collocated datasets
ϑ = unknown "ground truth" signal of the target
α_i = systematic additive bias
β_i = systematic multiplicative bias
ϵ_i = random bias with zero mean

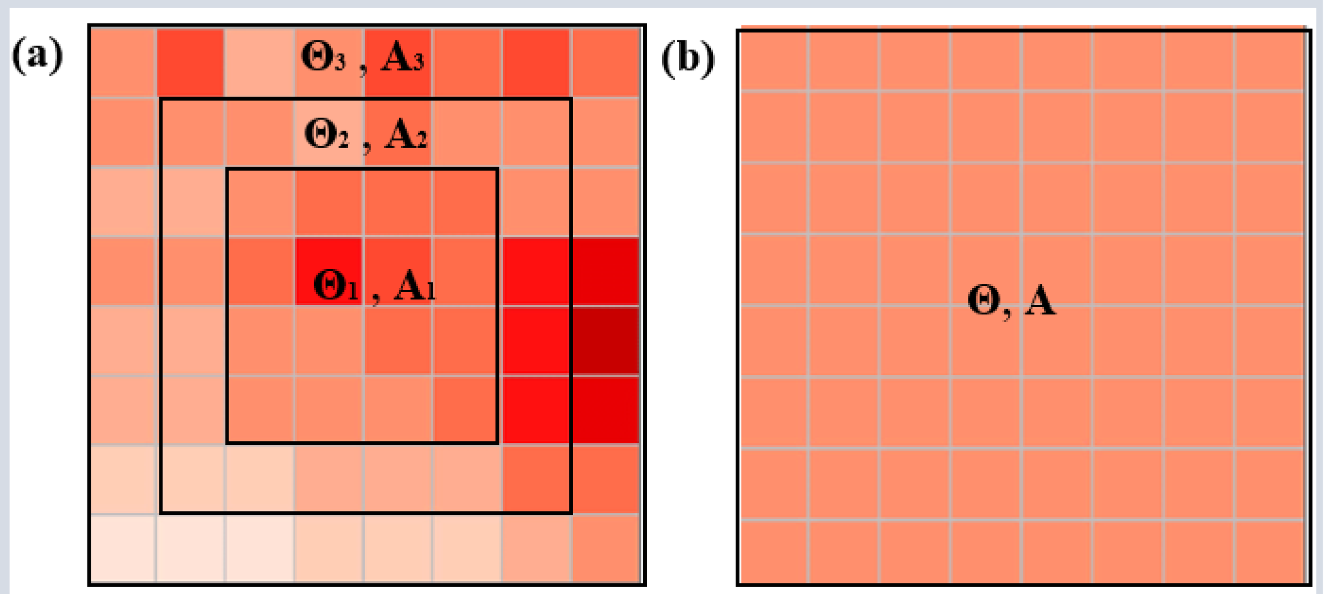

Figure 1. *Space showing the ground truth of the variable θ for the two cases: **(a)** Spatial Homogeneity) and **(b)** Spatial Heterogeneity. Black line indicates different spatial resolutions.*

How spatial heterogeneity effect the error variance estimates using Triple Collocation?

2. Data and Methods

- Synthetic soil moisture dataset with varying spatial heterogeneity is generated, as shown in Fig. 2.

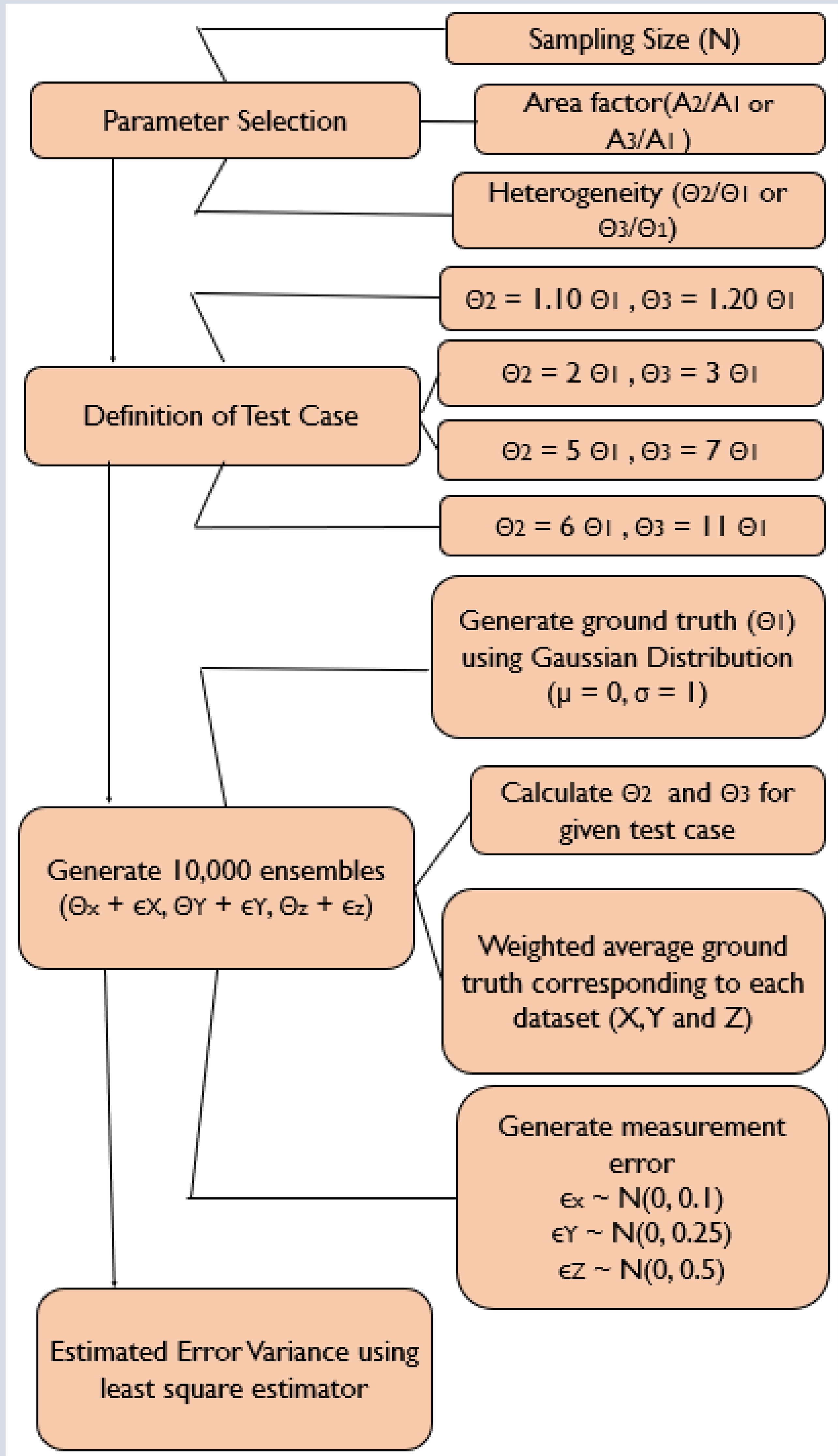

Figure 2. *Flowchart showing the generation of the synthetic dataset.*

3. Results and Discussion

- The error variance varies significantly with the parameters of spatial heterogeneity and area factor.
- The influence of spatial heterogeneity decreases significantly with increasing *N* (Fig. 3).
- TC framework exhibits equal likelihood of overestimation and underestimation (Fig. 4).

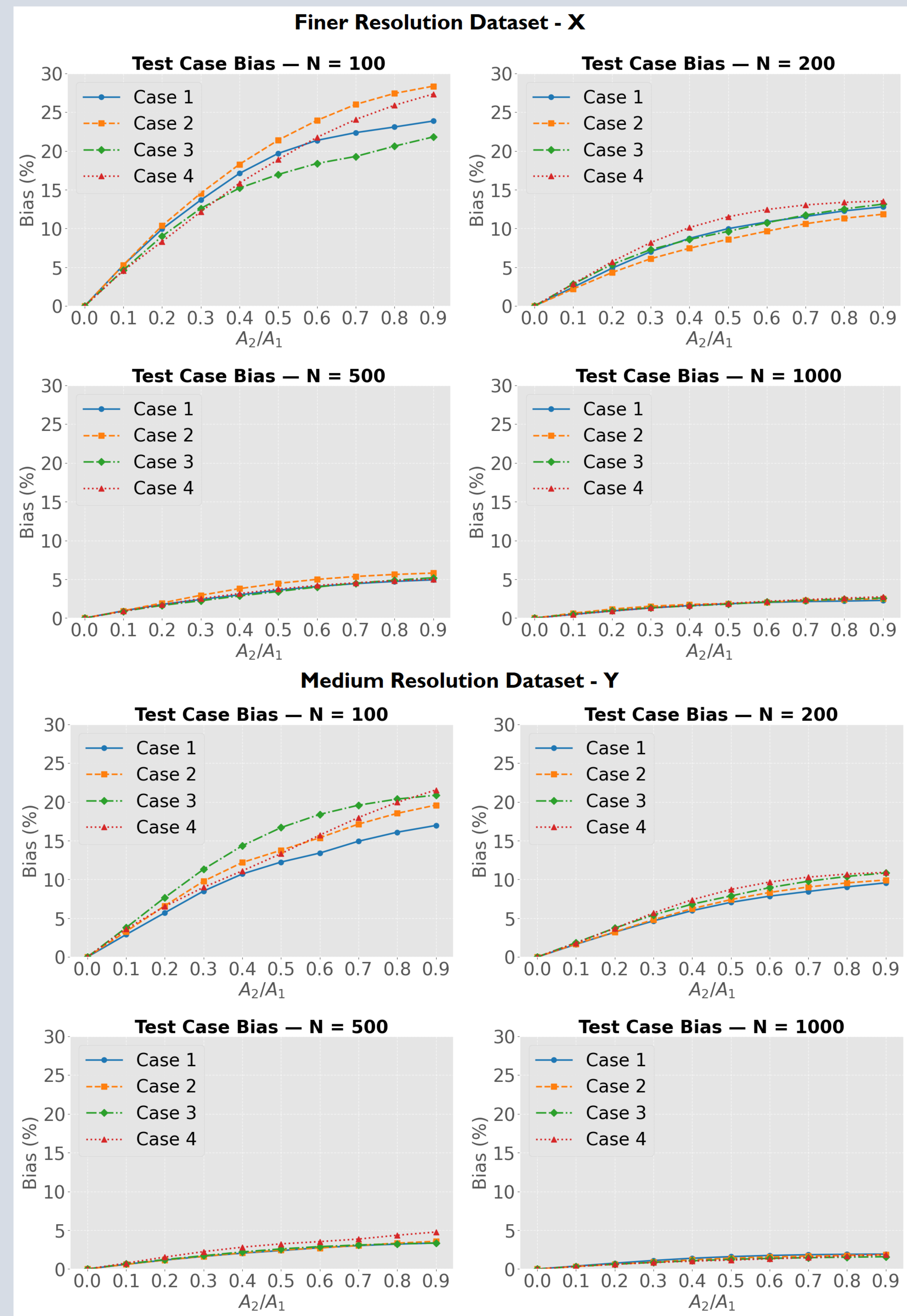

Figure 3. *Range of the % bias variation over 10,000 synthetically generated ensembles of the soil moisture dataset for finer and medium resolution dataset.*

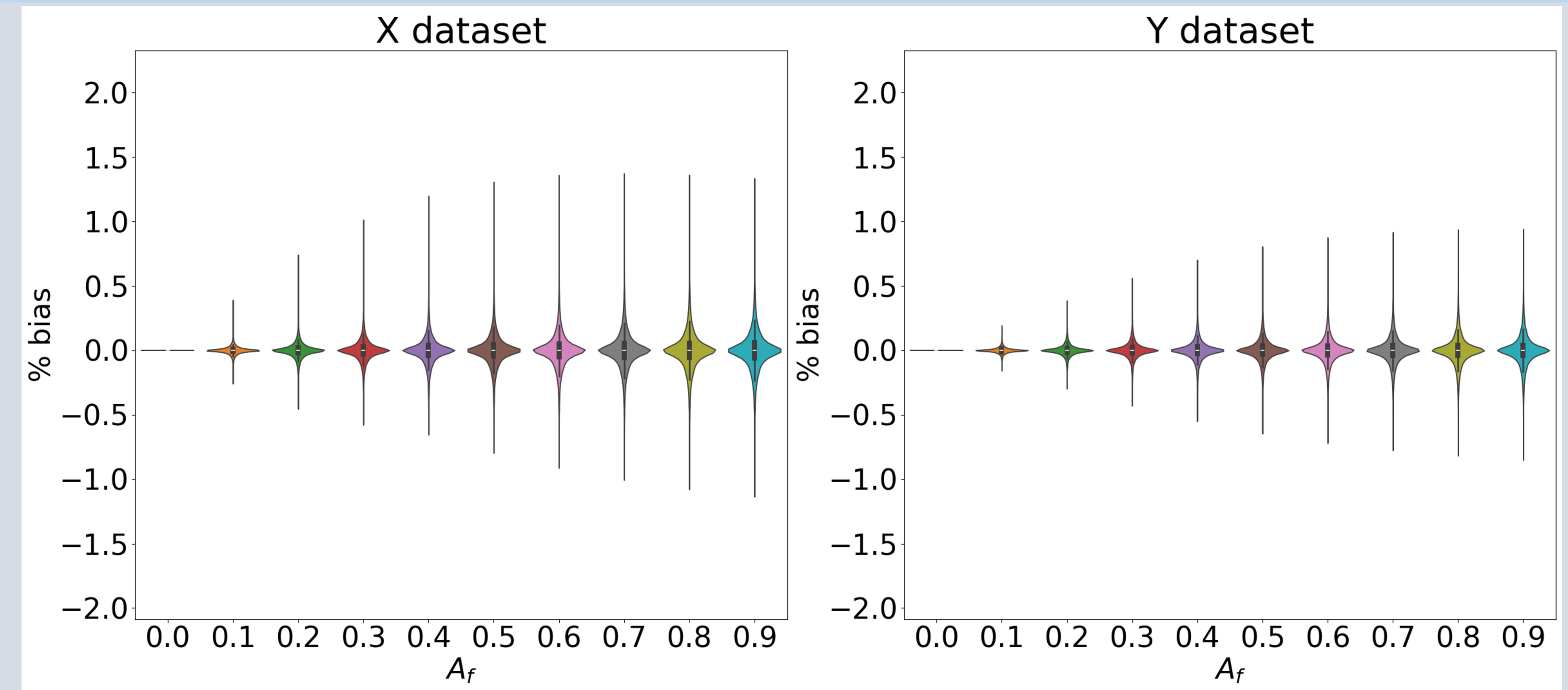

Figure 4. *Violin plot showing the variation of % bias for test case 1 and N = 1000, for X and Y dataset.*

- % bias decreases with increasing sampling sizes (*N*), from 30% to 3%.
- For smaller *N*, % bias significantly increases with the area factor (A_2/A_1), ranging from 0% to around 25% - 30%.
- Lesser % bias for the medium resolution dataset compared to the coarser dataset.

4. Conclusions & Future Scope

- The existence of spatial heterogeneity in the "ground truth" affect the estimated error variance using triple collocation.
- Selection of larger sampling size (*N*>500) is important to improve error variance estimates.
- Future works can progress on incorporating the factor for spatial heterogeneity to the triple collocation framework.
- The framework will aid to enhance data accuracy for data producers and data users.

References: *[1] Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. J. Geophys. Res.-Oceans 103 (C4), 7755–7766.*

Acknowledgement: *The authors would like to express sincere gratitude to Indian Institute of Technology Delhi, India for supporting this work. The authors would also like to thank all the data providers for making datasets available for research.*