Supporting Information for "Systematic Error in Flood Hazard Aggregation"

Seth Bryant[1,2], Heidi Kreibich[1], Bruno Merz[1,2]

[1]GFZ German Research Centre for Geosciences, Section 4.4. Hydrology, Potsdam, Germany

[2]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

S1. Introduction

This supplement provides an analytical treatment of errors introduced by flood hazard data aggregation described in the main text. This employs a novel *resample case* framework to investigate two typical aggregation routines. Aggregating or upscaling involves transferring data grids from fine ($s1$) to coarse ($s2$) scales where the support of the two domains can be expressed as:

$$s1 < s2$$

This is closely related to the linear dimension or resolution ($\lambda_1 < \lambda_2$) of the corresponding grid cells, often expressed in meters. From this, and the fact that both grids have the same extents, we can say:

$$\frac{\lambda_2^2}{\lambda_1^2} = \frac{N_1}{N_2} \tag{S1}$$

where N is the total cell count of the corresponding grid. See Section 2 of the main text for further context and equations.

S1.1. Aggregation Routines

The two aggregation routines considered here are summarized in the main text Section 3.1. These routines can generally be formulated as:

$$DEM_{s2}, WSH_{s2}, WSE_{s2} = f[DEM_{s1}, WSH_{s1}, WSE_{s1}, s2] \tag{S2}$$

where f is some aggregation routine. All aggregation routines first act on local groups of $s1$ cells, who we index with i, to obtain a new $s2$ value with index j. In this way, each i cell maps to a j index with a many:1 relation. The following sections elaborate on the two routines.

S1.1.1. First Routine: *WSH* Averaging

In the "WSH Averaging" routine, local $s1$ groups of DEM and WSH grids are simply averaged to yield new $s2$ cells, using Equations 4 and 5. Applying these to the full $s1$ domain yields aggregated DEM_{s2} and WSH_{s2} grids. Equation 1 is then used to compute WSE_{s2}:

$$\begin{aligned} DEM_{s2,j} &= \overline{DEM_{s1,i}} \\ WSH_{s2,j} &= \overline{WSH_{s1,i}} \\ WSE_{s2,j} &= DEM_{s2,j} + WSH_{s2,j} \end{aligned} \tag{S3}$$

S1.1.2. Second Routine: *WSE* Averaging

To satisfy Equation 3, "WSE Averaging" is more complicated, requiring a two step process: first, a wet averaging via Equation 6, then the cells violating Equation 3 are masked before computing WSH_{s2} via a modified Equation 1:

$$
\begin{aligned}
DEM_{s2,j} &= \overline{DEM_{s1,j}} \\
WSE_{s2,j} &= \begin{cases} null & \text{if } \overline{WSE_{s1,j}} \leq \overline{DEM_{s1,j}} \\ \overline{WSE_{s1,j}} & \text{else} \end{cases} \\
WSH_{s2,j} &= \begin{cases} 0 & \text{if } WSE_{s2,j} = null \\ WSE_{s2,j} - DEM_{s2,j} & \text{else} \end{cases}
\end{aligned}
\tag{S4}
$$

Both routines are summarized in Figure 1.

S2. Method

Using the *resample case* framework defined in Section 3.2 and summarized in Figure 2, we investigate systematic errors introduced by the two aggregation routines presented above on four metrics of importance to flood models: two primary metrics, water depth (WSH) and water surface elevation (WSE), and two derivative metrics, inundation area (A), and volume (V).

S2.0.1. Global Bias

To attribute errors to some aggregation routine, we define "global" bias as the difference between some metric computed with the aggregated vs. the raw grid. This can be formulated for some metric M, which is a reducing function of grid G (e.g., $M[G] = mean[G]$) and the aggregation routine f as:

$$
Bias_{global}[M, f, s2] = M[f[G_{s1}, ...]] - M[G_{s1}] \tag{S5}
$$

S2.0.2. Local Bias

For the primary grid metrics (WSH and WSE), Equation S5 can alternatively be computed locally, by first calculating the difference of each i cell, before applying the reducing function M to obtain the grid bias:

$$Bias_{local}[M, f, s2] = M[f[G_{s1}, ...]]_i - G_{s1,i}] \quad \text{where} \quad i \neq null \tag{S6}$$

For metrics computed from the WSE grid, this local bias can of course only be computed in regions inundated by both $s1$ and $s2$ grids (see Equation 2). For consistency, we apply this same constraint to WSH metrics (i.e., cells where $WSH = 0$ are excluded). While this masks the performance of a routine in dry regions, it provides a consistent way to separate the reporting of bias in local variables from bias in inundation area (which is reported as a secondary metric). Extending the "resample case" framework to these two definitions of bias, and assuming that M is linear in the wet domain, it follows that:

$$Bias_{local} = \begin{cases} n/a & \text{if } DD \\ \neq Bias_{global} & \text{if } DP \\ \neq Bias_{global} & \text{if } WP \\ = Bias_{global} & \text{if } WW \end{cases} \tag{S7}$$

In other words, when computed on specific "resample case" regions, $Bias_{local}$ may only differ from $Bias_{global}$ in partial regions (DP and WP) and is undefined in DD regions.

S3. Bias in Flood Depths (WSH)

For computing bias in flood depths (WSH), we focus on the grid average at support $s = s1$ or $s = s2$:

$$M[s] = \frac{1}{N_s} \sum_{i=1}^{N_s} WSH_{s,i} \tag{S8}$$

where N_s is the count of cells i within the grid at support s, which is the global version of Equation 5. Expanding the global bias in Equation S5 with this yields:

$$Bias_{global}[f, s2] = \frac{1}{N_2} \sum_{j=1}^{N_2} WSH_{s2,j} - \frac{1}{N_1} \sum_{i=1}^{N_1} WSH_{s1,i} \tag{S9}$$

A similar expansion for the local bias in Equation S6 yields:

$$Bias_{local}[f, s2] = \frac{1}{N_2} \sum_{j=1}^{N_2} (WSH_{s1,i,j} - WSH_{s2,i,j}) \tag{S10}$$

S3.1. First Routine: *WSH* Averaging

Comparing Equation S9 with our definition of the "WSH Averaging" routine (Equation S3) shows this routine has no systematic bias in global WSH.

S3.1.1. Local Bias

To explore the local bias of this routine, we first examine regions classified by the WW resample case defined by Equation 7, where $min(WSE_{s1,i}) > max(DEM_{s1,i})$. This can be re-written using Equation 1 in terms of WSH for convenience as:

$$WW \equiv min(WSH_{s1,i}) > 0 \tag{S11}$$

In other words, all i cells within j are wet. Expanding Equation S9 for a j group of i cells, then substituting with Equation S3 yields:

$$\begin{aligned} Bias[f, s2] &= \frac{1}{N_2} \sum_{j=1}^{N_2} (\frac{1}{N_{12}} \sum_{i=1}^{N_{12}} WSH_{s1,j,i} - WSH_{s1,j,i}) \\ &= \frac{1}{N_2} \sum_{j=1}^{N_2} (\frac{N_{12}}{N_{12}} (WSH_{s1,j,1} + WSH_{s1,j,2} + ...) - (WSH_{s1,j,1} + WSH_{s1,j,2} + ...)) \\ &= 0 \end{aligned} \tag{S12}$$

The terms in line two cancel from because in the WW region (Equation S11) i blocks and j blocks are equivalent. This is intuitive if we consider the absence of *null* values in these WW regions.

Taking a similar approach to evaluate the WP region, where $min(WSE_{s1,i}) \geq max(DEM_{s1,i})$ and $\overline{DEM_{s1,i}} < \overline{WSE_{s1,i}}$, re-written again in terms of WSH:

$$WP \equiv min(WSH_{s1,i}) = 0 \quad \text{and} \quad \overline{DEM_{s1,i}} < \overline{WSE_{s1,i}} \tag{S13}$$

in other words, some i cells now are dry, but the group average is still higher than the DEM average. Following our definition of local bias (Equation S6), the domain of computation for this metric is further constrained to cells which are non-*null* in both the $s1$ and $s2$ grids:

$$\text{for } i \text{ where } WSH_{s1,i} > 0 \text{ and } WSH_{s2,i} > 0 \tag{S14}$$

Starting from Equation S12 for the combined domain of Equation S13 and S14 and adopting $i = 2$ as an illustrative dry cell yields:

$$\begin{aligned} Bias_{local}[f, s2] &= \frac{1}{N_2}\sum_{j=1}^{N_2}\left(\frac{N_{wet,j}}{N_{12}}(WSH_{s1,j,1} + \cancelto{0}{WSH_{s1,j,2}} + ...) - (WSH_{s1,j,1} + ...)\right) \\ &= \frac{1}{N_2}\sum_{j=1}^{N_2}\left(\frac{N_{wet,j}}{N_{12}} - 1\right) \\ &< 0 \end{aligned}$$

because $N_{wet} < N_{12}$ by definition. In other words, if the calculation is limited to the wet domain (per our definition of local bias in Equation S6), the $s2$ values are systematically low, as the aggregation does *not* have the same limitation, and the $s2$ value is therefore pulled down by the dry neighbours. The same result holds for DP regions, with the bias likely being more severe assuming that $N_{wet,DP} < N_{wet,WP}$.

S3.2. Second Routine: $\boldsymbol{WSE}$ Averaging

S3.2.1. Global Bias

Looking now at the global bias introduced by the WSE preserving routine described in Equation S4, we substitute this into Equation S9:

$$Bias_{global}[f, s2] = -\frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} + \begin{cases} 0 & \text{if } WSE_{s2,j} = null \\ \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i}} - \overline{DEM_{s1,i}}) & \text{else} \end{cases} \tag{S15}$$

For the DD case ($WSH = 0$ and $WSE = null$), all terms reduce to zero. For the DP case, we can re-write the domain condition from Equation 7 by substituting in WSH using Equation 1:

$$DP \equiv max(WSH_{s1,i}) > 0 \text{ and } \overline{DEM_{s1,j}} \geq \overline{WSE_{s1,j}} \tag{S16}$$

In other words, there are some wet i cells, but their (wet) average is less than the (wet+dry) average of the terrain. Equation S4 states that $WSE_{s2,j} = null$ for this condition, reducing Equation S15 to:

$$\begin{aligned} Bias_{global}[f, s2] &= -\frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} + 0 \\ &< 0 \end{aligned}$$

because $max(WSH_{s1,i}) > 0$ implies $\sum_{i=1}^{N_1} WSH_{s1,i} > 0$ (and from Equation 2 we know $WSH \geq 0$). In other words, because this routine always yields a dry $WSH = 0$ value in DP cells, the bias is always negative in this region.

For the WP case, the domain condition is stated above in Equation S13. This is the most interesting case as $WSE_{s2,j}$ is non-null and the second part of Equation S15 therefore reduces to the non-zero term:

$$Bias_{global}[f, s2] = \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i}} - \overline{DEM_{s1,i}}) - \frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} \tag{S17}$$

To evaluate this, we separate DEM averaging into *wet* and *dry* regions for later comparison knowing $\overline{DEM_{all}} = \overline{DEM_{wet}} + \overline{DEM_{dry}}$:

$$
\begin{aligned}
Bias_g[f, s2] = {} & \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i,wet}} + \cancelto{0}{\overline{WSE_{s1,i,dry}}} - \overline{DEM_{s1,i,wet}} - \overline{DEM_{s1,i,dry}}) \\
& - \frac{1}{N_{wet}}\sum_{i=1}^{N_{wet}} WSH_{s1,i} - \cancelto{0}{\frac{1}{N_{dry}}\sum_{i=1}^{N_{dry}} WSH_{s1,i}} \\
& = \cancelto{0}{\frac{1}{N_2}\sum_{j=1}^{N_2}((\overline{WSE_{s1,i,wet}} - \overline{DEM_{s1,i,wet}}) - \frac{1}{N_{wet}}\sum_{i=1}^{N_{wet}} WSH_{s1,i})} \\
& - \frac{1}{N_2}\sum_{j=1}^{N_2}\overline{DEM_{s1,i,dry}} \\
& = -\frac{1}{N_2}\sum_{j=1}^{N_2}\overline{DEM_{s1,i,dry}} \\
& < 0
\end{aligned}
\qquad \text{(S18)}
$$

in other words, in WW regions this routine introduces a negative bias equivalent to the average value of the dry DEM cells.

For the WW case and the domain condition ($min(WSH_{s1,i}) > 0$), Equation S17 reduces to zero.

S3.2.2. Local Bias

Examining the local bias (Equation S10) of WSH produced by the "WSE Averaging" routine, Equation S7 states WW will also have no local bias. For DP and DD regions, recall that the routine (Equation S3) returns *dry* values for j cells, therefore these are excluded per our definition of local bias (Equation S6). For the remaining WP regions, Equation S18 still holds; however, the summation domain will differ and therefore so will the magnitude.

S4. Bias in Water Surface Elevation ($\boldsymbol{WSE}$)

Like flood depths (WSH), WSE is a primary variable and we therefore focus on the grid average at support $s = s1$ or $s = s2$:

$$M[s] = \frac{1}{N_{s,wet}} \sum_{i=1}^{N_{s,wet}} WSE_{s,i}$$

Like Equation 6, dry values are ignored. Expanding the global bias with this as in Equation S9 yields:

$$Bias_{global}[f, s2] = \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} WSE_{s2,j} - \frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} WSE_{s1,i} \tag{S19}$$

And for the local bias:

$$Bias_{local}[f, s2] = \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (WSE_{s1,i,j} - WSE_{s2,i,j}) \tag{S20}$$

S4.1. First Routine: $\boldsymbol{WSH}$ Averaging

S4.1.1. Global Bias

Substituting our definition of the "WSH Averaging" routine from Equation S3 into Equation S19 yields:

$$\begin{aligned} Bias_{global}[f, s2] = {} & \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\overline{DEM_{s1,i}} + \overline{WSH_{s1,i}}) \\ & - \frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} (DEM_{s1,i} + WSH_{s1,i}) \end{aligned} \tag{S21}$$

For the DD case, WSE is not defined; while for the WW all terms cancel to zero. For the DP case, the domain is provided in Equation S16. Expanding Equation S21 with this and separating into wet and dry regions again yields:

$$
\begin{aligned}
Bias_{global}[f, s2] = {} & \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\overline{DEM_{s1,i,wet}} + \overline{DEM_{s1,i,dry}} + \overline{WSH_{s1,i,wet}} + \cancelto{0}{\overline{WSH_{s1,i,dry}}}) \\
& - \frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} (DEM_{s1,i,wet} + WSH_{s1,i,wet}) \\
= {} & \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\cancelto{0}{\overline{DEM_{s1,i,wet}} - \frac{1}{N_{1,wet}} DEM_{s1,i,wet}} \\
& + \cancelto{0}{\overline{WSH_{s1,i,wet}} - \frac{1}{N_{1,wet}} WSH_{s1,i,wet}} + \overline{DEM_{s1,i,dry}}) \\
= {} & \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} \overline{DEM_{s1,i,dry}} \\
> {} & 0
\end{aligned} \tag{S22}
$$

Similar to the $Bias_{global}[WSH]$ of the "WSE averaging" routine derived above (Equation S18), the magnitude of the bias is related to $\overline{DEM_{s1,dry}}$, but with opposite directions.

A similar result holds for the WP case; however, at a lesser magnitude assuming $\overline{DEM_{s1,dry,WP}} < \overline{DEM_{s1,dry,DP}}$.

S4.1.2. Local Bias

Examining the local bias of WSE (Equation S20) produced by the "WSH Averaging" routine, again Equation S7 shows WW will also have no local bias (and DD cells are *null*). Similarly, Equation S22 holds for the partial regions (DP and WP).

S4.2. Second Routine: *WSE* Averaging

S4.2.1. Global Bias

Looking now at the global bias introduced by the WSE preserving routine described in Equation S4, we substitute this into Equation S19:

$$Bias_{global}[f, s2] = -\frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} WSE_{s1,i} + \begin{cases} null & \text{if } \overline{WSE_{s1,j}} = null \\ \overline{WSE_{s1,j}} & \text{else} \end{cases} \tag{S23}$$

For DD and DP regions, all terms are $null$. For WW regions, all terms reduce to zero. For WP regions, all terms also reduce to zero per Equation 6.

S4.2.2. Local Bias

Given that both local bias and WSE are only defined in wet regions, WSE global bias is equivalent to local bias for the "WSE Averaging" routine.

S5. Bias in Inundation Area (A)

Inundation area (A) is an important secondary metric for flood models and can be simply computed with a binary transformation from either the WSE or the WSH grid using Equation 2:

$$A_i = \begin{cases} 0 & \text{if } WSH_i = 0 \text{ or } WSE_i = null \\ 1 & \text{else} \end{cases} \tag{S24}$$

A_i can further be multiplied by λ^2 to obtain a geospatial inundation area (e.g., in square meters). For computing bias from aggregation routines, we focus on the total grid inundation area:

$$M[s] = \sum_{i=1}^{N_s} A_{s,i} \tag{S25}$$

We select this metric, rather than average area, to align with standard metrics in flood literature. However, because the grid sizes do not change, total and average area only differ by a scalar ($\frac{1}{N_s}$). By combining Equation S24 with the "resample case" framework (Equation 7), the total inundation area of some group j, computed directly on the fine

($s1$) grid, can be written as:

$$A_{s1,j} = \begin{cases} 0 & \text{if } DD \\ 0 < x < s1 & \text{if } DP \\ 0 < x < s1 & \text{if } WP \\ s1 & \text{if } WW \end{cases} \quad \text{(S26)}$$

In other words, partial regions have some dry cells, DD regions have all dry cells, and WW regions have no dry cells. With this, we can compare against the area $A_{s2,j}$ which is computed on the aggregated grids to calculate the bias of this metric. For this metric, local bias and global bias are equivalent by definition.

S5.1. First Routine: *WSH* Averaging

Combining Equation S3 and Equation 7, inundation area for this routine can be written for each j in terms of "resample case" as:

$$A_{s2,j} = \begin{cases} 0 & \text{if } DD \\ s1 & \text{if } DP \\ s1 & \text{if } WP \\ s1 & \text{if } WW \end{cases}$$

Comparing this to Equation S26 shows that the partial zones have a positive bias and WW and DD have no bias.

S5.2. Second Routine: *WSE* Averaging

Combining Equation S4 and Equation 7, inundation area can be written as:

$$A_{s2,j} = \begin{cases} 0 & \text{if } DD \\ 0 & \text{if } DP \\ s1 & \text{if } WP \\ s1 & \text{if } WW \end{cases}$$

The only difference with the previous routine being the DP region. From this, it follows that DP has a negative bias and WP has a positive bias, while the remaining have no bias.

S6. Bias in Flood Volume (V)

Flood volume (V) is a metric of interest to hydrodynamic models which assume volume conservation. For our evaluation, flood volume is computed from the depth grid (WSH) and the geospatial area (A):

$$\begin{aligned} V_i &= WSH_i * \cancelto{1 \text{ for wet}}{A_i} * \lambda^2 \\ &= WSH_i * \lambda^2 \end{aligned}$$

Like inundation area, here we focus on total grid volume:

$$\begin{aligned} M[s] &= \sum_{i=1}^{N_s} V_{s,i} \\ &= \lambda_s^2 \sum_{i=1}^{N_s} WSH_i \end{aligned}$$

Expanding Equation S5 with this yields:

$$Bias[f, s2] = \lambda_2^2 \sum_{j=1}^{N_2} WSH_{s2,j} - \lambda_1^2 \sum_{i=1}^{N_1} WSH_{s1,i} \tag{S27}$$

This is equivalent to the WSH bias multiplied by a constant.

S6.1. First Routine: $\boldsymbol{WSH}$ Averaging

To evaluate the V bias for the "WSH Averaging" routine, we substitute Equation S3 into Equation S27 which yields:

$$\begin{aligned} Bias[f, s2] &= \lambda_2^2 \sum_{j=1}^{N_2} \overline{WSH_i} - \lambda_1^2 \sum_{i=1}^{N_1} WSH_i \\ &= \lambda_2^2 (N_2 \overline{WSH_i}) - \lambda_1^2 (N_1 \overline{WSH_i}) \\ &= \overline{WSH_i} \cancelto{0}{(N_2 \lambda_2^2 - N_1 \lambda_1^2)} \end{aligned}$$

which cancels to zero following Equation S1.

S6.2. Second Routine: $\boldsymbol{WSE}$ Averaging

As discussed above, the "WSE Averaging" routine has no WSH bias in the WW and DD domains, so it follows V bias is similarly absent. For the DP and WP case, WSH bias is negative, so it follows V bias will also be negative.

S7. Summary

Here we have presented the novel "resample case" framework with which we could evaluate the direction of bias on four metrics under two aggregation routines. The resulting biases are summarized in Table 2.

S8. Computational Results: Additional Figures

Additional figures for the computational analysis are provided below. See the main text for details.

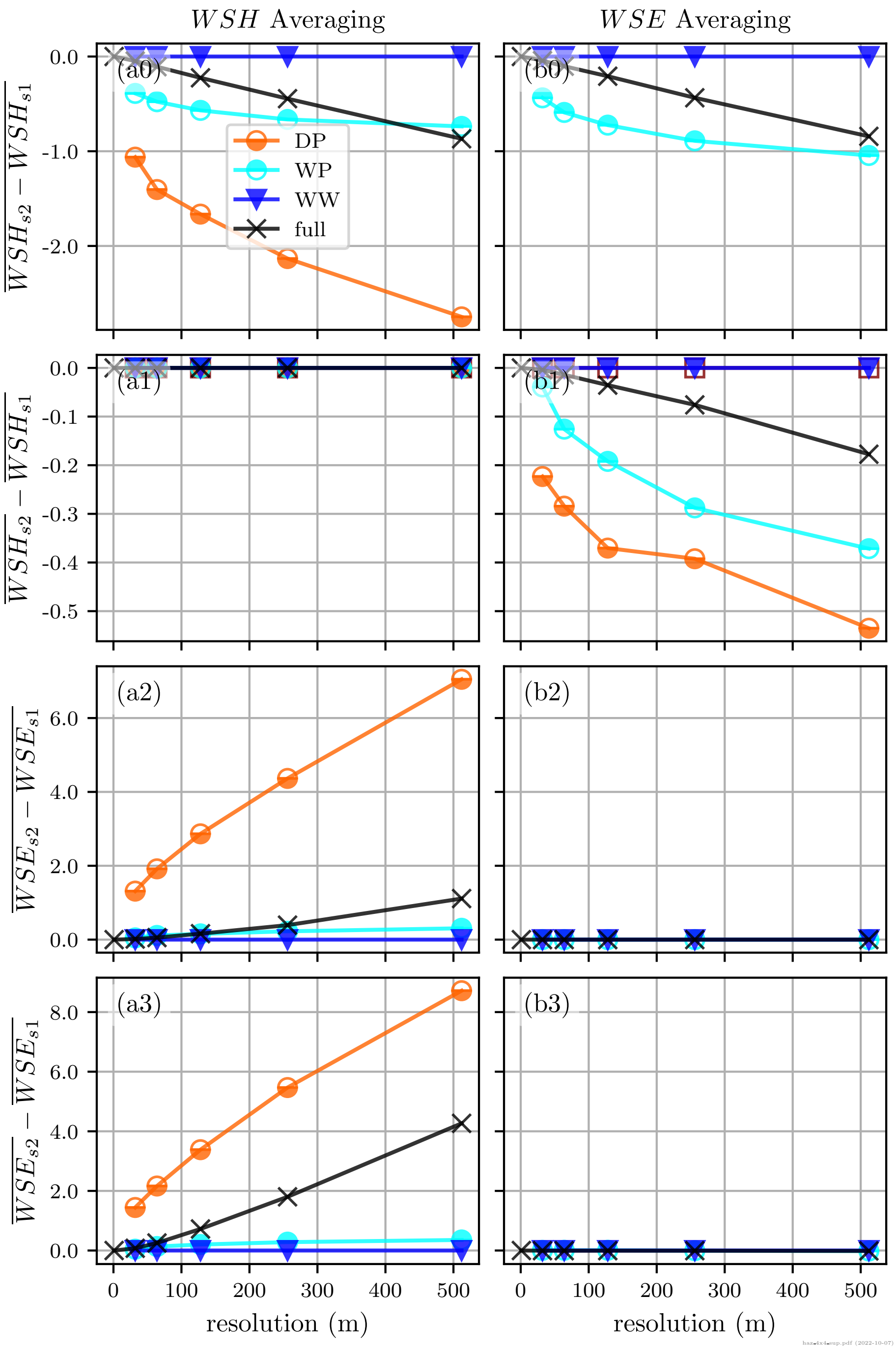

Figure S1. Full domain computation results. See main text for details

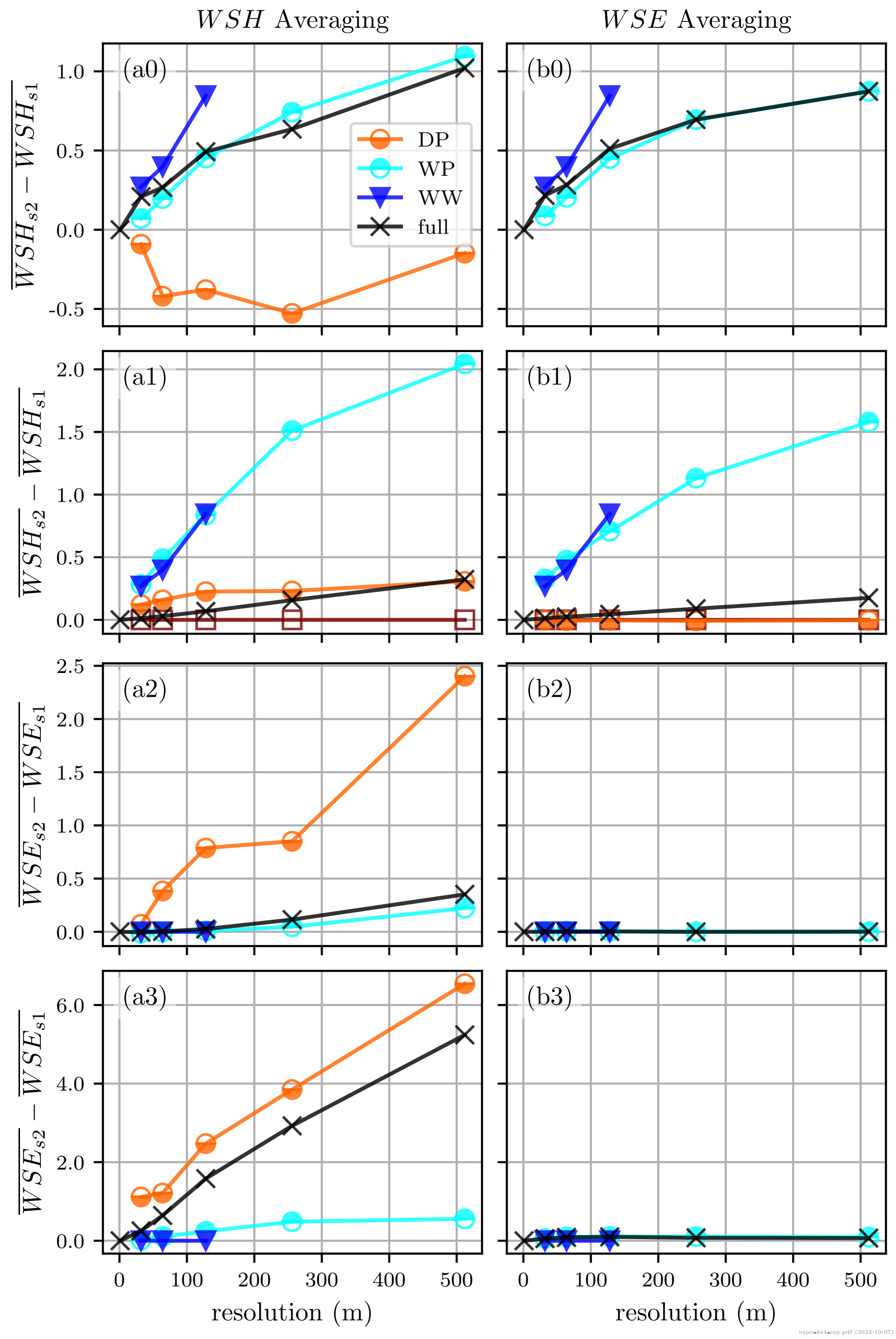

Figure S2. Exposed domain computation results. See main text for details