Supplement of Earth Syst. Sci. Data, 16, 5753–5766, 2024
https://doi.org/10.5194/essd-16-5753-2024-supplement

Supplement of

Global tropical cyclone size and intensity reconstruction dataset for 1959–2022 based on IBTrACS and ERA5 data

Zhiqi Xu et al.

Correspondence to: Jianping Guo (jpguocams@gmail.com)

S1. The hyperparameter selections for each model

S1.1 Random forecast

When optimizing the random forecast model for regression tasks, we utilize Randomized Search Cross-Validation (RandomizedSearchCV) to systematically explore a wide range of hyperparameters.

Specifically, we define hyperparameter distributions that encompass a range of values for the number of trees in the forest (100 to 300), maximum depth of the tree (10 to 30), maximum number of features (1 to 7), minimum number of samples (2 to 10), minimum number of samples (2 to 20), and maximum number of leaf nodes (800 to 1200).

With 500 iterations and 5-fold cross-validation, we search for the optimal hyperparameter combination that minimized the mean squared error (MSE), which is a common choice for regression problems due to its ability to penalize large errors. Leveraging parallel computing, we efficiently fit the model to the training data and obtain the best-performing estimator.

S1.2 Support vector machine

Similarly, we utilize RandomizedSearchCV with an extensive hyperparameter grid to optimize a support vector machine (SVM) regression model.

This grid comprises key parameters such as the regularization parameter C, ranging from 1.0 to 1000, to strike a balance between training error and margin. We explore the kernel, encompassing radial basis function, polynomial, and sigmoid options, with each option influencing the model's decision function. Further, we adjust the influence of each training sample on the decision function through the gamma parameter, which varies from 0.001 to 1. The epsilon parameter (0.01 to 0.1) defines the epsilon-insensitive zone, thereby controlling the margin width.

Leveraging 500 iterations and 5-fold cross-validation, we systematically search for the optimal hyperparameter combination that minimized the MSE, thereby refining the SVM regression model for enhanced performance.

S1.3 Artificial neural network

We also optimize the artificial neural network model by RandomizedSearchCV in our research to tackle the regression task, featuring a series of densely interconnected layers.

For the number of hidden layers, we explore between 1 and 4 layers to find the optimal layer depth that could capture the complexity of the data without causing overfitting. We adjust the size of each layer (number of neurons) between 1 and 500 to optimize the model's expressive power and generalization ability. The Adam optimization algorithm initializes with a learning rate through random search using reciprocal distribution within a wide range (from 0.0001 to 0.01). Across all layers, we employ the rectified linear unit activation function to introduce non-linearity.

To optimize the model's performance, we opt for the MSE as the loss function. During the training phase, the model undergoes rigorous training leveraging 500 iterations and 5-fold cross-validation, ensuring optimal performance and robustness.

S1.4 Convolutional neural network

For performance evaluation of the convolutional neural network, we adopt the MSE loss function. We also optimize the model by RandomizedSearchCV.

Leveraging 500 iterations and 5-fold cross-validation, the model undergoes rigorous training on the designated training dataset. Specifically, we define a range of values for each hyperparameter (e.g., dropout rate between 0.001 and 0.5, learning rate between 0.0001 and 0.01, filter size between 2 and 5, number of filters between 32 and 128, and number of dense units between 64 and 256). The Adam optimizer dynamically adjusts the learning rate throughout training, further enhancing the optimization process.

S1.5 Multivariate linear regression

We also apply a multivariate linear regression model to investigate the relationship between the predictor variables and the response variable. We train the multiple linear regression model on the training set to estimate the coefficients of the linear relationship, which is subsequently used to predict the response variable on the test set.

S2. Typical cyclone radial wind profile models

We utilize six widely used wind field models (Holland, 1980; DeMaria, 1987; Willoughby et al., 2006; Emanuel and Rotunno, 2011; Frisius and Scgönemann, 2013; Chavas et al., 2015) to estimate the reconstructed the radial distances from the cyclone center to locations where sustained wind speeds of 34, 50 and 64 knots (~17, 26, and 33 m/s). In the models, r, R_{max} and V_{max} represent the distance from the cyclone center, radius of maximum wind and maximum sustained wind speed, respectively.

The wind profile model proposed by Holland (1980) is formulated as follows:

$$V(r) = V_{max}\sqrt{\left(\frac{R_{max}}{r}\right)^{b} e^{1-\left(\frac{r}{R_{max}}\right)^{-b}}} \quad \text{(S1)}$$

where V is the wind speed at distance r from the TC center, and b = 2, according to Kowaleski and Evans (2016).

The model developed by DeMaria (1987) is formulated as follows:

$$V(r) = V_{max}\left(\frac{R_{max}}{r}\right) e^{\frac{\frac{1}{c}-\frac{1}{c}\left(\frac{r}{R_{max}}\right)^{c}}{d}} \quad \text{(S2)}$$

where c = 0.63 and d = 1, following Kowaleski and Evans (2016).

The model proposed by Willoughby et al. (2006; hereinafter, W06) is formulated as follows:

$$V(r) = \begin{cases} V_i = V_{max}\left(\frac{r}{R_{max}}\right)^{n}, \; 0 \le r \le R_1 & \text{(S3a)} \\ V_i(1-w) + V_0 w, \;\; R_1 \le r \le R_2 & \text{(S3b)} \\ V_0 = V_{max} e^{-\frac{r-R_{max}}{X_1}}, \quad R_2 \le r & \text{(S3c)} \end{cases}$$

where V_i and V_0 are the tangential wind components in the eye and beyond the transition zone, which lies between $r = R_1$ and $r = R_2$, respectively. The transition zone is defined as the radius of maximum wind from the cyclone inner to outer profiles. w, X_1, and n are the weight function, exponential decay length in the outer vortex, and power law exponent within the eye, respectively. R_2 is the location of the transition zone, and determined by requiring the radial derivative of (S3c) to vanish at $r = R_{max}$. R_1 is the location of the eye, and can be solved as follows:

$$w(\xi) = 126\xi^{5} - 420\xi^{6} + 540\xi^{7} - 315\xi^{8} + 70\xi^{9} \quad \text{(S3d)}$$

$$w\left(\frac{R_{max}-R_1}{R_{max}}\right) = \frac{nX_1}{nX_1 + R_{max}} \quad \text{(S3e)}$$

The model proposed by Emanuel and Rotunno (2011) is formulated as follows:

$$V(r) = \frac{2r(R_{max}V_{max} + 0.5fR_{max}^{2})}{R_{max}^{2} + r^{2}} - \frac{fr}{2} \quad \text{(S4)}$$

where f is the Coriolis parameter.

The model developed by Frisius and Scgönemann (2013) is formulated as follows:

$$V(r) = V_{max}\frac{r}{R_{max}}\left[\frac{2(\frac{R_{max}}{r})^2}{2-(\frac{C_H}{C_d})[1-(\frac{r}{R_{max}})^2]}\right]^{\frac{1}{2}-\frac{C_H}{C_d}} - \frac{fr}{2} \tag{S5}$$

where C_H and C_D are the surface enthalpy transfer and drag coefficients, respectively, and $\frac{C_H}{C_d}$ = 1, according to Frisius and Scgönemann (2013).

The model proposed by Chavas et al. (2015; hereinafter, CLE15) is formulated as follows:

$$(\frac{M_{inner}}{M_m})^{2-\frac{C_k}{C_d}} = \frac{2(\frac{r}{R_{max}})^2}{2-(\frac{C_k}{C_d})+(\frac{C_k}{C_d})(\frac{r}{R_{max}})^2} \tag{S6}$$

$$\frac{\partial M_{outer}}{\partial r} = \frac{C_d(rV)^2}{0.001({r_o}^2-r^2)}$$

where M_{inner}, M_{outer}, and M_m are the angular moment of the inner and outer wind regimes and at R_{max}, respectively; and C_k and C_d are the exchange surface enthalpy and momentum coefficients, respectively.

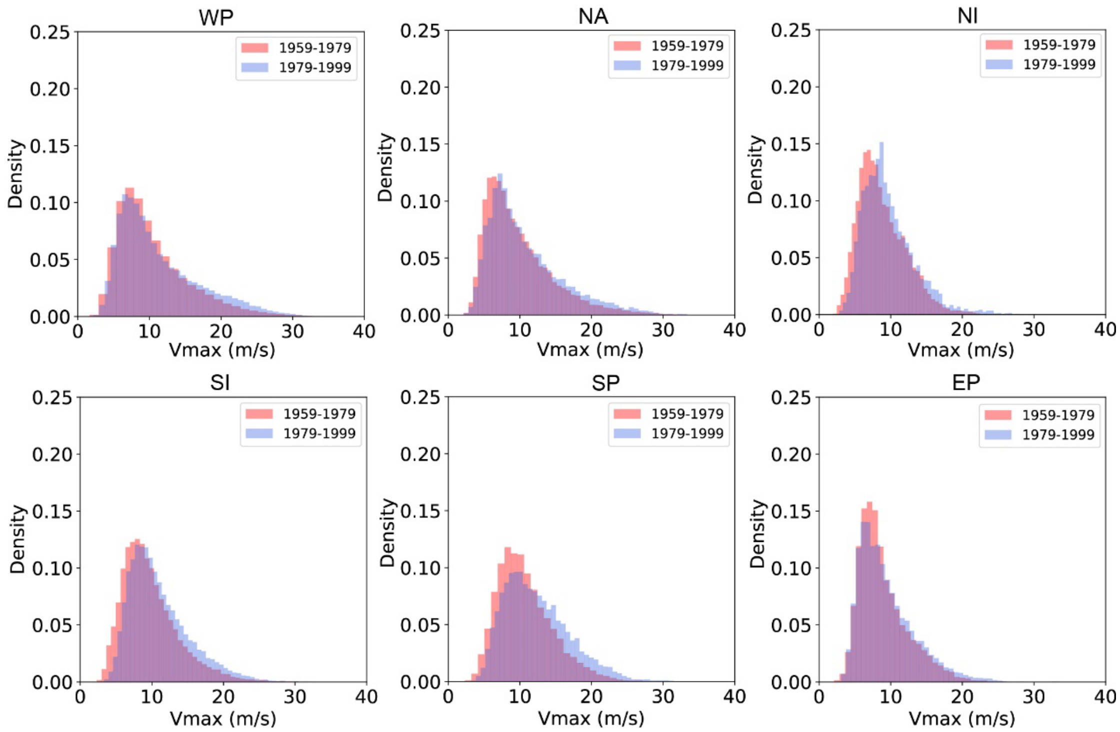

Figure S1: Histogram comparison of ERA5 V_{max} for tropical cyclones from 1959-1979 (red) and from 1979-1999 (blue).

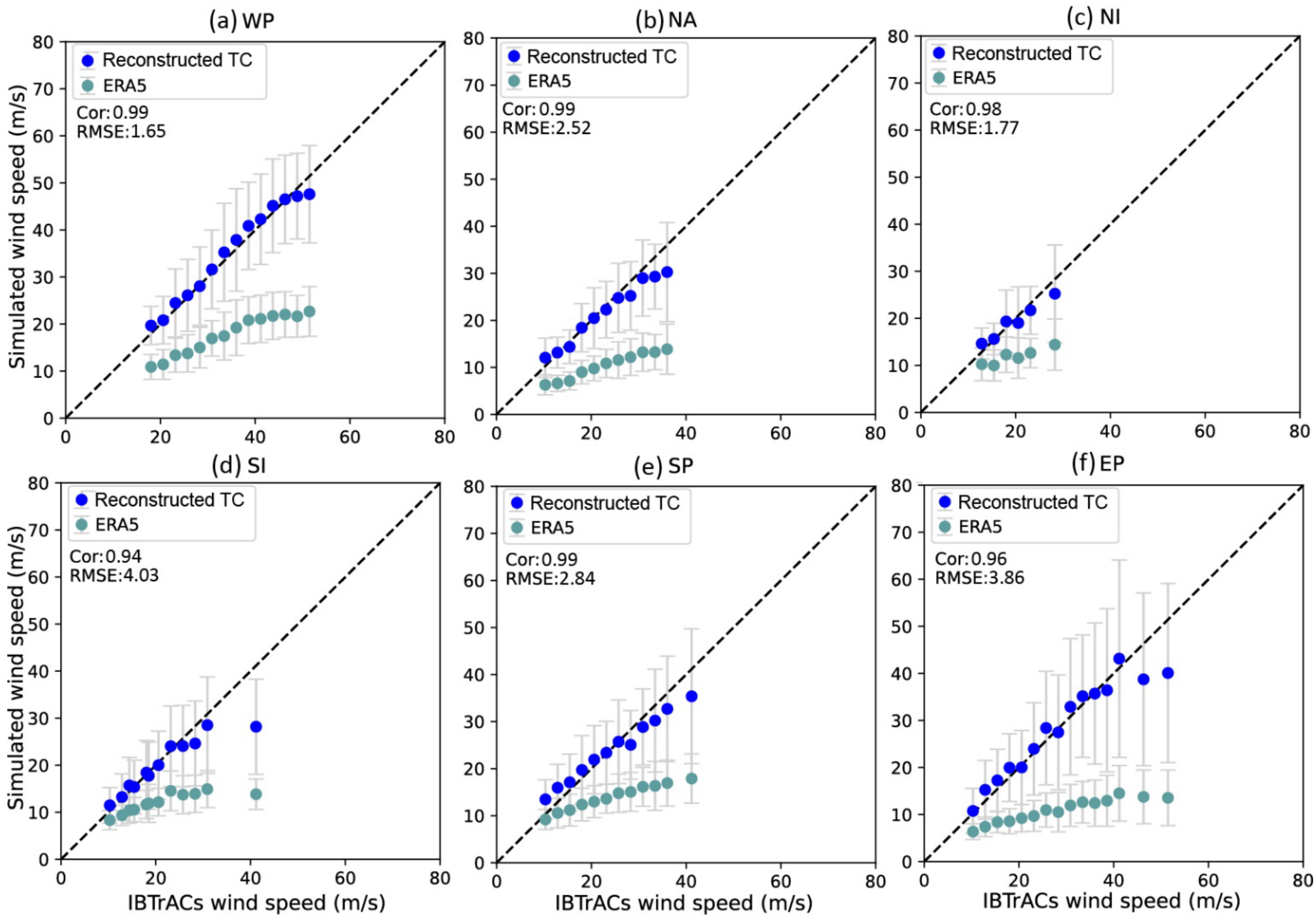

Figure S2: Comparison between value-averaged maximum wind speeds (V_{max}) from ERA5-derived and reconstructed (ERA5 + Random forest) data and IBTrACS maximum wind speeds for tropical cyclones during El Niño years in (a) Western Pacific, (b) North Atlantic, (c) North Indian, (d) South Indian, (e) South Pacific and (f) Eastern Pacific basins. Grey lines represent the error bar, given as one standard deviation from the mean.

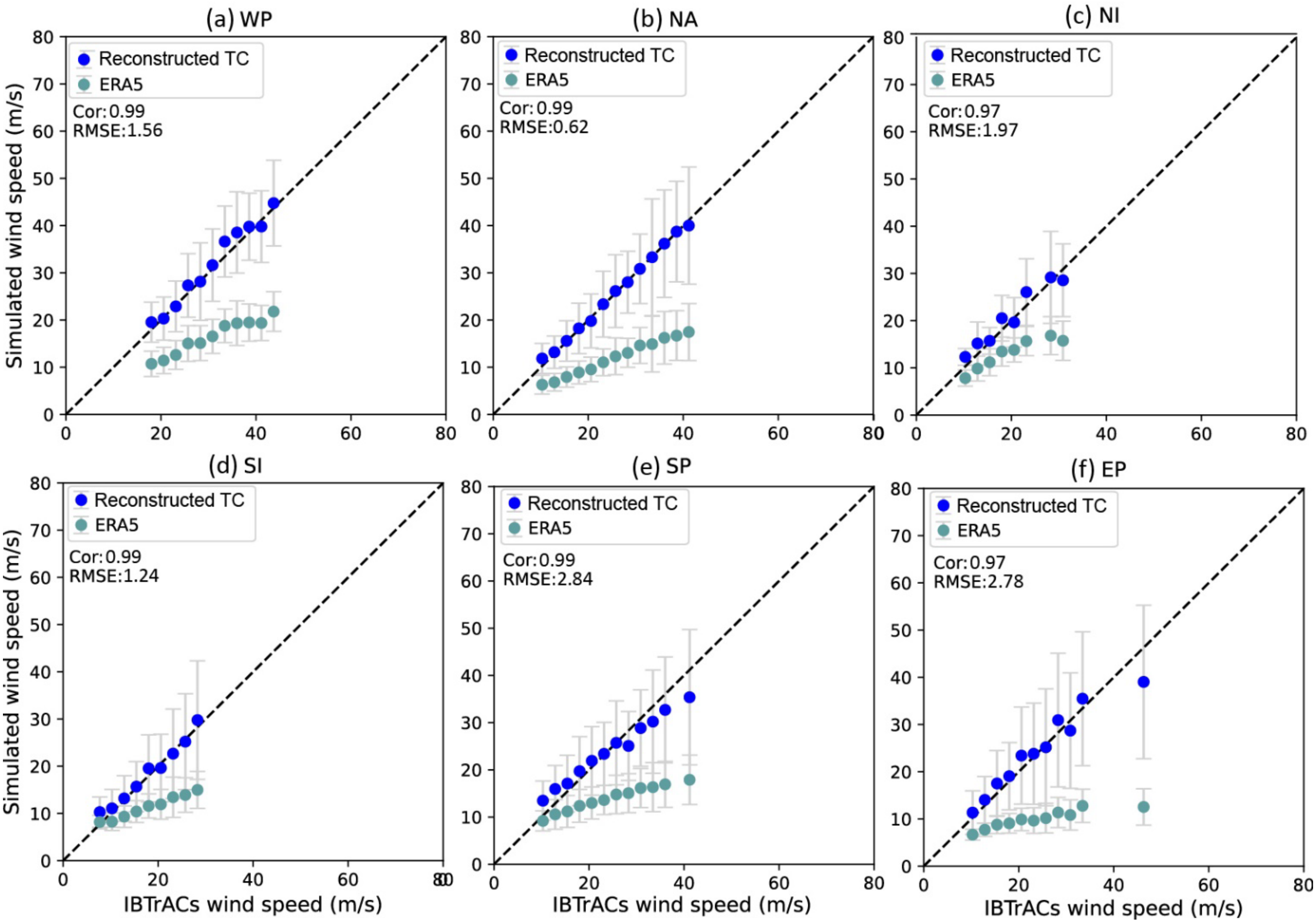

Figure S3: Similar to Figure S2, but during La Niña years.

Table S1. Basic information on the comparison of the reconstructed with the observed R_{34}, R_{50} and R_{64} in Western Pacific. ME, mean errors; MAE, mean absolute error; RMSE, root mean square error; CE, correlation coefficients. H80, D87, W06, E11, F13 and CLE15 refer to the wind field models proposed by Holland (1980), DeMaria (1987), Willoughby et al. (2006), Emanuel and Rotunno (2011), Frisius and Scgönemann (2013) and Chavas et al. (2015)

	ME (km)	MAE (km)	RMSE (km)	CE
$H80_{R34}$	-80.39	89.34	110.54	0.68
$D87_{R34}$	-66.36	82.35	103.63	0.60
$W06_{R34}$	-24.79	46.75	64.54	0.89
$E11_{R34}$	-60.24	73.72	93.24	0.71
$F13_{R34}$	-106.04	110.49	132.57	0.68
$CLE15_{R34}$	-60.60	74.80	93.53	0.72
$H80_{R50}$	-50.42	54.66	64.76	0.54
$D87_{R50}$	-39.40	47.11	56.91	0.52
$W06_{R50}$	-14.60	26.00	33.27	0.82
$E11_{R50}$	-28.29	46.90	56.26	0.25
$F13_{R50}$	-65.31	67.30	77.81	0.50
$CLE15_{R50}$	-39.87	46.93	56.52	0.56
$H80_{R64}$	-32.13	33.66	39.06	0.60
$D87_{R64}$	-23.97	26.75	32.27	0.62
$W06_{R64}$	-14.14	18.28	22.71	0.78
$E11_{R64}$	-26.36	28.71	34.01	0.62
$F13_{R64}$	-41.06	41.86	47.22	0.56
$CLE15_{R64}$	-24.32	27.25	32.73	0.61

Table S2. Similar to Table S1, but in North Atlantic.

	ME (km)	MAE (km)	RMSE (km)	CE
$H80_{R34}$	-50.20	63.41	89.89	0.87
$D87_{R34}$	-22.85	58.87	82.83	0.85
$W06_{R34}$	-41.75	67.06	112.55	0.81
$E11_{R34}$	-19.90	54.17	80.59	0.85
$F13_{R34}$	-91.60	94.33	122.41	0.87
$CLE15_{R34}$	-25.19	53.00	78.77	0.87
$H80_{R50}$	-25.33	43.75	62.34	0.82
$D87_{R50}$	-6.81	45.40	66.86	0.79
$W06_{R50}$	-11.58	32.71	57.39	0.84
$E11_{R50}$	25.41	61.73	89.89	0.82
$F13_{R50}$	-50.70	54.88	72.66	0.83
$CLE15_{R50}$	-10.75	41.50	60.52	0.82
$H80_{R64}$	-15.35	27.55	39.46	0.82
$D87_{R64}$	-6.39	26.98	39.04	0.81
$W06_{R64}$	2.67	18.52	30.37	0.87
$E11_{R64}$	-8.73	26.15	37.58	0.83
$F13_{R64}$	-27.81	31.47	42.17	0.85
$CLE15_{R64}$	-5.52	26.38	37.92	0.83

Table S3. Similar to Table S1, but in North Indian.

	ME (km)	MAE (km)	RMSE (km)	CE
$H80_{R34}$	-71.01	72.12	87.65	0.17
$D87_{R34}$	-62.28	64.64	81.20	0.18
$W06_{R34}$	-23.19	31.19	41.59	0.74
$E11_{R34}$	-59.22	61.89	77.41	0.27
$F13_{R34}$	-88.25	88.42	102.12	0.12
$CLE15_{R34}$	-59.04	61.77	78.04	0.22
$H80_{R50}$	-40.23	41.11	49.98	-0.17
$D87_{R50}$	-32.57	35.01	45.31	-0.24
$W06_{R50}$	-14.66	20.49	25.69	0.63
$E11_{R50}$	-20.03	37.42	43.99	-0.57
$F13_{R50}$	-49.46	50.07	57.74	-0.29
$CLE15_{R50}$	-32.77	34.95	44.63	-0.18
$H80_{R64}$	-24.54	27.28	33.41	-0.18
$D87_{R64}$	-19.30	24.63	30.47	-0.23
$W06_{R64}$	-11.63	16.62	21.17	0.62
$E11_{R64}$	-22.20	25.82	31.92	-0.19
$F13_{R64}$	-29.87	31.81	38.26	-0.47
$CLE15_{R64}$	-18.94	24.54	30.71	-0.33

Table S4. Similar to Table S1, but in South Indian.

	ME (km)	MAE (km)	RMSE (km)	CE
$H80_{R34}$	-52.37	66.93	86.54	0.40
$D87_{R34}$	-39.35	65.36	83.62	0.24
$W06_{R34}$	3.57	45.71	56.68	0.74
$E11_{R34}$	-32.10	58.83	75.18	0.42
$F13_{R34}$	-77.33	82.86	103.96	0.40
$CLE15_{R34}$	-34.01	58.82	75.05	0.46
$H80_{R50}$	-16.33	33.34	42.84	0.06
$D87_{R50}$	-4.94	33.45	41.81	-0.01
$W06_{R50}$	14.35	29.69	36.18	0.46
$E11_{R50}$	10.50	40.17	49.96	-0.13
$F13_{R50}$	-31.74	39.51	50.65	-0.01
$CLE15_{R50}$	-6.01	33.58	41.64	0.02
$H80_{R64}$	-6.23	18.45	23.88	-0.01
$D87_{R64}$	2.01	18.92	23.27	0.05
$W06_{R64}$	9.68	18.54	21.57	0.43
$E11_{R64}$	-0.55	18.31	22.70	0.12
$F13_{R64}$	-17.11	21.82	28.37	-0.06
$CLE15_{R64}$	0.36	18.21	22.42	0.13

Table S5. Similar to Table S1, but in South Pacific.

	ME (km)	MAE (km)	RMSE (km)	CE
H80$_{R34}$	-59.42	67.85	82.65	0.66
D87$_{R34}$	-46.49	61.37	77.34	0.57
W06$_{R34}$	-5.00	33.51	46.25	0.83
E11$_{R34}$	-39.66	53.81	67.68	0.69
F13$_{R34}$	-85.65	88.49	104.22	0.68
CLE15$_{R34}$	-40.36	53.51	67.32	0.71
H80$_{R50}$	-21.77	30.51	36.71	0.64
D87$_{R50}$	-11.07	26.03	32.12	0.63
W06$_{R50}$	11.75	21.53	27.25	0.77
E11$_{R50}$	2.86	29.22	38.00	0.42
F13$_{R50}$	-38.11	41.99	49.05	0.62
CLE15$_{R50}$	-10.19	24.78	31.18	0.65
H80$_{R64}$	-2.51	13.33	16.38	0.64
D87$_{R64}$	5.80	14.17	17.05	0.66
W06$_{R64}$	12.75	15.60	18.56	0.77
E11$_{R64}$	4.97	14.14	17.02	0.68
F13$_{R64}$	-13.60	16.42	20.91	0.66
CLE15$_{R64}$	7.00	14.58	17.05	0.69

Table S6. Similar to Table S1, but in Eastern Pacific.

	ME (km)	MAE (km)	RMSE (km)	CE
H80$_{R34}$	-18.13	48.97	62.66	0.52
D87$_{R34}$	-7.32	55.07	69.07	0.41
W06$_{R34}$	32.25	44.43	51.31	0.81
E11$_{R34}$	4.02	50.50	64.00	0.53
F13$_{R34}$	-43.59	55.90	70.91	0.54
CLE15$_{R34}$	5.97	48.36	61.56	0.57
H80$_{R50}$	-9.60	27.55	36.73	0.32
D87$_{R50}$	-2.51	28.81	38.34	0.27
W06$_{R50}$	27.19	31.77	36.61	0.68
E11$_{R50}$	10.68	35.00	47.12	0.12
F13$_{R50}$	-23.54	31.68	40.92	0.31
CLE15$_{R50}$	1.73	27.70	37.00	0.34
H80$_{R64}$	-5.67	17.90	23.18	0.14
D87$_{R64}$	-0.32	18.69	23.92	0.11
W06$_{R64}$	18.74	21.66	25.24	0.51
E11$_{R64}$	0.59	17.86	22.98	0.19
F13$_{R64}$	-14.41	19.89	25.70	0.12
CLE15$_{R64}$	1.31	18.27	23.54	0.15