Global Underwater Geolocation from Time-Lapse Polarization Imagery Supplementary Material

Supplementary Note-1

Underwater polarization data preprocessing. The raw sensor image is 2048×2448 pixels with a 2×2 superpixel layout that records linear polarization intensities at $0°$, $45°$, $90°$, and $135°$. We extract those four subimages, downsample by $4\times$ to 256×306 pixels, and stack them to obtain a single four-channel frame.

Using the four channels we compute the first three Stokes parameters,

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} I_0 + I_{90} \\ I_0 - I_{90} \\ I_{45} - I_{135} \end{bmatrix}, \qquad \mathrm{AoP} = \tfrac{1}{2} \tan^{-1}\left(\frac{S_2}{S_1}\right), \tag{1}$$

yielding AoP values in the range 0–π.

The omnidirectional lens introduces a radial offset in the reconstructed AoP. Following [3], we correct each pixel (i, j) by

$$\phi_{i,j} = \tan^{-1}\left(\frac{w/2 - j}{h/2 - i}\right), \qquad \mathrm{AoP}_{\mathrm{cal}}(i, j) = \mathrm{AoP}(i, j) - \phi_{i,j}, \tag{2}$$

where h and w are the frame height and width. The calibrated AoP frames, expressed as 32-bit floats, form the input to all subsequent processing.

Supplementary Note-2

Optical model of underwater polarization. Underwater polarization can be described with a single-scattering (Rayleigh) model that depends on only two angles: the solar elevation, or incident angle θ_i, and the scattering angle θ that directs light toward the camera. Unpolarized sunlight, represented by the Stokes vector S_{in}, undergoes three sequential operations—refraction at the air–water interface, single-Rayleigh scattering from suspended particles, and two coordinate rotations—yielding the detected Stokes vector

$$S_{\mathrm{d}} = R_{S\to D}\, M_{\mathrm{scat}}\, R_{R\to S}\, M_{\mathrm{refr}}\, S_{\mathrm{in}}. \tag{3}$$

In this expression, M_{refr} is the Müller matrix for refraction at the interface,

$$M_{\mathrm{refr}} = \begin{bmatrix} \alpha+\beta & \alpha-\beta & 0 & 0 \\ \alpha-\beta & \alpha+\beta & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}, \tag{4}$$

with

$$\alpha = \tfrac{1}{2}\left[\frac{2\sin\theta_t \cos\theta_i}{\sin(\theta_i+\theta_t)\cos(\theta_i-\theta_t)}\right]^2, \quad \beta = \tfrac{1}{2}\left[\frac{2\sin\theta_t \cos\theta_i}{\sin(\theta_i+\theta_t)}\right]^2, \quad \gamma = \frac{4\sin^2\theta_t \cos^2\theta_i}{\sin^2(\theta_i+\theta_t)\cos^2(\theta_i-\theta_t)}. \tag{5}$$

where θ_t is the transmitted angle linked to θ_i by Snell's law, $\sin\theta_i = n \sin\theta_t$, with n the air-to-water refractive-index ratio. The matrix M_{scat} captures single-Rayleigh scattering,

$$M_{\mathrm{scat}} = \tfrac{1}{2}\begin{bmatrix} 1+\cos^2\theta & \cos^2\theta - 1 & 0 & 0 \\ \cos^2\theta - 1 & 1+\cos^2\theta & 0 & 0 \\ 0 & 0 & 2\cos\theta & 0 \\ 0 & 0 & 0 & 2\cos\theta \end{bmatrix}, \tag{6}$$

and each change of reference plane uses the rotation matrix

$$R(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\phi & \sin 2\phi & 0 \\ 0 & -\sin 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

Two practical insights follow. First, the Sun's azimuth appears only inside the rotation matrices; for a $360°$ fisheye system it merely rotates the radial AoP pattern without changing its shape. Second, the shape itself is set by the solar elevation θ_i through Snell's law, and by bulk water optics—the refractive-index ratio n in M_{refr} and the particle size and concentration embedded in M_{scat}.

Supplementary Note-3

Angle of polarization, intensity, and degree of linear polarization at matched solar elevations. Figs. S1–S3

present angle of polarization (AoP), degree of linear polarization (DoLP), and raw-intensity frames recorded at the same five solar elevations (25°, 35°, 45°, 55°, 65°) in two optical end-members: turbid Río Ceballos and ultra-clear Lake Ohrid.

AoP images: Each image was first rotated by the measured solar heading so that all frames share a common reference. Pairs captured hours apart but at the *same* elevation collapse onto an identical radial pattern, whereas frames from different elevations remain distinct (Fig. S1). This confirms that heading contributes only an in-plane rotation, while solar elevation governs the AoP geometry, independent of water clarity.

DoLP images: These intensity differences translate into contrasting DoLP maps (Fig. S2). Peak DoLP reaches 0.65–0.75 in Lake Ohrid but seldom exceeds 0.45 in Río Ceballos, and the depolarized zone around the Sun is broader in the turbid water, indicating stronger multiple scattering. In both sites the high-DoLP ring lies $\sim 90^\circ$ from the Sun and rotates rigidly between morning and afternoon, mirroring the single-scattering model.

Intensity images: Both sites show the characteristic brightening toward the Sun (Fig. S3), yet Lake Ohrid exhibits a sharper specular lobe and darker background, visually reflecting its greater clarity and weaker multiple scattering.

These images show that AoP—set chiefly by solar elevation—is robust to water type, whereas intensity and DoLP vary strongly with optical clarity. This underlines why AoP is the primary cue for our physics-guided simulator and transformer network, and why intensity alone is unreliable for geolocation in optically diverse waters.

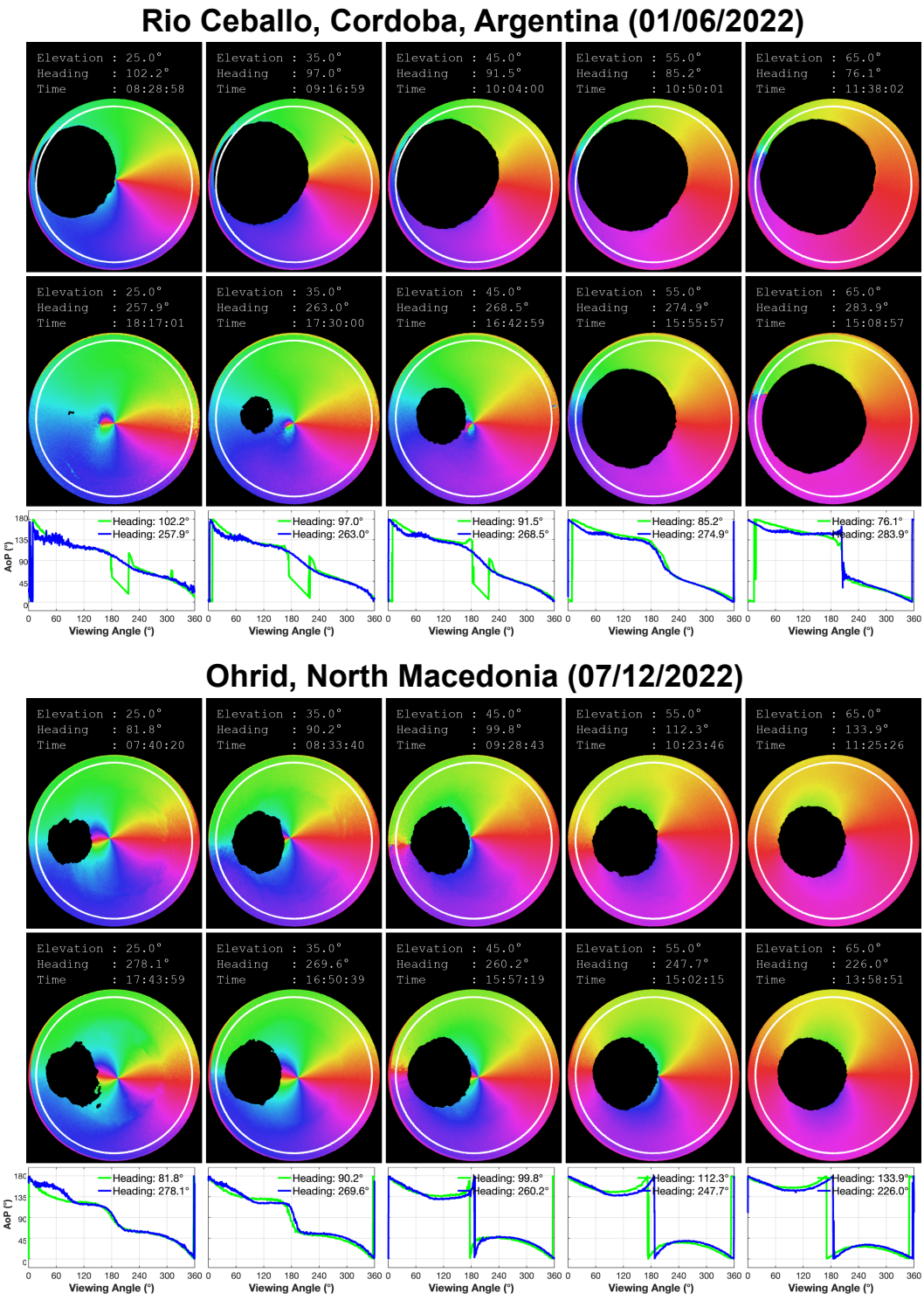

Figure S1. **Solar elevation, not heading, shapes the underwater angle of polarization (AoP). Rows 1-3,** Río Ceballos (turbid water). **1**, AoP frames at five morning elevations (25°, 35°, 45°, 55°, 65°); **2**, matching elevations recorded hours later, after the Sun has moved in azimuth. All images are rotated by the measured solar heading so that they share a common reference direction. **3**, AoP traces sampled along the white ring in each frame: curves recorded at the same elevation (colors) coincide, whereas curves from different elevations remain distinct. **Rows 4-6,** Lake Ohrid (clear water) in the same morning–afternoon–plot order. Despite the large contrast in turbidity, morning and afternoon pairs align once heading is removed, confirming that AoP geometry depends chiefly on solar elevation. The central gray mask marks saturated pixels around the solar disc and is excluded from analysis.

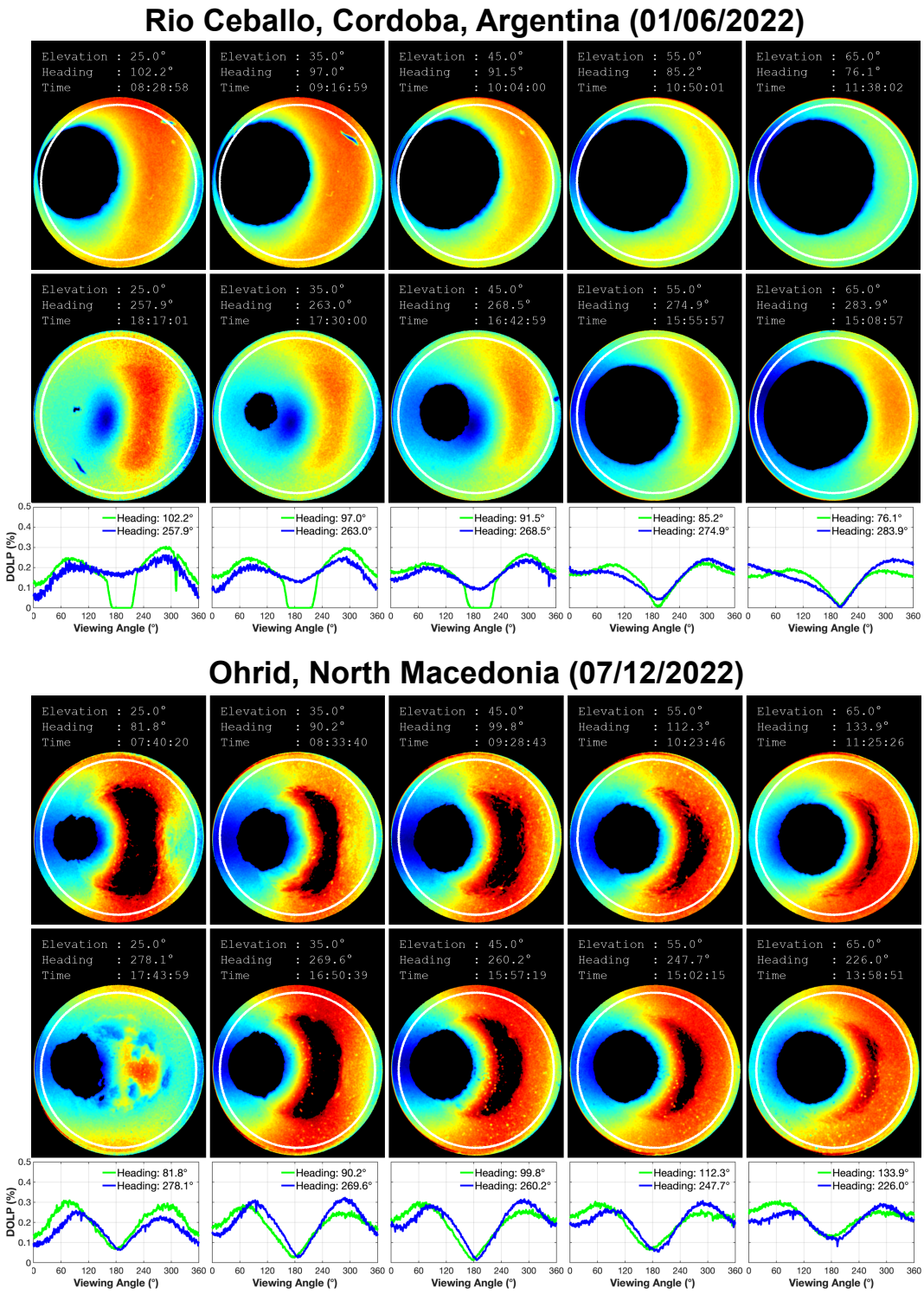

Figure S2. **Degree of linear polarization distinguishes clear and turbid waters.** Each row shows five solar elevations—25°, 35°, 45°, 55°, 65°—captured once in the morning (row 1) and again in the afternoon (row 2). **Top,** Río Ceballos, Argentina. Peak DoLP seldom exceeds 0.45 and the depolarized zone around the Sun is broad, indicating strong multiple scattering in turbid water. **Bottom,** Lake Ohrid, North Macedonia. Peak DoLP reaches 0.65–0.75 and the high-DoLP ring $\sim 90^\circ$ from the Sun is sharply defined, reflecting clearer water and weaker scattering. Line profiles extracted along the outer rings confirm the contrast: at every elevation the Ohrid curve lies well above its Río Ceballos counterpart. Morning and afternoon maps differ only by an in-plane rotation, showing that solar heading alters orientation but not DoLP magnitude, in agreement with single-scattering model.

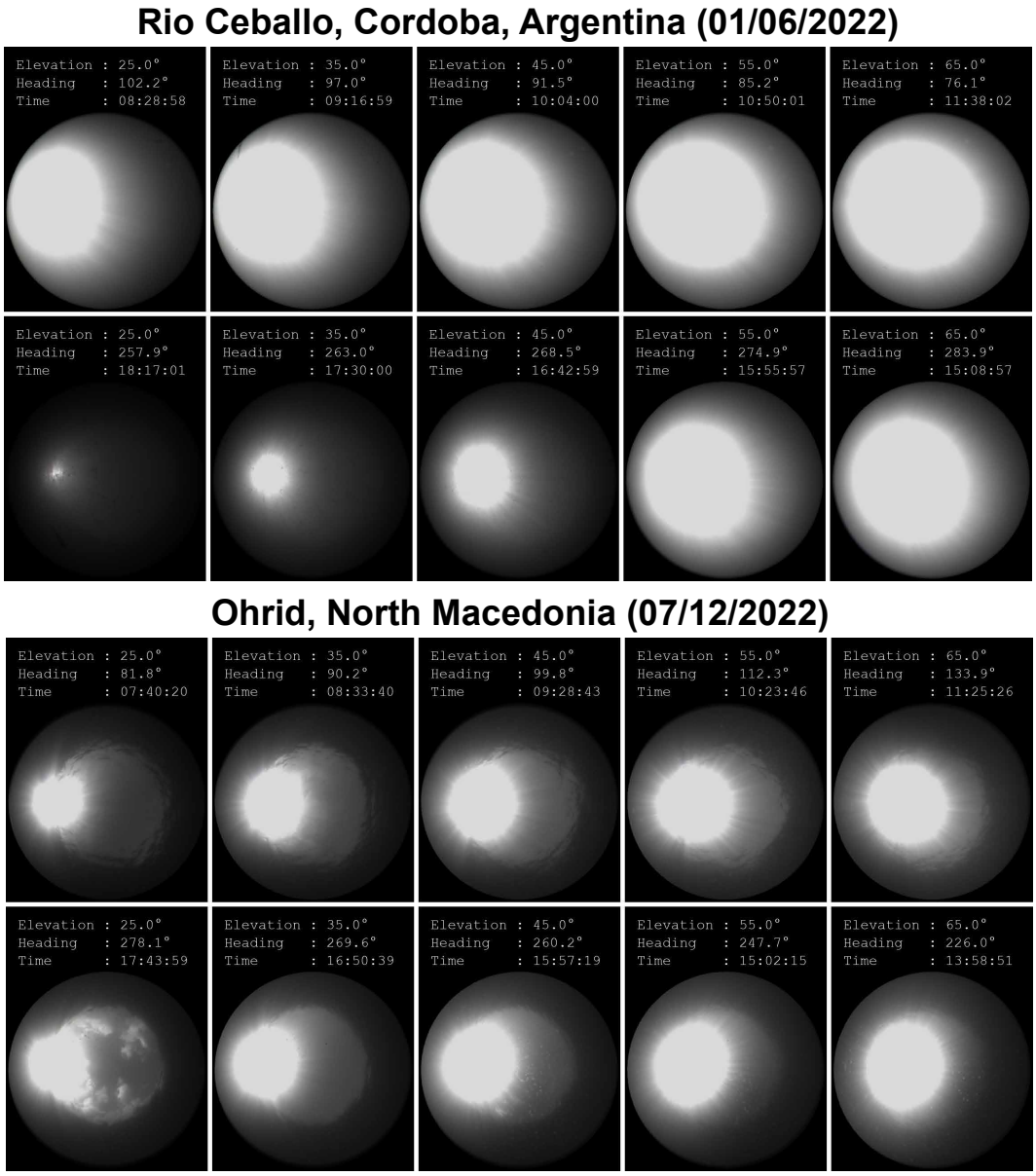

Figure S3. **Raw intensity at matched solar elevations.** Each block pairs the AoP examples in Fig. S1 with their corresponding intensity frames, sampled at identical solar elevations (25°, 35°, 45°, 55°, 65°). **Top,** Río Ceballos, Argentina (turbid water). **Bottom,** Lake Ohrid, North Macedonia (clear water). The bright, saturated Sun patch traverses the fisheye field as heading changes, while the underlying radial illumination pattern remains similar. Lake Ohrid shows a sharper specular lobe and darker background than Río Ceballos, revealing its higher clarity; the flatter contrast in Río Ceballos reflects stronger multiple scattering. These site-specific amplitude differences underscore why intensity alone is an unreliable geolocation cue and motivate our focus on AoP, which is far less sensitive to scattering strength.

Supplementary Note-4

Physics-guided simulator and curated training corpus. Fig. S4 shows representative AoP frames from the five training sites and one sample sequence produced by the simulator (the Florida Keys frames appear only in the test set and are omitted here).

Supplementary Note-5

Solar-trajectory variation with latitude, longitude, and season. Fig. S5 illustrates how solar-elevation curves vary with geography and time of year. The center–top block plots solar-elevation curves for 2 January and 2 July at those latitude offsets, holding longitude fixed. As latitude increases, the midday peak drops from 32° to 12° and daylight shortens by almost four hours, showing that latitude chiefly scales the height and duration of the curve. The center–bottom block presents curves for the same two dates at five longitudes, holding latitude fixed; the curves shift horizontally by up to two hours yet keep the same peak height, revealing that longitude mainly translates the curve in time. The right panel demonstrates the stratified 64-point sampling applied to one trajectory (48.089°N, $-88.316°$, Day 1). Red dots indicate the evenly spaced timestamps that are paired with the nearest polarization frames in the empirical library. This sampling guarantees uniform temporal coverage regardless of daylight length and produces training sequences that remain physically consistent while varying in heading and water type.

Supplementary Note-6

Selecting polarization data by solar elevation. To build realistic cross-site training sets we first simulate year-long solar-elevation curves for uniformly sampled locations within the $20° \times 20°$ spherical triangle centered on each held-out site (panels A of Figs. S6–S11). For every elevation sample in these curves we select the polarization frame in our database whose solar elevation is closest, excluding all frames from the test site; the resulting target-to-frame match quality is *summarized* in panels C. For four of the six experiments—Argentina, Hawaii, Florida Keys, and Tampa—more than 95 % of matches lie within 0.01° (gray boxes), while the worst outliers occur only in bins that are poorly represented in the training pool; for example, the Tampa experiment shows errors up to 8° in the $> 80°$ bin because Tampa itself supplies most of the high-elevation data and is withheld from training. In North Macedonia the maximum deviation is 0.5°, and in Champaign it is 1°, both confined to the $< 10°$ bin. These statistics show that the elevation-based selection delivers near-exact matches for the bulk of the trajectory and degrades gracefully at the extremes, demonstrating that the network *generalizes* well to unseen waters.

Volume and diversity of polarization data. Panels D–F in the same figures quantify how much real data each non-test site contributes to the simulated set. For example, when Argentina is the test site, Champaign supplies 39.6 % of all frames used, followed by North Macedonia at 20.3 % (Fig. S6, panel F); 98.3 % of Champaign's unique frames and 98.9 % of North Macedonia's are actually selected (panel E). Similar *utilization* rates—between 94 % and 99 %—are observed in every cross-site split, indicating that the sampler exploits nearly the entire inventory of available data. As the geographic triangle widens, the simulator reuses an increasing fraction of frames, which boosts sample diversity but could lead to over-saturation unless model capacity scales accordingly. These observations underscore the value of continued field collection at underrepresented elevations ($< 10°$ and $> 80°$) and in additional water types to further enhance simulation fidelity and *generalization* performance.

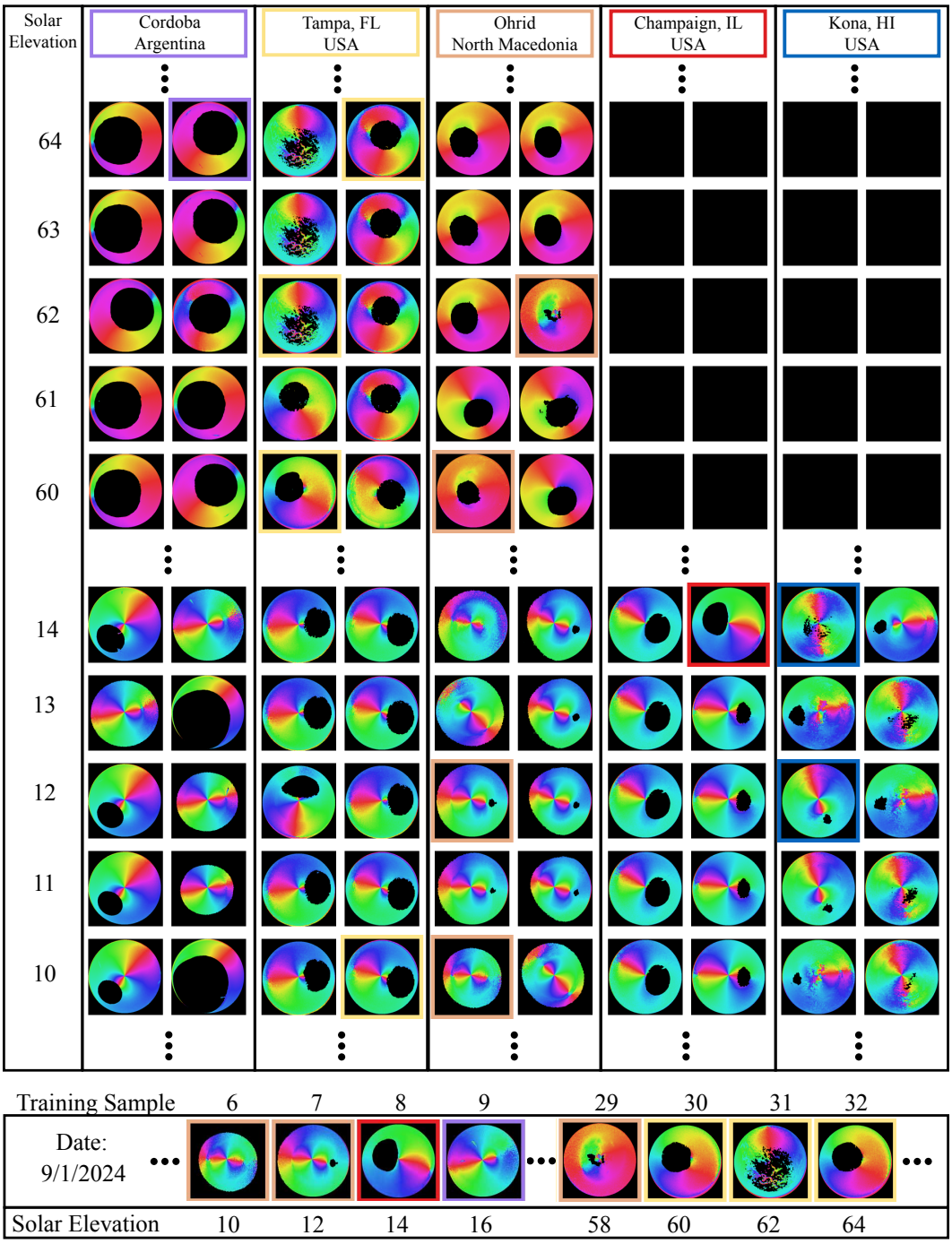

Figure S4. **Simulated training sequences assembled from real polarization images. Top rows:** Representative AoP frames from the five training sites—Lake Ohrid, Champaign, Tampa Bay, Río Ceballos, and Kona—sorted by solar elevation. Data from the Florida Keys (the test site) are withheld. Black squares in the Champaign and Kona columns mark elevations that could not be recorded because the Sun never reached those angles during our collection period. **Bottom row:** Portion of a 64-frame simulated sequence generated for a random location on September 1, 2024. For each requested elevation the simulator draws a frame at a random heading and from a random site, so the masked solar patch drifts within the image and water types vary across the sequence. This diversity suppresses heading and water-type biases and prevents the network from overfitting to any single optical regime, contributing to the accuracy gains reported in the paper.

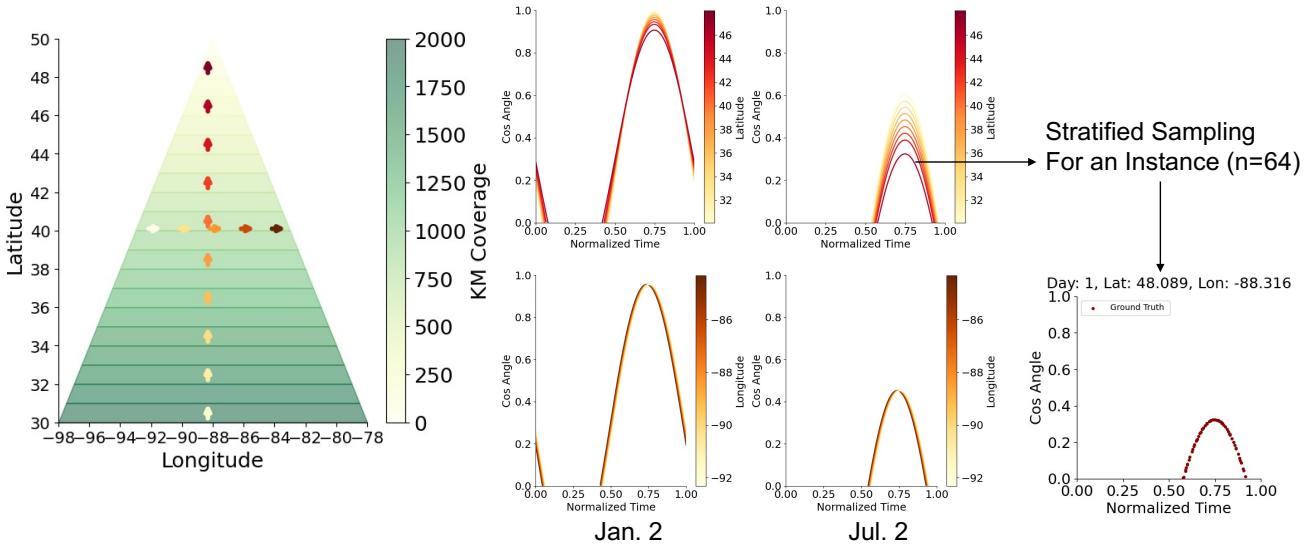

Figure S5. **Latitude, longitude, and season shape solar trajectories. Left**, Equal-area projection of the $20° \times 20°$ spherical triangle centered on Champaign. Arrows indicate selected locations whose solar elevation curves are shown in the right. **Center-top**, Simulated solar trajectories for the arrowed locations on 2 January and 2 July. Varying latitude (longitude fixed) shifts the curves vertically, with higher latitudes lowering the midday peak. **Center-bottom**, Varying longitude (latitude fixed) translates the curves horizontally without changing peak height. **Right**, Stratified 64-point sampling along one trajectory (48.089°N, $-88.316°$, Day 1); red dots mark evenly spaced timestamps that are paired with the nearest polarization frames in the empirical library. This scheme guarantees uniform temporal coverage irrespective of day length and produces physically consistent training sequences while varying heading and water type.

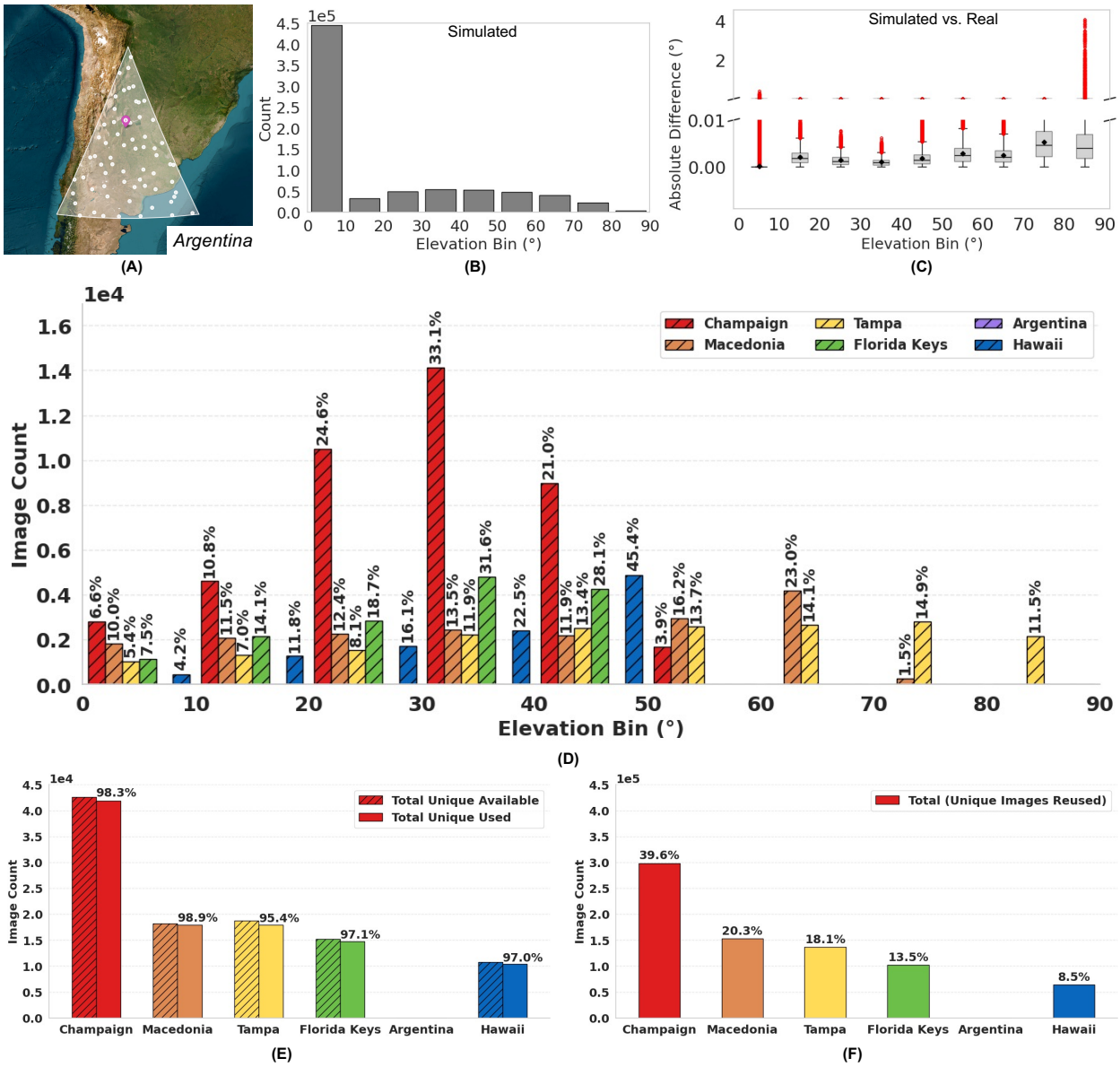

Figure S6. **Statistics of the simulated training set for the Río Ceballos (Argentina) cross-site experiment.** **(A)** Spherical triangle (gray) centered on the Río Ceballos collection site; random points within the triangle provide training locations whose solar trajectories are simulated for different days of the year. **(B)** Histogram of the resulting solar-elevation samples, binned in 10° intervals from 0° to 90°. **(C)** Absolute error between each target elevation in (B) and the nearest real AoP frame used to assemble a training sequence. More than 95 % of matches are within 0.01°; the largest outliers ($>4°$) occur only in the highest elevation bin ($>80°$). **(D)** Availability of polarization frames for the five non-test sites, binned by their solar-elevation labels. **(E)** Total number of unique frames per site (open bars) and the subset selected for simulation (solid bars). **(F)** Site-wise contribution of frames to the final simulated training set; as required for cross-site evaluation, no frames from Argentina are included.

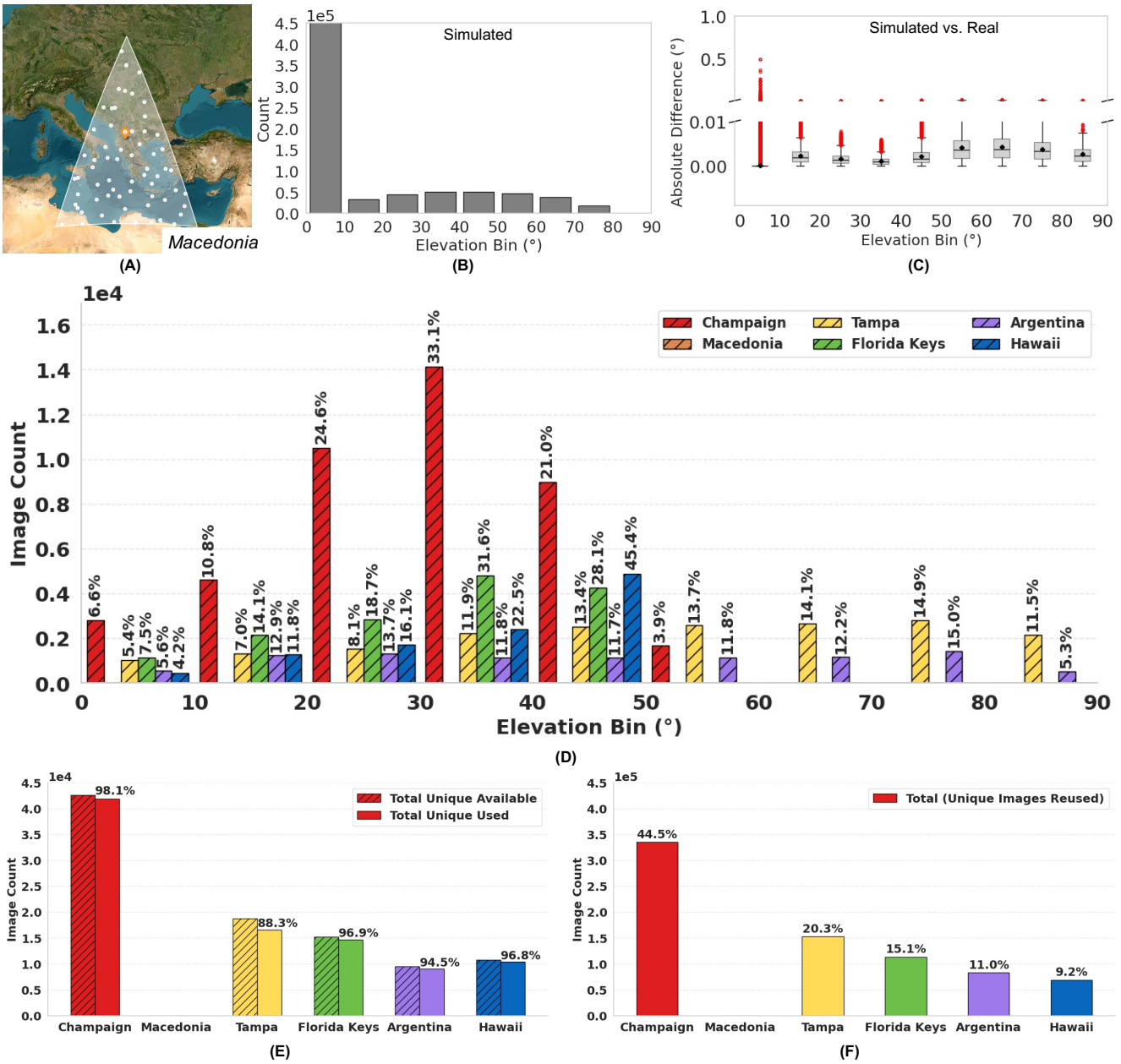

Figure S7. **Statistics of the simulated training set for the Lake Ohrid (North Macedonia) cross-site experiment.** **(A)** Spherical triangle (gray) centered on Lake Ohrid; random points inside the triangle supply training locations, and a year-long solar-elevation curve is generated for each point. **(B)** Histogram of all target elevations, binned in 10° intervals from 0° to 90°. **(C)** Absolute error between every target elevation in (B) and the nearest real AoP frame used to build a sequence. More than 98 % of matches are within 0.01°; the largest deviations (up to 0.5°) appear only in the lowest bin ($< 10°$). **(D)** Availability of polarization frames at the five non-test sites, binned by solar-elevation label. **(E)** Total unique frames per site (open bars) and the subset selected for simulation (solid bars). **(F)** Site-wise contribution of frames to the final simulated training set; no data from North Macedonia are included, as required for cross-site evaluation.

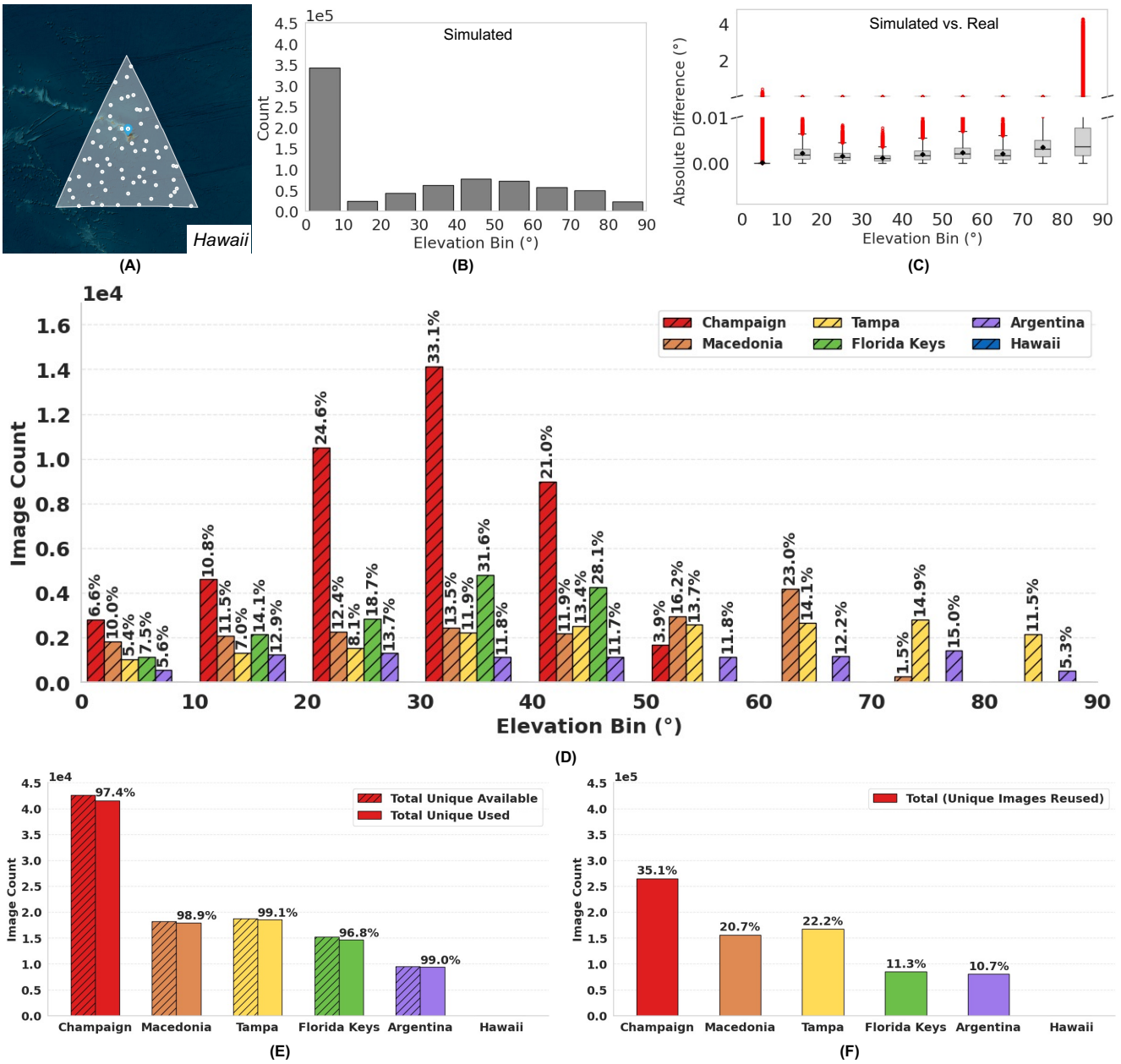

Figure S8. **Statistics of the simulated training set for the Kona (USA) cross-site experiment. (A)** Spherical triangle (gray) centered on the Kona collection site; random points within the triangle serve as training locations, and year-long solar-elevation curves are generated for each point. **(B)** Histogram of target elevations, binned in 10° increments from 0° to 90°. **(C)** Absolute error between each target elevation in (B) and the closest real AoP frame used during sequence assembly. More than 95 % of matches fall within 0.01°; the largest deviations (4°) appear only in the highest bin (> 80°). **(D)** Availability of polarization frames at the five non-test sites, grouped by solar-elevation bin. **(E)** Total frame count per site (open bars) and the subset chosen for simulation (solid bars). **(F)** Contribution of each non-test site to the final simulated training set; no Hawaii data are included, as required for cross-site evaluation.

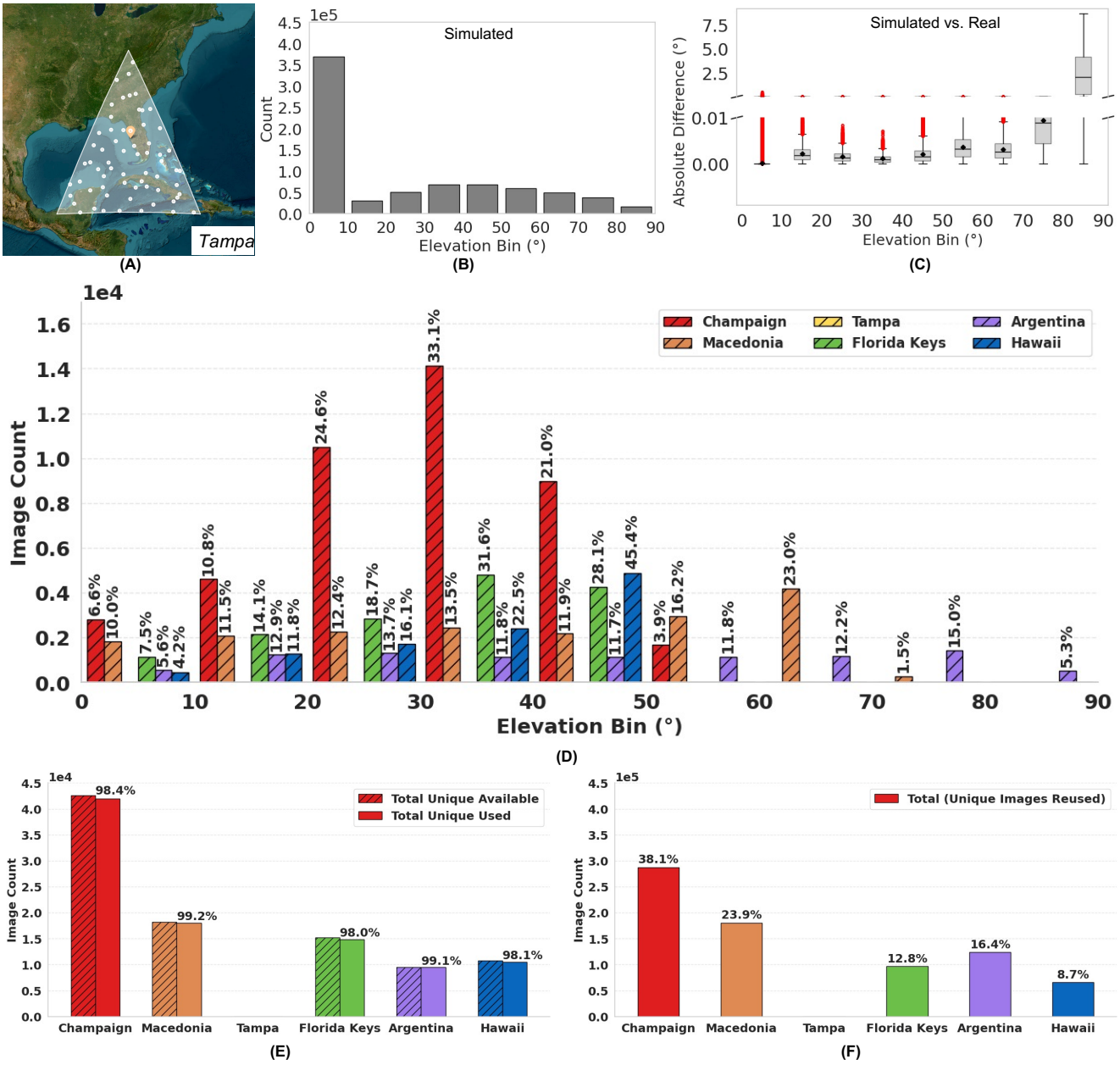

Figure S9. **Statistics of the simulated training set for the Tampa Bay (USA) cross-site experiment.** **(A)** Spherical triangle (gray) centered on the Tampa Bay collection site; random points inside the triangle provide training locations, and solar-elevation curves are generated for every day of the year. **(B)** Histogram of target elevations, binned in 10° intervals from 0° to 90°. **(C)** Absolute error between each target elevation in (B) and the nearest real AoP frame used during sequence assembly. More than 95 % of matches lie within 0.01°, while deviations up to 8° appear only in the highest bin ($>80°$). **(D)** Availability of polarization frames at the five non-test sites, grouped by solar-elevation bin. **(E)** Total frame count per site (open bars) and the subset selected for simulation (solid bars). **(F)** Contribution of each non-test site to the final simulated training set; no Tampa data are included, as required for cross-site evaluation.

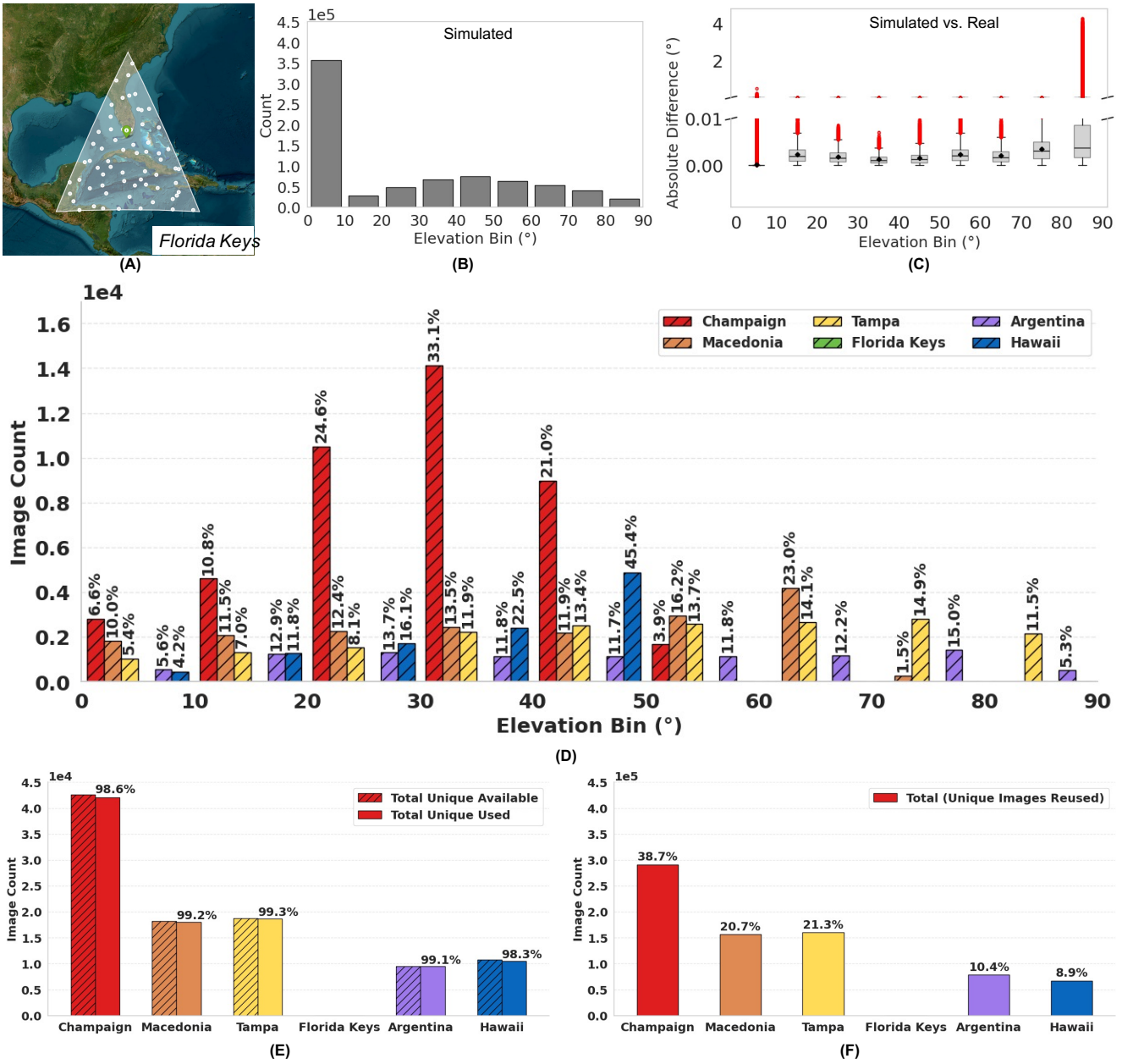

Figure S10. **Statistics of the simulated training set for the Florida Keys (USA) cross-site experiment.** **(A)** Spherical triangle (gray) centered on the Florida Keys collection site; random points inside this triangle supply training locations, and solar-elevation curves are generated for every day of the year. **(B)** Histogram of target elevations, binned in 10° intervals from 0° to 90°. **(C)** Absolute error between each target elevation in (B) and the closest real AoP frame used to build a sequence. More than 95 % of matches fall within 0.01°, while deviations up to 4° occur only in the highest bin ($> 80°$). **(D)** Availability of polarization frames at the five non-test sites, grouped by solar-elevation bin. **(E)** Total frame count per site (open bars) and the subset chosen for simulation (solid bars). **(F)** Contribution of each non-test site to the final simulated training set; no Florida Keys data are included, as required for cross-site evaluation.

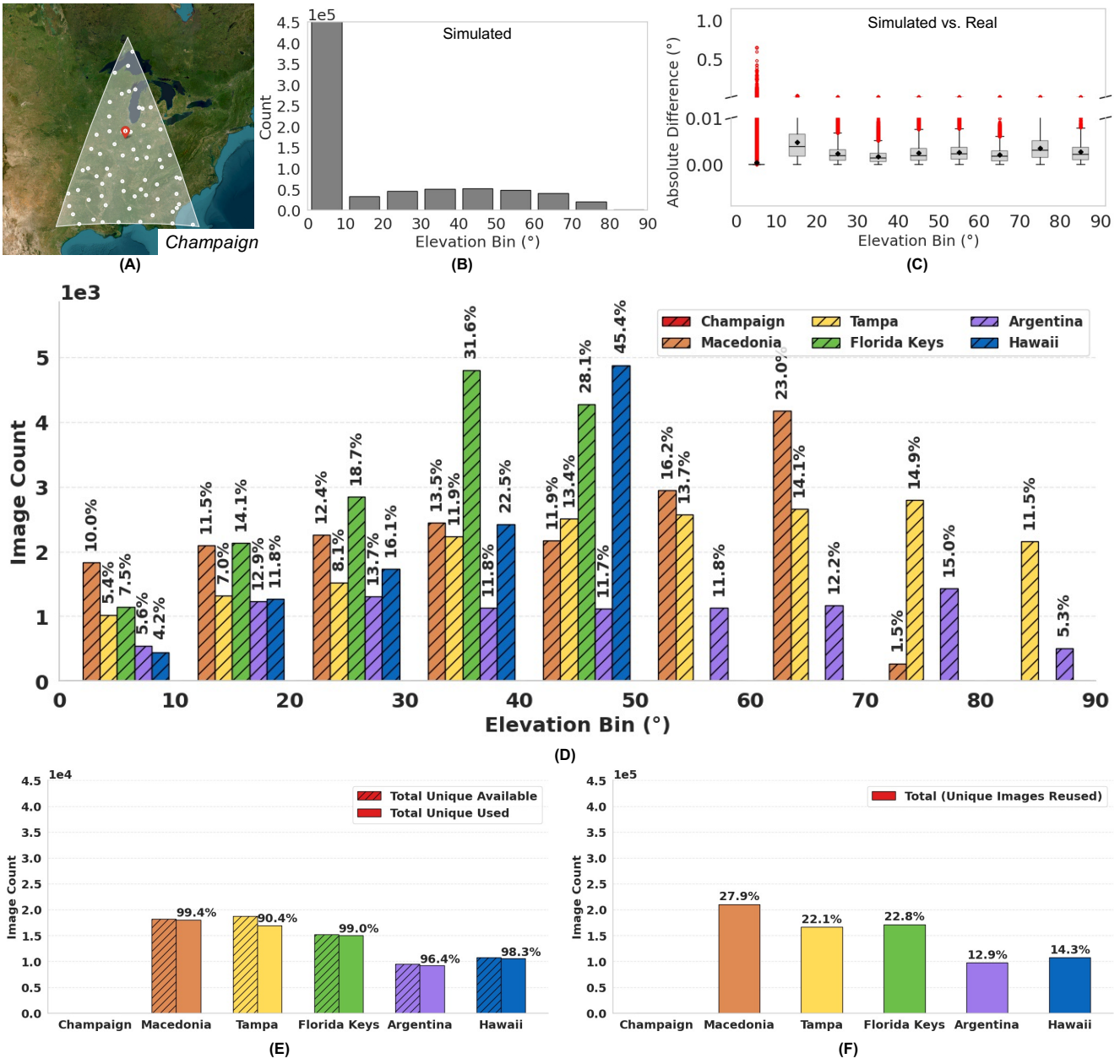

Figure S11. **Statistics of the simulated training set for the Champaign (USA) cross-site experiment.** **(A)** Spherical triangle (gray) centered on the Champaign collection site; random samples within the triangle define training locations, and full-year solar-elevation curves are generated for each point. **(B)** Histogram of target elevations, binned in 10° increments from 0° to 90°. **(C)** Absolute error between every target elevation in (B) and the nearest real AoP frame used during sequence assembly. More than 95 % of matches are within 0.01°; the largest deviations (up to 1°) appear only in the lowest bin ($< 10°$). **(D)** Availability of polarization frames at the five non-test sites, grouped by solar-elevation bin. **(E)** Total number of unique frames per site (open bars) and the subset selected for simulation (solid bars). **(F)** Contribution of each non-test site to the final simulated training set; no Champaign data are included, as required for cross-site evaluation.

Supplementary Note-7

Architecture and optimization details. Our model comprises two coupled modules trained end-to-end. Each frame in the input sequence is processed by a lightweight CNN stem (two 3×3 stride-2 convolutions, channels 1→64→128, ReLU), followed by adaptive average pooling and a linear projection to a 121-D embedding. This embedding is concatenated with temporal features (normalized time-of-day, sine/cosine day-of-year; 3-D) to form a 124-D token. We prepend a learnable `[summary]` token, add a learned positional embedding to the resulting 65-token sequence, apply dropout (p=0.1), and feed the sequence to a 4-layer transformer encoder (4 heads, hidden size d_{model}=124, feedforward size 256, GELU, pre-norm). After a final LayerNorm, the output is concatenated back to each input token, yielding 248-D representations decoded by a 4-layer MLP (64→64→32→1, ReLU) to predict solar elevation for every frame. The second module treats the 64 per-frame predictions as a point set with 4-D features (predicted angle, normalized time, sin-day, cos-day) and maps them to a 3-D geolocation. We employ a Point Transformer with 3 blocks, neighborhood size 16, and width 512. We optimize all parameters jointly with Adam (initial learning rate 10^{-3}) and a One-Cycle schedule peaking at 10^{-2} (warm-up fraction 0.3) over 100 epochs, using batch size 32. The loss is the weighted sum with $\lambda_{\text{elev}} = 2$ and $\lambda_{\text{geo}} = 1$. Within the transformer we apply dropout (p=0.1). Optimizer betas and initialization follow PyTorch defaults.

Supplementary Note-8

Site-resolved geodesic error maps. Figs. S12 and S15 zoom in on each test site to *visualize* the median geodesic error circles for Polar Transformer, SecTran-MIM, and RI-ResNet-RDM. Each circle is drawn on satellite imagery, centered on the site's true coordinates, and its radius equals the method's median distance error for that site. Fig. S12 covers the cross-site task; our circles range from 275 km at Lake Ohrid to 829 km at Tampa, whereas SecTran-MIM spans 1,733–2,815 km and RI-ResNet-RDM exceeds 2,000 km everywhere it is reported. Fig. S15 repeats the analysis for the same-site task and shows an even sharper contrast. Polar Transformer contracts the error to single- or double-digit kilometers (1 km at Río Ceballos, 2.8 km at Tampa, 5.4 km at Champaign, 9.2 km at Lake Ohrid), while RI-ResNet-RDM and SecTran-MIM remain two orders of magnitude larger, between 96 km and 515 km. No baseline results are available for Río Ceballos and Kona; at these sites, our model still attains 1 km and 27 km accuracy, respectively.

These maps translate the quantitative gains reported in the main paper into an intuitive geographic scale: what was a continental-sized uncertainty for the baselines shrinks to a regional (cross-site) or neighborhood (same-site) footprint with Polar Transformer.

Supplementary Note-9

Directional error analysis. Fig. S13 reports the north–south (latitude) errors, while Fig. S14 presents the east–west (longitude) errors for the cross-site evaluation. Polar Transformer maintains tight latitude accuracy: median errors range from roughly 250 km at the Florida Keys and Kona sites to about 650 km at Río Ceballos, yielding a six-site average of 430 km. By contrast, the best published baselines are an order of magnitude worse. SecTran-MIM reports a mean latitude error of 2,032 km and RI-ResNet-RDM rises to 3,429 km, values computed on the four sites for which baselines report.

Longitude results show an even wider gap. Polar Transformer's median errors lie between 50 km at Río Ceballos and less than 200 km at the Florida Keys, for a six-site average of 181 km. SecTran-MIM averages 867 km, while RI-ResNet-RDM climbs to 1,813 km on the reported sites. The figures use the same color code for the six test locations; baseline panels are plotted at eight times the vertical scale of the Polar Transformer panels, yet several RI-ResNet-RDM outliers still reach the axis limit. Black diamond markers denote the site averages. Together, these results show that the transformer's trajectory-level representation sharply reduces systematic latitude bias and longitude drift, achieving sub-degree accuracy even when all data from the test water are withheld during training.

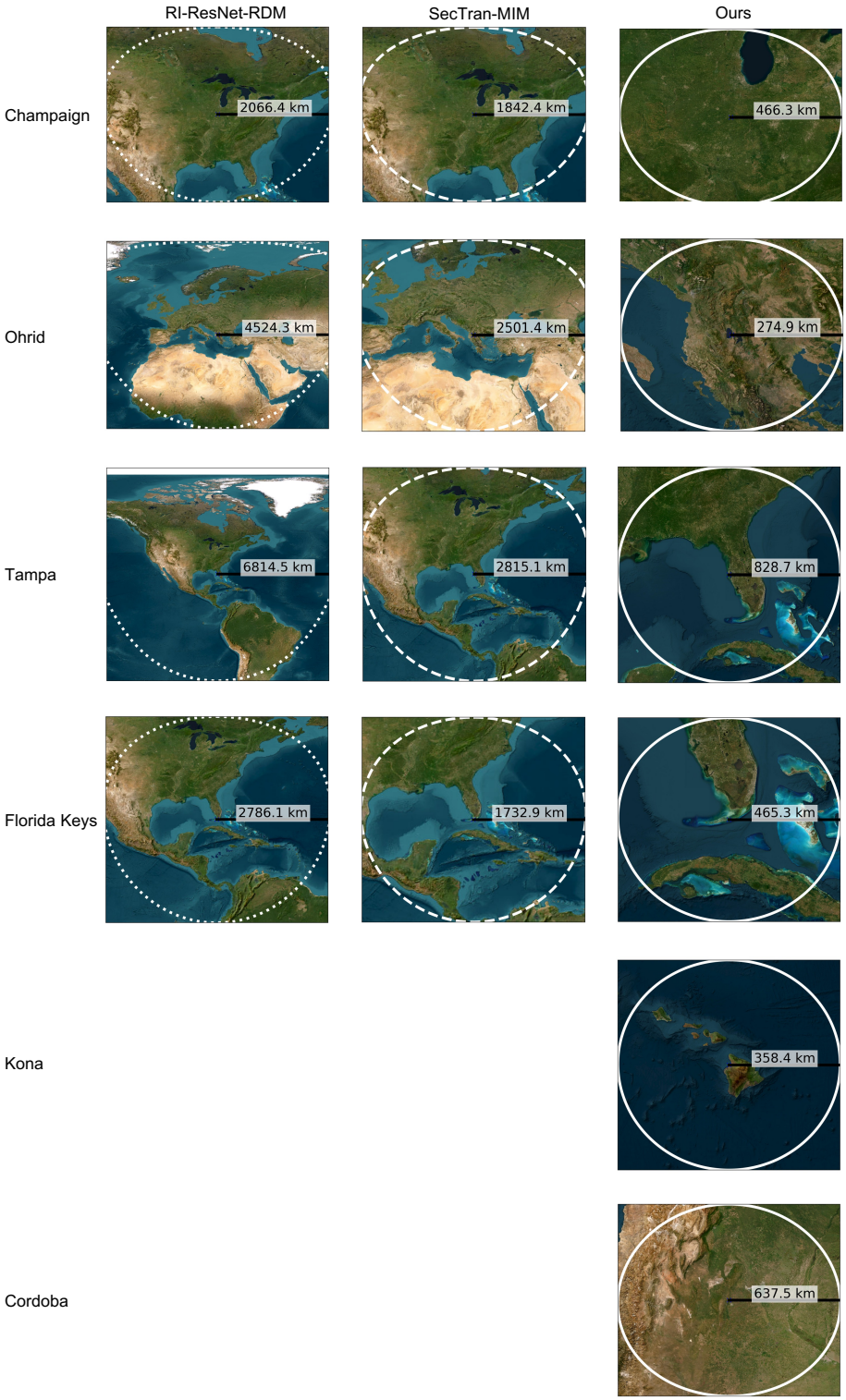

Figure S12. **Zoomed geodesic-error maps for the cross-site evaluation.** Each panel shows the median error circle for Polar Transformer (solid), SecTran-MIM [4] (dashed), and RI-ResNet-RDM [2] (dotted) at the corresponding test site. The circle is centered on the site's true coordinates, and its radius equals the method's median great-circle distance error (in kilometers) over all test sequences. Where a baseline is missing (Kona and Río Ceballos), no result was reported in the original publication. Across the six sites, Polar Transformer consistently shrinks the error circle from continental to regional scale, reducing the median distance by a factor of four relative to SecTran-MIM and by an order of magnitude relative to RI-ResNet-RDM.

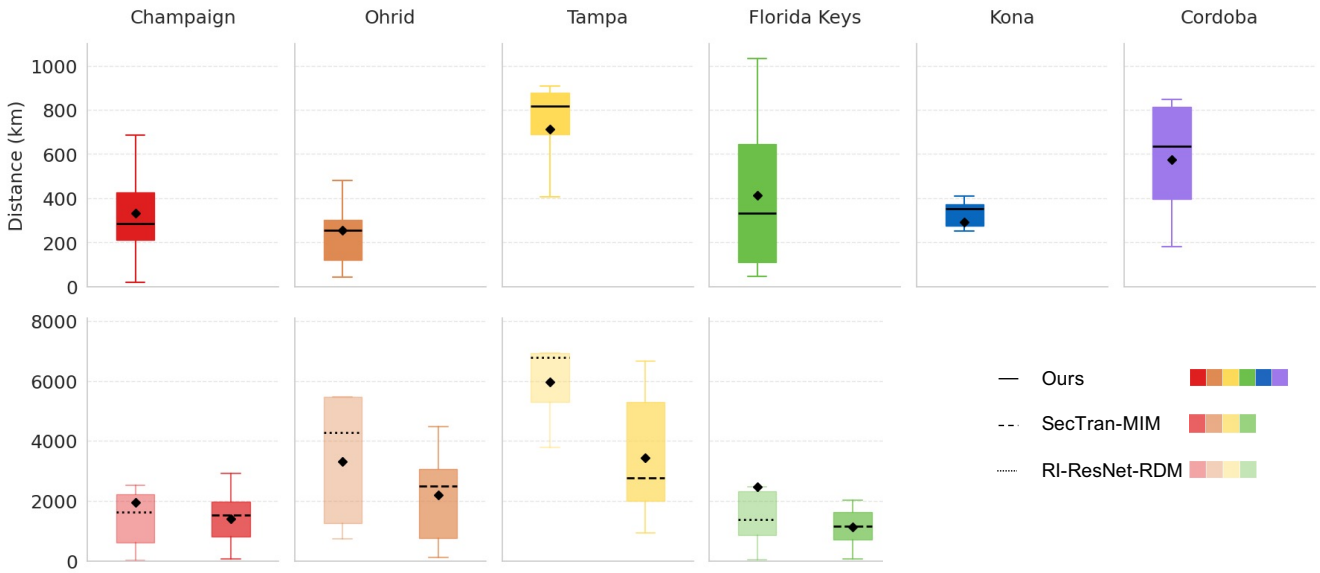

Figure S13. **North–south geolocation accuracy in the cross-site task.** Panels report the geodesic distance between estimated and true latitude for every test sequence; lower values are better. Our Polar Transformer (top row, solid lines) is compared with SecTran-MIM [4] (bottom row, dashed) and RI-ResNet-RDM [2] (bottom row, dotted). All maps share the same color code for the six test sites, but the baseline panels are shown at eight times the vertical scale to accommodate their larger errors. Across all sites, Polar Transformer yields the smallest latitude error; SecTran-MIM is intermediate, and RI-ResNet-RDM performs worst. Black diamond markers (⧫) indicate the site-average error for each method.

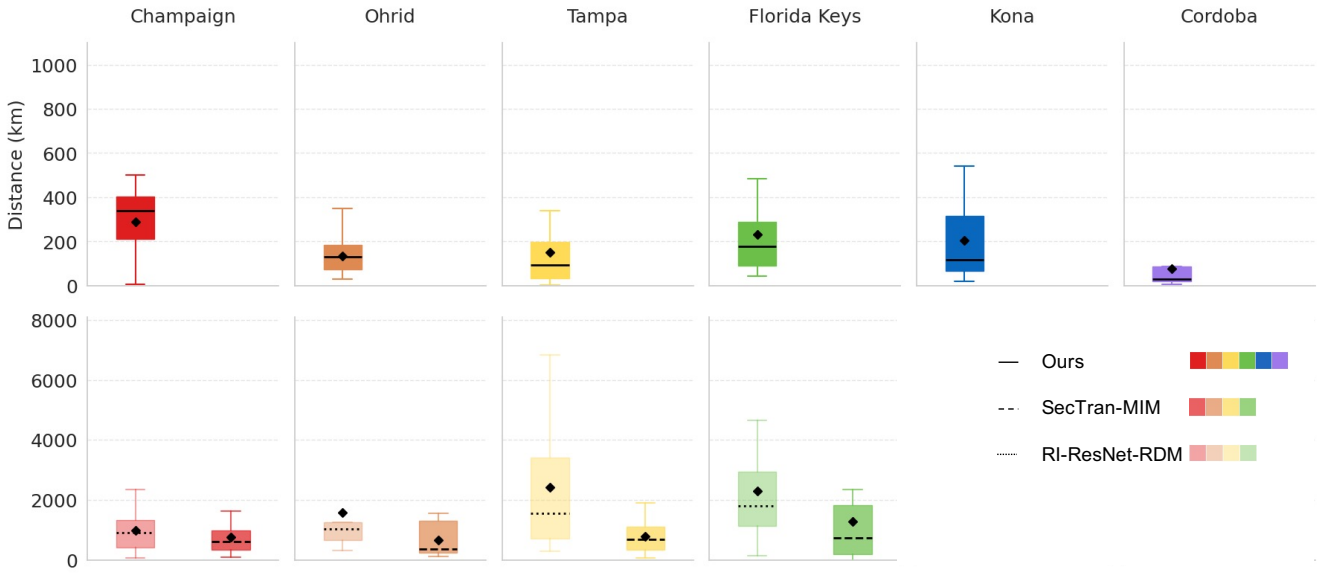

Figure S14. **East–west geolocation accuracy in the cross-site task.** Panels show the geodesic distance between estimated and true longitude for every test sequence; lower values are better. Polar Transformer results appear in the top row (solid lines), while SecTran-MIM [4] and RI-ResNet-RDM [2] occupy the bottom row with dashed and dotted lines, respectively. All six test sites share the same color code, but the baseline panels are plotted at eight times the vertical scale to accommodate their larger errors. Polar Transformer consistently delivers the smallest longitude error, followed by SecTran-MIM, with RI-ResNet-RDM performing worst. Black diamonds (⧫) mark the average error per site for each method.

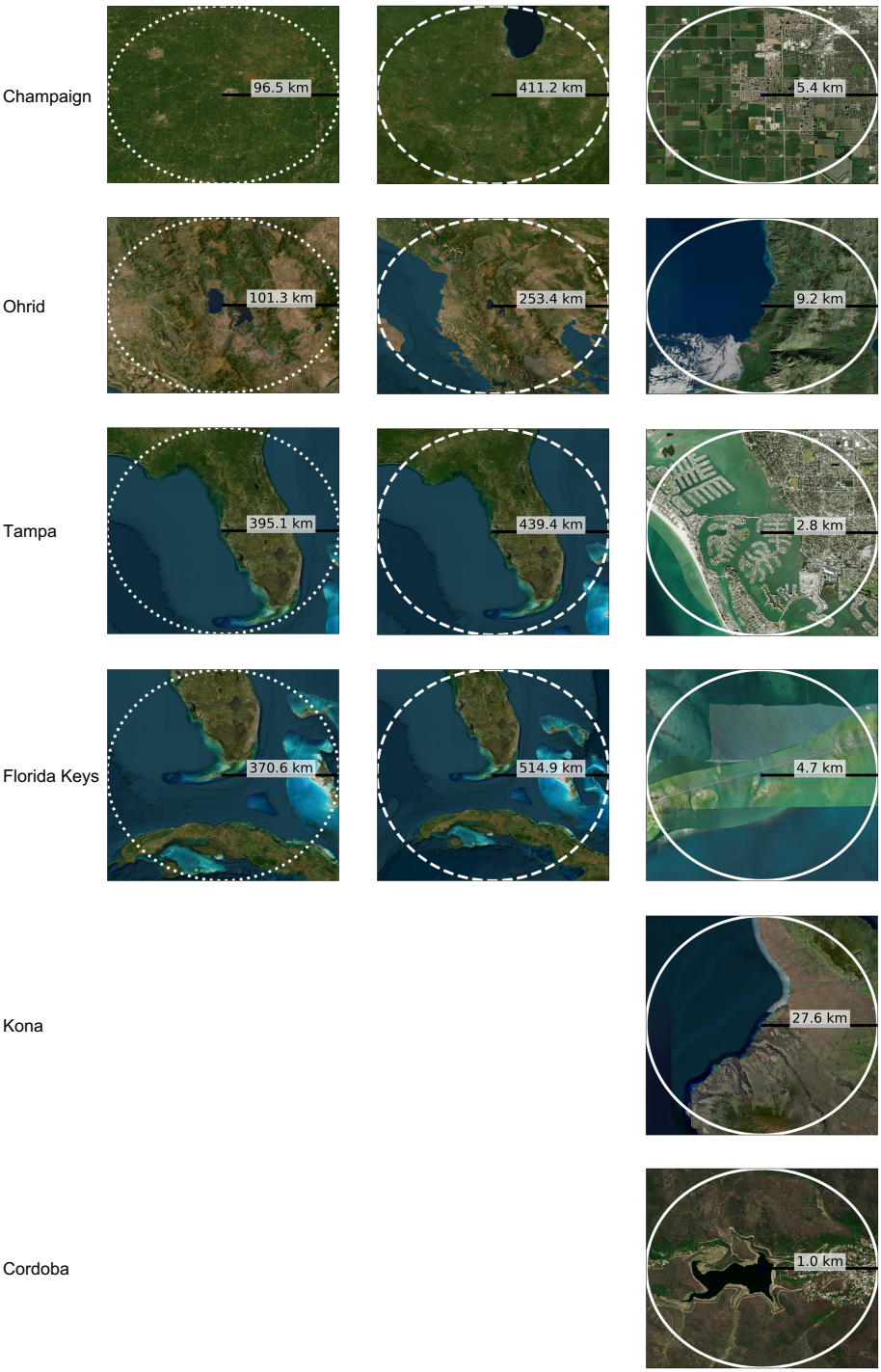

Figure S15. **Zoomed geodesic-error maps for the same-site evaluation.** For each test site, the satellite inset shows the median error circle obtained by RI-ResNet-RDM [2] (dotted), SecTran-MIM [4] (dashed), and Polar Transformer (solid). The circle is centered on the ground-truth coordinates; its radius equals the method's median great-circle distance error across all test sequences collected at that site. Compared with the deep baselines, Polar Transformer compresses the error from regional scale ($\geq$ 100 km) to neighborhood scale ($\leq$ 10 km) at Champaign, Lake Ohrid, Tampa, and the Florida Keys, and delivers 1–30 km accuracy at Río Ceballos and Kona, where baseline results are unavailable.

Supplementary Note-10

Solar angle accuracy with incomplete daylight arcs. Many real test sequences span only a fraction of the daylight period. A model trained solely on complete sunrise-to-sunset trajectories therefore struggles when it encounters shorter or differently timed sequences at inference. To quantify this issue, we first analyze the completeness of our real data. Only a handful of trajectories from the Florida Keys, Río Ceballos, and Champaign approximate full-day coverage; the vast majority capture just a segment of the solar cycle (Fig. S16, right).

Defining trajectory completeness. To turn "partial" coverage into a structured notion, we derive ten canonical categories from each full-day curve (Fig. S16, left). We vary (i) the start point—0%, 25%, 50%, or 75% of the daylight span—and (ii) the portion that follows that start point—25%, 50%, 75%, or 100% of the span. Category 4, for example, starts at sunrise (0%) and runs to sunset (100%), while Category 10 begins three-quarters of the way through the day and covers only the final 25%.

Classifying real sequences. For every real test sequence, we simulate the complete sunrise-to-sunset trajectory for the same location and day, generate the ten category variants, and assign the real sequence to the variant with the smallest squared error. The distribution of categories across sites is shown in Fig. S16 right.

Augmentation strategy. To make the network robust to incomplete input, we introduce a dropout-style temporal augmentation (Fig. S17a). Starting from the full, minute-resolved solar-elevation curve for a given day, we partition the trajectory into six equal segments and randomly discard one, two, or three of them, so that the retained portion spans 50–83 % of the day. Only after this pruning step do we perform our usual stratified sampling, selecting 64 samples evenly spaced within the surviving segments. The procedure exposes the model to a wide spectrum of partial-context inputs while preserving the physical consistency of the solar trajectory.

Effectiveness. Fig. S17b compares the model with and without augmentation on two Champaign test sequences (21 and 26 September, Categories 1 and 9). Without augmentation, the solar-elevation RMSE and geodesic error are high; with augmentation, RMSE falls by more than 60% in Category 1 and 20% in Category 9, and the location errors by over 40%, demonstrating that the model has learned to handle partial trajectories.

Extended data Fig. S18b shows that predicted trajectories become noticeably jagged if removing transformer encoder in solar elevation module.

Supplementary Note-11

Influence of simulation region size. For cross-site evaluation the simulator must draw training locations from a geographic region known *a priori*. We select points uniformly from the spherical triangle region with Arvo's sampling[1]. To keep computation tractable we fix the sample count at 64 locations per experiment—enough to cover a $20^\circ \times 20^\circ$ triangle ($\approx$ 2 M km^2) while keeping the full simulate–train–test cycle below 30 h on a single GPU.

Fig. S19 explores what happens as this prior expands. Starting from the baseline 20° triangle around Champaign, we enlarge the latitude and longitude bounds in 5° steps, up to a 60° span ($\approx$ 23 M km^2). Because the number of samples is held constant, wider triangles imply lower spatial density. The box plots show that performance degrades gracefully. Median solar-elevation RMSE climbs from 2.5° to 11°, and median position error from 450 km to 1,800 km, with most of the increase coming from the north–south component; east–west error remains roughly constant. Linear fits account for less than a quarter of the variance ($R^2 \leq 0.22$), indicating that error growth is sublinear in region size. These results suggest that extending accuracy to continental or global priors will require either denser sampling, additional real-world data at extreme elevations, or greater model capacity—directions we leave for future work.

Supplementary Note-miscellaneous

Quantitative experimental results in table format, corresponding to plots in the main paper Figs.1,6-7, are provided in Tab. 1.

Depth of camera ranges from 5 to 10 meters across the sites.

Using AoP instead of raw multi-channel intensity frames: AoP is physically meaningful and largely independent of the imaging sensor, while raw intensity frames depend on the camera. Using AoP means the model does not need to be limited to a particular camera.

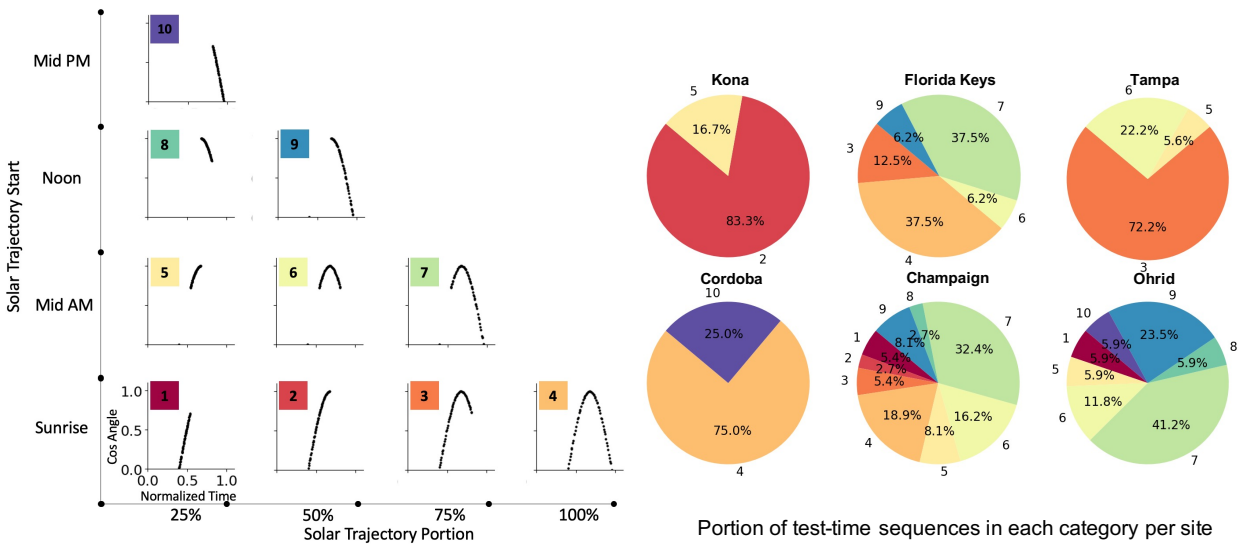

Figure S16. **Daylight coverage of test sequences. Left:** A full sunrise–to–sunset solar trajectory is truncated along two axes: start time (0%, 25%, 50%, or 75% of the daylight span) and retained portion (25%, 50%, 75%, or 100% after the start). The ten resulting categories are illustrated with example curves; Category 4 (0%, 100%) is a complete day, whereas Categories 1 and 10 capture only the early-morning or near-sunset segment. Real test sequences are assigned to a category by simulating the full-day curve for the same location and date, generating all ten variants, and choosing the variant with the smallest squared error. **Right:** Distribution of categories by site. Champaign and Lake Ohrid span nearly the entire set, while Río Ceballos and Kona contain mostly short segments.

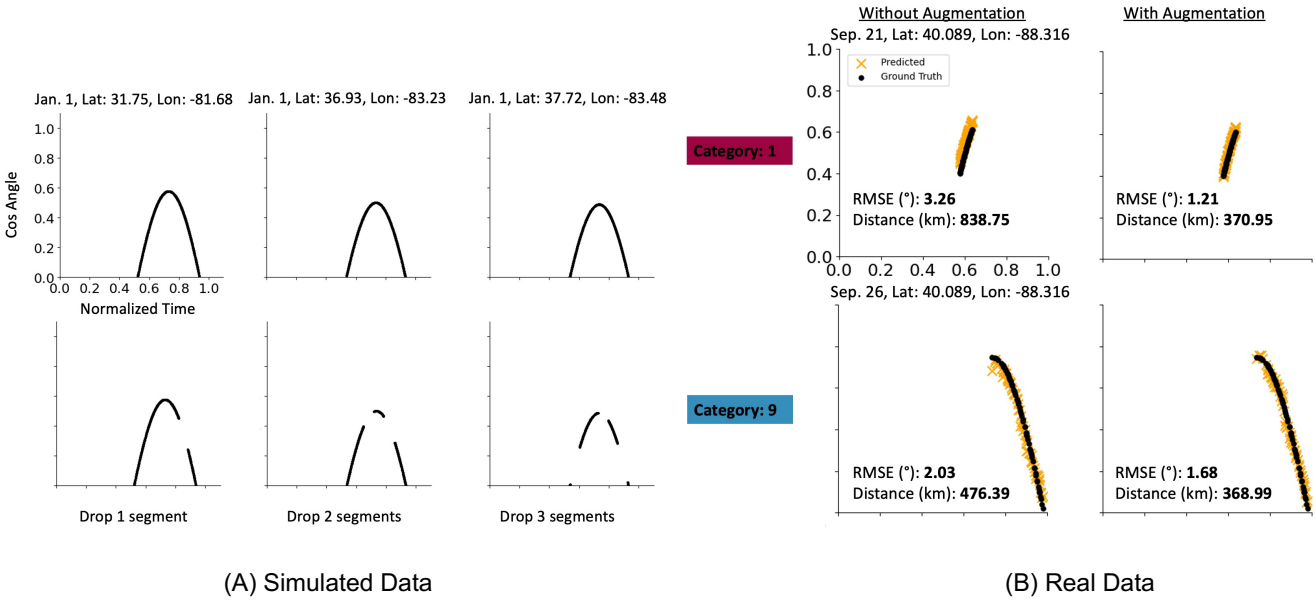

Figure S17. **Trajectory-dropout augmentation improves robustness. (A)** Schematic of the augmentation strategy. Each simulated sunrise-to-sunset trajectory is divided into six equal segments; during training we randomly remove one, two, or three contiguous segments, exposing the network to both complete and incomplete inputs. The top row shows three full trajectories generated for the same calendar day at different locations; the lower row shows their augmented counterparts. **(B)** Two real test sequences from the Champaign site (21 and 26 September), classified as incomplete Categories 1 and 9. Left column: predictions from a model trained without augmentation, with the resulting RMSE and geodesic errors. Right column: corresponding predictions from the augmented model, which reduce solar-angle RMSE and geolocation error for both examples.

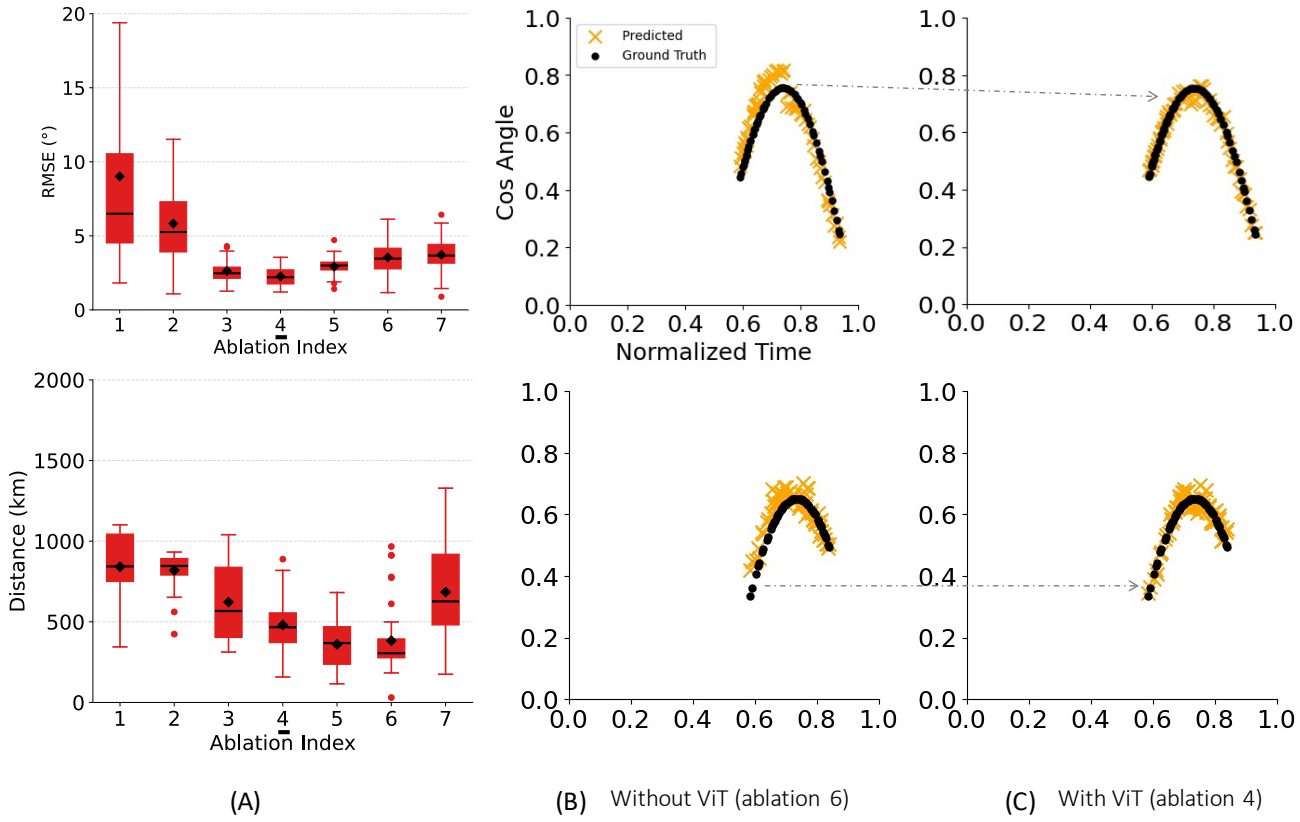

Figure S18. **Impact of individual components on cross-site generalization (Champaign test site).** **(A)** Box-and-whisker plots summarizing model performance with each configuration in the ablation study. **(B–C)** Two representative days (Sept. 24 and Oct. 16) show solar-elevation predictions with and without the ViT encoder; arrows mark deviations from the ground truth.

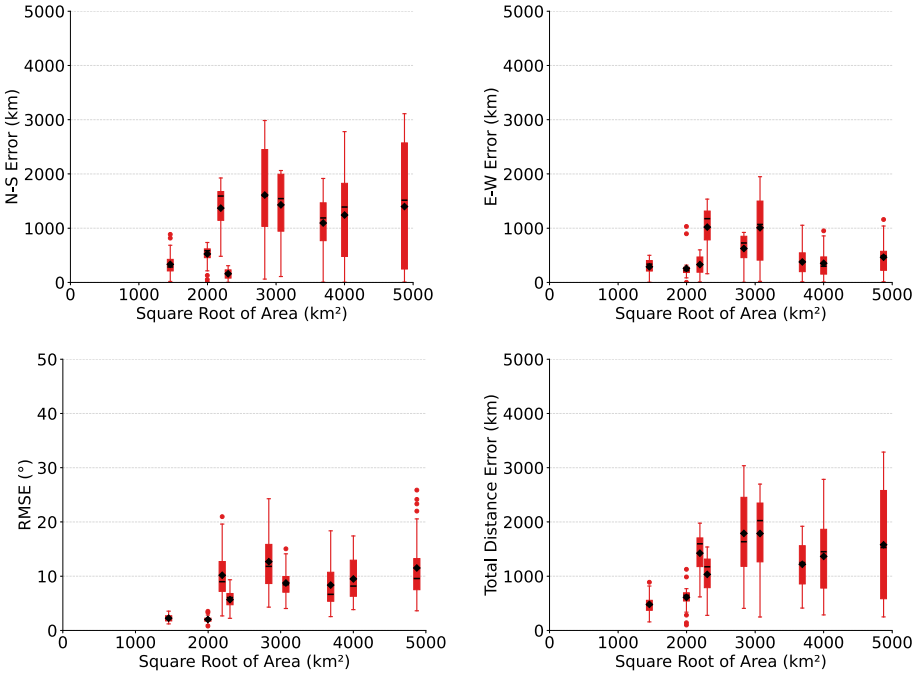

Figure S19. **Model performance on cross-site geolocation, evaluated over spherical-triangle regions of increasing size centered on Champaign.** Plots show RMSE for solar-trajectory prediction (top left) and geodesic-distance errors (north–south, east–west, total) for coordinate estimation. Each box plot represents a region, with the x-axis showing the square root of its area. To expand spatial context, we increase latitude bounds from 30–50° to 20–60° and 10–70°, and longitude from $-78°$ to $-98°$, to $-68°$ to $-108°$, and $-58°$ to $-118°$, increasing area from 2 to 23 M km^2. Errors generally increase with region size, but not linearly. Linear regressions show: solar RMSE (slope = 0.0023, $R^2 = 0.22$, $p < 0.001$) and geodesic error (slope = 0.2635, $R^2 = 0.14$, $p < 0.001$). Low R^2 values suggest error growth is not well explained by linear trends. Performance degrades primarily owing to shifts in solar-trajectory peaks over wider latitude ranges, increasing north–south errors. East–west errors remain stable, but total error grows because of north–south dominance. Each region includes 64 simulated locations. Improving performance in large regions may require greater sample density, model capacity, and diverse real-world data at extreme elevations—left for future work.

Metric	Test Site	Mean				Median		
		P	Se	RI	Ours	Se	RI	Ours
Cross-site								
D(↓)	Champaign	–	1674.62	2404.11	**480.41**	1842.40	2066.38	**466.32**
	Lake Ohrid	–	2370.05	3546.44	**298.32**	2501.41	4524.29	**274.89**
	Tampa Bay	–	3512.91	6155.83	**779.94**	2815.12	6814.49	**828.66**
	Florida Keys	–	2007.88	3566.91	**524.69**	1732.88	2786.05	**465.29**
	Kona	–	–	–	**389.27**	–	–	**358.36**
	Rio Ceballos	–	–	–	**605.99**	–	–	**637.49**
R(↓)	Champaign	–	9.11	11.63	**2.26**	8.79	11.29	**2.20**
	Lake Ohrid	–	12.01	14.72	**4.61**	12.25	16.38	**4.58**
	Tampa Bay	–	23.98	30.58	**8.94**	23.84	33.11	**8.60**
	Florida Keys	–	9.30	14.58	**5.29**	10.07	12.68	**4.86**
	Kona	–	–	–	**4.85**	–	–	**4.61**
	Rio Ceballos	–	–	–	**2.78**	–	–	**2.54**
Same-site								
D(↓)	Champaign	1554.30	391.25	114.05	**5.44**	411.24	96.46	**5.43**
	Lake Ohrid	1224.44	197.46	147.38	**9.18**	253.43	101.31	**9.18**
	Tampa Bay	5203.17	666.69	580.92	**2.62**	439.36	395.12	**2.82**
	Florida Keys	1756.65	864.79	864.45	**4.94**	514.92	370.58	**4.68**
	Kona	–	–	–	**27.65**	–	–	**27.65**
	Rio Ceballos	–	–	–	**0.96**	–	–	**0.96**
R(↓)	Champaign	14.58	3.48	1.60	**1.24**	3.45	1.47	**1.19**
	Lake Ohrid	11.11	2.62	**1.79**	3.62	2.48	**0.85**	3.62
	Tampa Bay	35.86	7.05	5.62	**3.86**	7.26	4.82	**3.49**
	Florida Keys	16.87	4.45	5.10	**1.45**	4.36	3.09	**1.23**
	Kona	–	–	–	**2.86**	–	–	**2.86**
	Rio Ceballos	–	–	–	**3.63**	–	–	**3.63**

Table 1. Performance comparisons. D: Distance (km); R: RMSE (°). P: Parametric [5]; Se: SecTran-MIM [4]; RI: RI-ResNet-RDM [2].

References

[1] James Arvo. Stratified sampling of spherical triangles. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, page 437–438, New York, NY, USA, 1995. Association for Computing Machinery. 19

[2] Xiaoyang Bai, Zuodong Liang, Zhongmin Zhu, Alexander Schwing, David Forsyth, and Viktor Gruev. Polarization-based underwater geolocalization with deep learning. *eLight*, 3(15), 2023. 16, 17, 18, 23

[3] Xiaoyang Bai, Zhongmin Zhu, Alexander Schwing, David Forsyth, and Viktor Gruev. Angle of polarization calibration for omnidirectional polarization cameras. *Optics Express*, 31 (4):6759–6769, 2023. 1

[4] Xiaoyang Bai, Zhongmin Zhu, Alexander Schwing, David Forsyth, and Viktor Gruev. Learning a global underwater geolocalization model with sectoral transformer. *Optics Express*, 32(12):20706–20718, 2024. 16, 17, 18, 23

[5] Samuel B Powell, Roman Garnett, Justin Marshall, Charbel Rizk, and Viktor Gruev. Bioinspired polarization vision enables underwater geolocalization. *Science Advances*, 4(4): eaao6841, 2018. 23