

SUPPLEMENTARY MATERIAL

Nonlinear-Constrained Geographically Weighted Regression: Enabling Multidimensional Dynamics Analysis Using Monostatic Terrestrial Radar Interferometry at Helheim Glacier

Chentong Zhang, Xianwei Wang, Yi Zhou, David M. Holland, Denise Holland

Contents of this file

Figure S1. Data processing results used in this study.

Figure S2. RMSE between NL-GWR output velocity (Site 2) and bistatic TRI versus different parameter settings.

Text S1. Solving Algorithms of Nonlinear Optimization Question.

Text S2. Weight Adaptive Method.

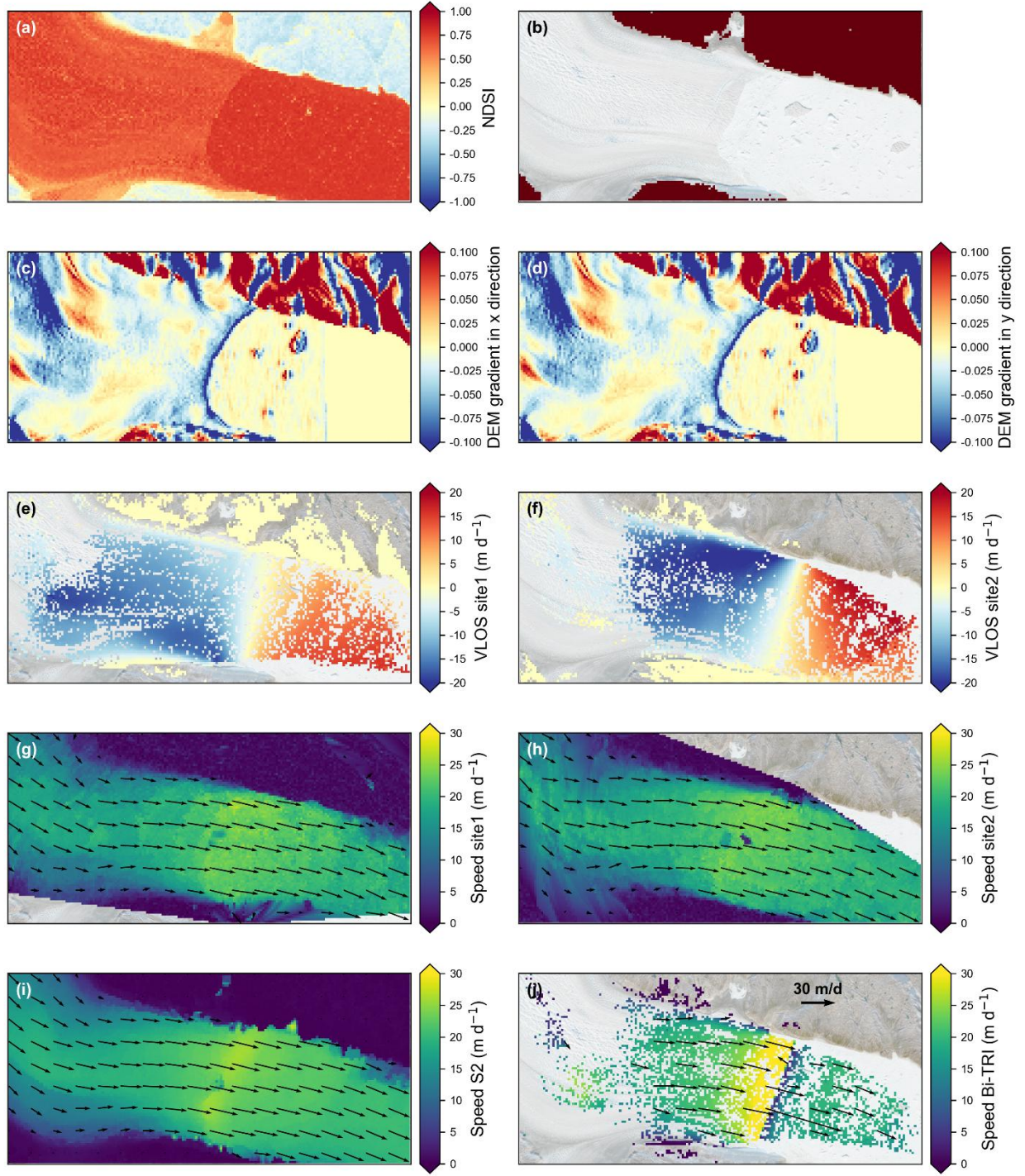


Figure S1. Data processing results used in this study, including (a) NDSI values, (b) bedrock outcrop mask created based on $\text{NDSI} < 0.2$ and marked by dark red mask, (c-d) spatial gradient of surface elevation in x and y direction, (e-f) interferometric LOS velocity for Site1 and Site2 (Sect. 2.2.1), (g-h) feature-tracking-based speed for Site1 and Site2 (Sect. 2.2.2), (i) Sentinel-2A/B-derived speed (Sect. 2.2.3), and (j) velocity extracted by bistatic TRI method. (e-f, j) use the average of 29 interferogram results with two-minute intervals on August 7. (e-f, g-i) undergo removal of points with speed $> 30 \text{ m/d}$ and bilinear interpolation to fill gaps before being input to NL-GWR calculations.

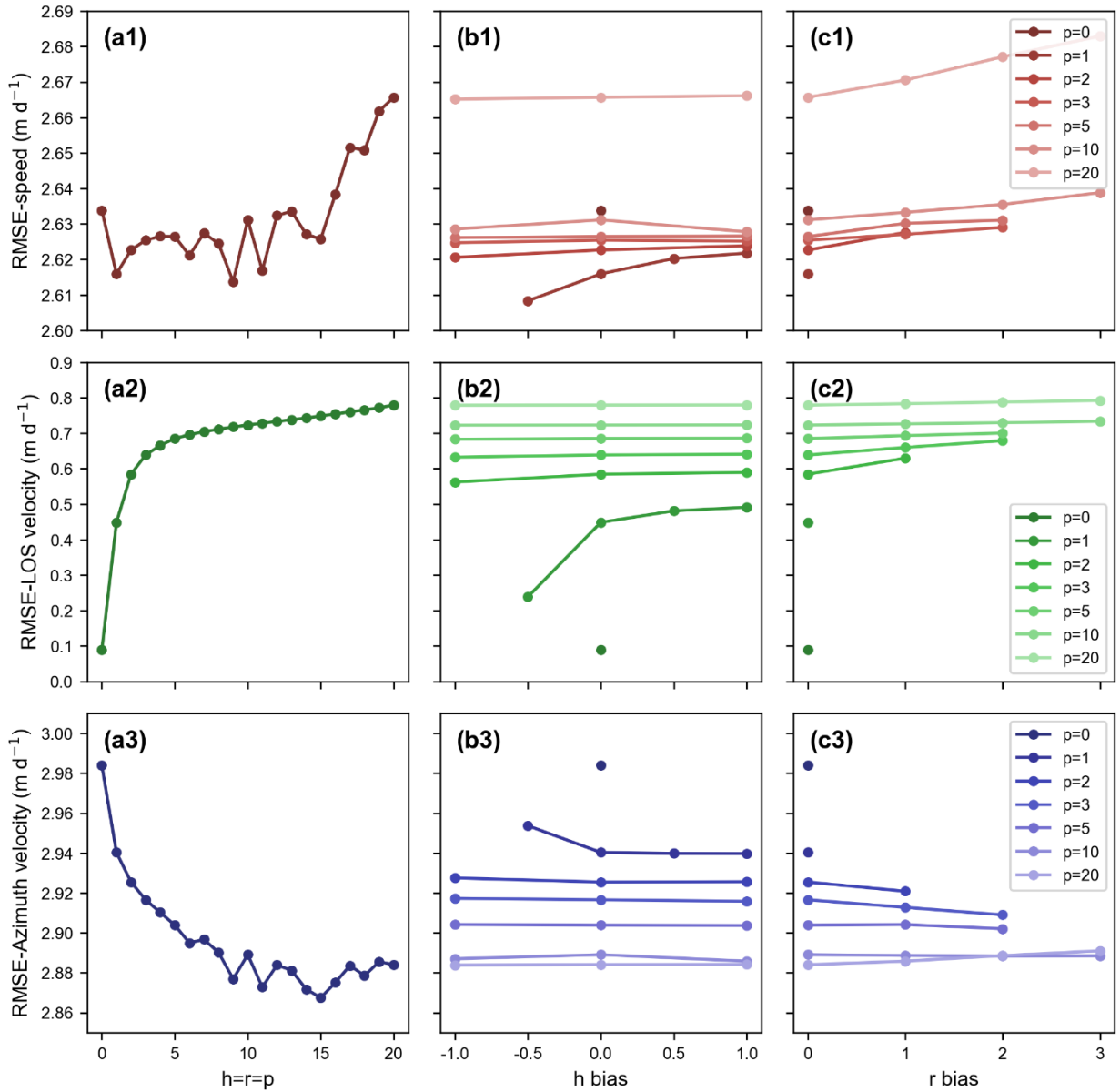


Figure S2. RMSE between NL-GWR output velocity (Site 2) and bistatic TRI versus different parameter settings. The figure's setting is the same as Figure 7 in the main text.

Text S1. Alternative Solving Algorithms of Nonlinear Optimization Question

This section primarily introduces the mathematical framework for solving nonlinear constrained regression problems:

$$\mathbf{y} = \mathbf{Ax} + \epsilon, \quad g(\mathbf{x}) = 0$$

The algorithmic approach is decomposed into linear operations or matrix operations that are readily transferable to code, with flowchart in Figure S3. The penalty-function approach for implementing the proposed algorithm has been introduced in the main text. Here, we further describe the alternative implementation based on the Lagrange multiplier method.

The Lagrange multiplier method is a hard-constraint approach. We design the Lagrangian function $L(\mathbf{x}, \mu)$ as:

$$L(\mathbf{x}, \mu) = \frac{1}{2}(\mathbf{y} - \mathbf{Ax})^T \mathbf{W}(\mathbf{y} - \mathbf{Ax}) + \mu g(\mathbf{x})$$

The solution satisfying the constraints of the original regression problem corresponds to an extremum point of this function. The conditions for this function to reach an extremum are:

$$\begin{cases} \mathbf{A}^T \mathbf{W} \mathbf{Ax} + \mu \nabla g - \mathbf{A}^T \mathbf{Wy} = 0 \\ g(\mathbf{x}) = 0 \end{cases}$$

The problem is thus transformed into solving this system of nonlinear equations. This method can strictly ensure that the final result satisfies the nonlinear constraint function.

Below, we employ the Newton-Raphson iterative method to solve this problem. Also known as Newton's tangent method, it is used to find zeros of nonlinear functions. For concise notation, we define the two-dimensional function $F_k := F(\mathbf{x}_k, \mu_k)$:

$$F(\mathbf{x}_k, \mu_k) = \begin{bmatrix} \mathbf{A}^T \mathbf{W} \mathbf{Ax}_k + \mu_k \nabla g - \mathbf{A}^T \mathbf{Wy} \\ g(\mathbf{x}_k) \end{bmatrix}$$

The problem then becomes finding the zero of this function. According to the algorithmic principle, we need to solve:

$$\mathbf{J}_k \Delta \mathbf{x}_k^\mu = -F_k, \quad \Delta \mathbf{x}_k^\mu = \begin{bmatrix} \Delta \mathbf{x}_k \\ \Delta \mu_k \end{bmatrix}$$

for the matrix solution $\Delta \mathbf{x}_k^\mu$, where $\mathbf{J}_k$ is the Jacobian matrix of F_k , with the form:

$$\mathbf{J}_k = \begin{bmatrix} \mathbf{A}^T \mathbf{W} \mathbf{A} + \mu_k \nabla^2 g & \nabla g \\ \nabla g^T & 0 \end{bmatrix}$$

where ∇g is the gradient vector of function g , and $\nabla^2 g$ is the Hessian matrix of function g .

This is a linear system that can be solved directly using matrix inversion algorithms. By setting an appropriate initial condition $\mathbf{x} = \mathbf{x}_0$ (we use GWR solution as this initial condition), we compute $\Delta \mathbf{x}$ from the

above equation and update the solution via $\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma \Delta \mathbf{x}_k$, where γ is relaxation factor used to avoid

numerical instability. Convergence is achieved when the norm of $\Delta \mathbf{x}_k$ falls below a specified threshold τ_L , and

the final solution $\mathbf{x}_{k+1}$ is output. We will also set a maximum iteration times M to determine divergence.

It should be noted that the Lagrange multiplier method's mandatory enforcement of constraint satisfaction inevitably makes its convergence more difficult than the penalty function method. In fact, during testing for this study, the NL-GWR method based on the Lagrange multiplier method generally exhibited algorithmic

divergence, that is, cannot achieve convergence within the iteration times of M . Interested readers may attempt this in the GitHub code provided with this paper.

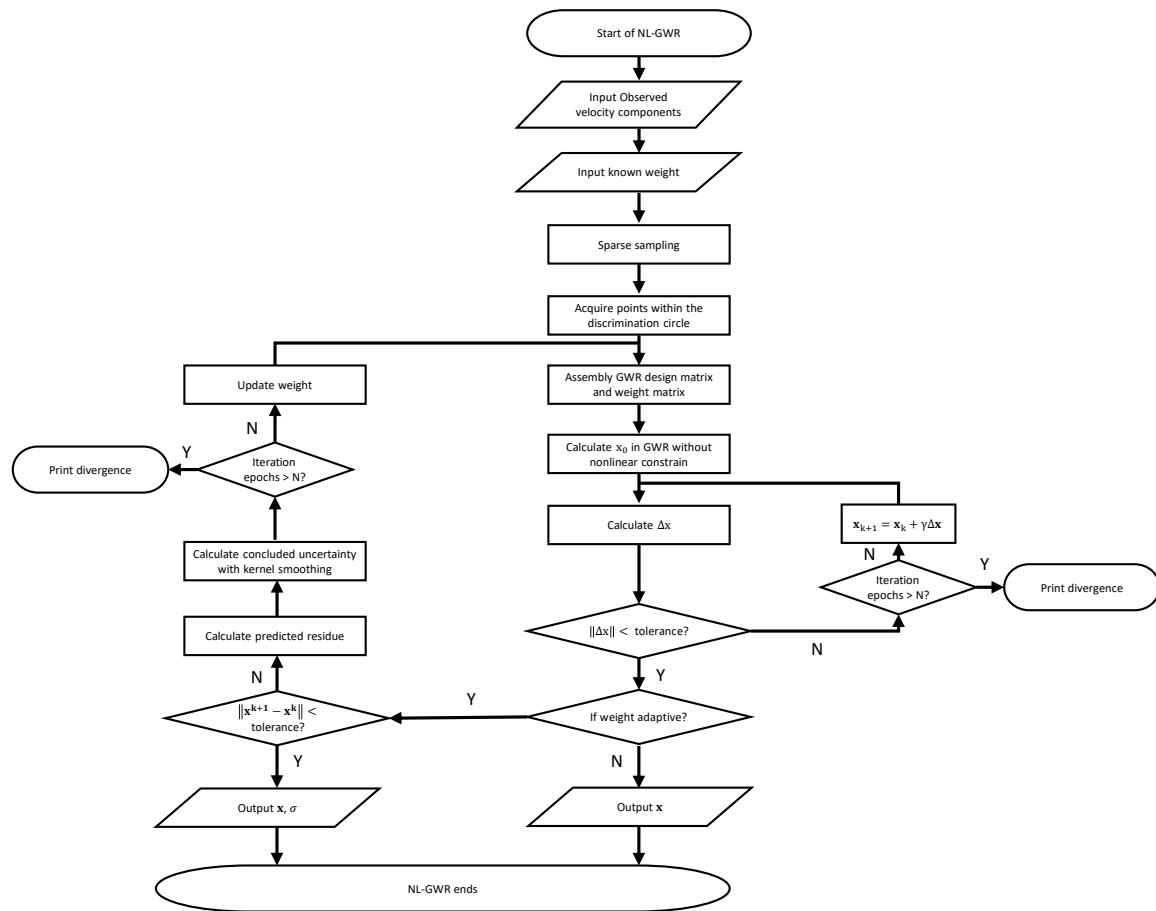


Figure S3. Flowchart of the NL-GWR core algorithm.

Text S2. Weight Adaptive Method

The weight matrix W_σ for weighted regression can be obtained through methods that infer statistical variance properties without requiring any prior information. Common approaches include the Variance Component Estimation (VCE) method [1], and the TC/MTC methods [2]. However, the VCE method is applicable to ordinary weighted regression but can only yield spatially uniform variance when applied to geographically weighted regression without modification; the TC/MTC methods impose strict requirements on the dimensions of observed quantities and target regression variables, making them inconvenient to adapt. In this paper, we design a weight adaptive method suitable for the proposed framework based on kernel smoothing techniques [3], from the perspective of the assumed properties of weights, with details in Figure S4.

Kernel smoothing is a nonparametric inference-based data fitting algorithm [3] that operates independently of the regression algorithm. Similar to the VCE method, we assume that the relationship between regression observation residuals and weights satisfies the distribution:

$$\epsilon \sim N(0, \sigma^2)$$

where σ corresponds to the observed uncertainty variable in weighted regression, which according to prior assumptions represents the reciprocal of the variance or the square of the uncertainty metric of the observed variable, and N denotes the normal distribution. Our objective is to use the concluded residuals $\hat{\epsilon}$ obtained from each regression to fit the two-dimensional distribution of weights/observation vector variance, which then serves as prior weight information for the next regression iteration, continuing until the deviation between successive regression results falls below a given tolerance.

Assuming that in the regression problem to be solved in this paper, we have no prior knowledge about the weights of any component of $\mathbf{y} = (y_i)$. In this case, we initialize the assumed statistical standard deviation of all observation components as $\sigma_i(s) = 1$ or any nonzero initial value. This standard deviation is then treated as the prior covariance for the observation vector and substituted into the NL-GWR framework in the main text to compute the state vector result $\mathbf{x}^{(0)}(s)$ and the predicted observation values $\hat{\mathbf{y}}^{(0)} = \mathbf{A}\mathbf{x}^{(0)}(s)$ based on this regression result. The residuals between this result and the true observation values are:

$$\epsilon^{(0)} = \mathbf{y} - \mathbf{A}\mathbf{x}^{(0)}(s)$$

Select an appropriate variance kernel bandwidth b , whose role is similar to the bandwidth h in the main text, characterizing the spatial correlation influence range of the variance. Update the latest state vector statistical standard deviation:

$$(\sigma^2)_i^{k+1}(s_i) = \frac{\sum_j K_b(d_{ji})(\epsilon^2)_i^k(s_j)}{\sum_j K_b(d_{ji})}$$

where $K_b(x) = K(x/b)$ is the kernel function representing the geographical correlation form of the variance kernel bandwidth, consistent with the form of kernel function K in the main text. Repeat the above operations until:

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\| < \tau_\sigma$$

where τ_σ is the tolerance that determines convergence before maximum iteration times M_σ . When convergence is achieved, it indicates that the final variance field represents an unbiased uncertainty magnitude of the observation vector inferred from statistical relationships.

Like in the main text, due to sparse sampling, variance calculation is also limited to these sparse sampling points. Therefore, after each computation, the variance is linearly resampled back to the original resolution.

In our actual calculations, we performed one iteration of adaptive variance computation, with parameter settings as follows:

Table S1. Parameter values used for adaptive weight NL-GWR

Symbol	Value	Brief Explanation
p	10	Sparse sampling interval for regression points
r	10	Discrimination radius for regression point sampling
h	10	GWR bandwidth
b	10	Geographic correlation kernel bandwidth for residuals
τ_{σ}	5	Relaxation factor for weight iteration
M_{σ}	100	Maximum number of iterations for weight iteration

After obtaining the variance, it is treated as a prior result for formal NL-GWR velocity solving, employing settings identical to those in the main text method (Table 2), rather than the parameter settings used during the adaptive process. The concluded weight by this method is shown as:

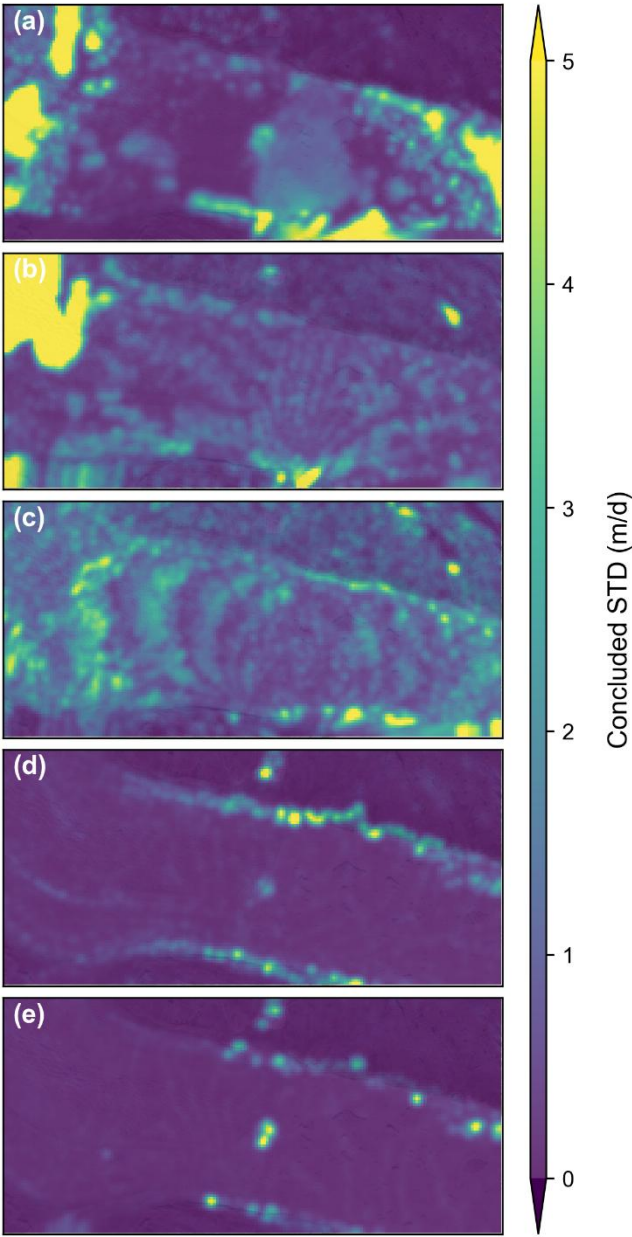


Figure S4. Magnitudes of uncertainty for various observation variable components inferred using the Weight Adaptive Method based on Kernel-Smooth, with random error/standard deviation (STD) serving as the metric,

121 including **(a)** interferometric LOS velocity, **(b-c)** TRI feature-tracking-derived LOS and azimuth velocities, and **(d-e)**
122 Sentinel-2A/B-derived easting and northing velocities.
123
124

References

- [1] J. Hu et al., "3-D movement mapping of the alpine glacier in Qinghai-Tibetan Plateau by integrating D-InSAR, MAI and Offset-Tracking: Case study of the Dongkemadi Glacier," *Global and Planetary Change*, vol. 118, pp. 62-68, 2014.
- [2] C. Chen et al., "Triple collocation-based error estimation and data fusion of global gridded precipitation products over the Yangtze River basin," *Journal of Hydrology*, vol. 605, p. 127307, 2022.
- [3] J. O. Ramsay, "Kernel smoothing approaches to nonparametric item characteristic curve estimation," *Psychometrika*, vol. 56, no. 4, pp. 611-630, 1991.