Response to Referee 1

Albertus J. Smit[1,2] and Neville Sweijd[3]

[1]Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, Cape Town 7535, South Africa

[2]South African Environmental Observation Network, Elwandle Coastal Node, Gqeberha 6031, South Africa

[3]School for Climate Studies, Stellenbosch University, Stellenbosch 7599, South Africa

28th July 2026

About this document

This document accompanies our author comment on the manuscript *A segmented-breakpoint sea-surface-temperature upwelling index for the Benguela Upwelling System*, in open discussion at *Ocean Science*. We respond to the general assessment (Section 1), the Methods passages rewritten as they would appear in the revised manuscript (Section 2), the walk-through figure the referee asked for (Section 3), and our replies to each of the specific comments (Section 4). Line numbers refer to the review PDF.

The replacement text you will see below gives the clarity the referee said was missing (yes, it was very dense!). We hope that each statistical step can now be followed on the page without recourse to the cited literature. The replacement text also demonstrates the new tone we propose for the revision as a whole, since the criticism was also about density and tone.

Three kinds of text appear below, and each item is set in the style it describes:

- Explanatory text, addressed to the referee, is in the normal body font, as in this sentence.
- Text proposed for the revised manuscript is in the same font, set in the red of the document style. Everything marked this way is meant to be read as manuscript replacement text rather than as a reply, and we will place it into the manuscript unchanged.
- AS SUBMITTED. Text quoted from the manuscript as submitted is set small and grey, and carries this label.

A caveat on length. Explaining a procedure on the page costs words, and the replacement passages are longer than the text they replace. We could not meet the referee's request while holding length constant, so we have written for clarity first and treated length as a separate problem. The passages are therefore *proposals* for content/tone. When the revision is assembled we will bring the manuscript within the length that *Ocean Science* permits, and we would welcome the editor's guidance on where the added explanation is best placed, whether in the main text, in the Supplement, or divided between the two.

1 Response to the general assessment

We thank the referee for a careful and helpful review, and for persisting through a manuscript that was harder work than it should have been. The central criticism is fair and we accept it. A methods paper whose method cannot be followed without consulting referenced papers has not done its job, whatever the merits of the method itself.

Two things... The first is that the statistical passages need to explain themselves on the page. We have rewritten them, and Section 2 gives the replacement text in full rather than describing our intention to produce it. When we reread the submitted version with the referee's comment in hand, we found the tone to be terse to the point of being unwelcoming, and defensive in places where the better thing is simply to say what we chose and why. We have tried to fix both.

The second is the walk-through figure. We agree, and we agree that it will do more for comprehension than any amount of added prose. Section 3 gives the proposed figure and its caption. This was a fun exercise and we think the result is worthwhile.

We are grateful for the comment on Figure 4 and for the assessment of the method. Every specific comment is addressed in Section 4; we accept almost all of them, and where we propose something different we say so and give our reasoning.

One clarification we should have made in the submitted version

We suspect that much of the difficulty came from a distinction we never stated. Apart from the ΔAIC guideline, the acceptance thresholds are our choices for our workflow rather than quantities derived from theory. We chose them to suit the data. The defence is not that they are the correct values but that the conclusions remain after considering reasonable alternatives, which is what the sensitivity assay in Table S2 was built to show. Leaving a reader to infer that from a table reference was a mistake, and the revised Methods says it now upfront (Section 2.1).

2 The Methods, rewritten

The passages below are *proposed* drop-in replacements for the corresponding paragraphs of the submitted manuscript, and are set in red as per the convention above. The text they replace is quoted in grey beneath each, so that the two can be compared.

2.1 Section 2.2, what counts as an observation

L. 85–91

Adjacent mesh transects are spaced at ca. 4 km, which is finer than the native pixel spacing of the L4 analyses, so neighbouring transects can sample the same pixel. Two rules therefore govern what counts as an observation.

Within a transect we sample one SST value per native pixel, and the model comparison uses that count, namely a median of ca. 91 to 101 values on OSTIA and OSPO and substantially fewer on AVHRR_OI and MW_OI (Table S1). Across transects, each L4 pixel is assigned to the first transect that covers it, and a transect contributes to the per-(UC, product, day) population only if at least one of its pixels remain after that deduplication.

The rules are independent, so the first fixes the sample size entering the model comparison at one transect, and the second fixes which transects enter the cell-scale aggregate, and leaves every individual fit unchanged.

> AS SUBMITTED. "We sample one SST observation per native pixel along the transect, so the Akaike information criterion (AIC) evidence uses the product's independent along-profile native-pixel count [...] Across-transect pixel uniqueness is then enforced at aggregation, so for each (UC, product, day), each L4 pixel is claimed by the first transect that covers it (first-occurrence-wins on col_id), and a transect contributes to the per-(UC, product, day) population only if at least one of its pixels survives this deduplication. The within-transect AIC test above is independent of the deduplication, since it uses each transect's own native-pixel sample."

2.2 Section 2.2, the four acceptance criteria

L. 91–100

A profile is accepted only when four conditions are met; three of the four thresholds are chosen for workflow justifications rather than values taken from theory.

The first is a model-evidence condition. We compare the two-segment fit against a single-segment fit using the Akaike information criterion (AIC), which rewards goodness of fit and penalises each additional parameter, and require $\Delta\mathrm{AIC} = \mathrm{AIC}_{\text{one-seg}} - \mathrm{AIC}_{\text{two-seg}} \geq 4$. A difference of four units is the

conventional point at which the poorer of two candidate models is regarded as having considerably less support (Burnham and Anderson, 2004). The condition checks whether the inflexion is worth the two parameters it costs.

The second is a slope-contrast condition, $|\beta_{\text{in}}|/|\beta_{\text{off}}| \geq 2.0$, so the inshore segment must be at least twice as steep as the offshore segment. Without it, a profile that warms almost uniformly offshore can satisfy the model-evidence test on a slight change in gradient, and be recorded as a cool tongue on that basis alone.

The third is an offshore-baseline condition, $\ell - x_b \geq 100\,\text{km}$, where ℓ is the streamline arc-length of the transect. It ensures that the background temperature $\widehat{T}_{\text{bg}}$ is estimated over a substantial stretch of open water rather than from the few pixels at the seaward end. The condition regularises the offshore fit, replacing the fixed reference distance of $UI_{\text{SST}}^{(d_0)}$ with a fitted breakpoint subject to an explicit constraint on the offshore extent.

The fourth is an edge-buffer condition, $x_b \geq 10\,\text{km}$ and $\ell - x_b \geq 10\,\text{km}$. A breakpoint at the coastline or the outer contour would leave one segment supported by too few points to estimate a slope, so the fit would be determined by the endpoints of the profile rather than by its shape.

The slope ratio, the offshore-baseline length, and the edge buffer are ours rather than the literature's. Their influence is bounded by the threshold-sensitivity assay of Table S2, which moves each threshold across a grid of alternatives. The cell-scale descriptions of Results Sect. 3.1 hold at every tested value.

> AS SUBMITTED. "We accept at fit when four joint criteria are met. First, $\Delta\text{AIC} = \text{AIC}_{\text{one-seg}} - \text{AIC}_{\text{two-seg}} \geq 4$, namely the 'considerably less support' guideline of Burnham and Anderson (2004) for the worse model. Second, the ratio of inshore-to-offshore slope magnitudes is at least 2.0, namely the coastal cooling segment is at least twice as steep as the offshore reference segment. Third, the offshore segment is at least 100 km in streamline arc-length ($\ell - x_b \geq 100\,\text{km}$), namely a minimum offshore-baseline constraint that regularises the offshore fit by preventing it from being estimated from a short tail of points. The 100 km value is a regularisation choice, which replaces the fixed offshore reference distance of $UI_{\text{SST}}^{(d_0)}$ with a fitted breakpoint x_b subject to this explicit regularisation. Fourth, the candidate breakpoint is at least an edge-buffer distance from each end of the transect ($x_b \geq 10\,\text{km}$ and $\ell - x_b \geq 10\,\text{km}$ in the grid), excluding breakpoint positions at the immediate coastline or at the outer contour where one segment would be supported by too few points to estimate a slope."

2.3 Section 2.3, counting the breakpoint as a parameter

L. 118–127

Treating x_b as a fourth free parameter alongside α_{in}, β_{in}, and β_{off} requires justification, because x_b is selected by search over a discrete grid of candidate positions rather than estimated within a fixed-grid likelihood. A search of this kind confers an advantage on the two-segment model that the standard penalty of two additional parameters may not fully offset. The change-point literature proposes stricter penalties for the same reason, namely the $k \ln n$ BIC-type penalty of Yao (1988), the structural-break criteria of Bai and Perron (1998), and the $(\ln n)^{1+\delta}$ modified-BIC correction of Zhang and Siegmund (2007).

We treat the choice of penalty as a sensitivity question. Re-scoring the test subset with the small-sample correction (AICc) and with the stricter BIC returns the same accept or reject decision as AIC on 100 % of fine-resolution profiles and on 99.5 % of coarse-resolution profiles (Table S3), and the few disagreements are stricter criteria rejecting marginal coarse-product fits. The four-parameter count is therefore adequate here, since the outcome is invariant across the penalty strengths tested.

> AS SUBMITTED. "Treating the breakpoint x_b as a fourth free parameter [...] is a choice that the standard AIC penalty does not strictly justify, because x_b is selected over a discrete grid rather than estimated within a fixed-grid likelihood. Stricter penalties have been developed for this case, namely the $k \ln n$ BIC-type penalty of Yao (1988) for change-point estimation, the structural-break information criteria of Bai and Perron (1998), and the $(\ln n)^{1+\delta}$ modified-BIC correction of Zhang and Siegmund (2007) for change-point detection on a discrete grid."

2.4 Section 2.3, how much independent information a profile carries

L. 128–139

Residual autocorrelation influences acceptance more heavily than the choice of penalty. The AIC comparison assumes that the observations entering the model are independent. However, cross-shore SST profiles violate that assumption, since adjacent native pixels covary strongly and n overstates the independent information a profile carries.

We quantify the dependence by fitting an AR(1) model to the OLS residuals along the unique-pixel sequence of each accepted profile, pooling lag-1 pairs within profiles per (UC, season) so that no across-transect pairs enter the estimate. The coefficient ϕ ranges 0.884 to 0.950 with a median of 0.915, so the dependence is strong and close to uniform across cells and seasons. Each ϕ converts to an effective sample size,

$$n_{\mathrm{eff}} = n\,\frac{1-\phi}{1+\phi}, \tag{1}$$

following Chelton et al. (1983), which reduces a 95-pixel profile to of order five effectively independent values, and we recompute AIC at n_{eff} for every (UC, product, season).

The correction severely reduces baseline acceptance by a median 20 to 24 % on the two fine-resolution products and leaves no accepted profiles on AVHRR_OI or MW_OI (Table S1). The coarse products therefore serve as resolution comparisons rather than as sources of breakpoint estimates, and every result below uses the AR(1)-corrected set.

> AS SUBMITTED. "The AR(1) coefficient ϕ ranges 0.884–0.950 with a median of 0.915, so the unique-pixel sample has a strong positive residual autocorrelation after the two-segment fit and the unique-pixel sample size n overestimates the independent-sample information by a large factor. Each ϕ converts to an effective sample size $n_{\mathrm{eff}} = n(1-\phi)/(1+\phi)$ (Chelton, 1983), and we recompute AIC with $n^{\star}_{\mathrm{eff}}$ at every (UC, product, season)."

2.5 Section 2.5, grouping days into events

L. 160

Days are grouped into events in two steps. A single inactive day flanked by active days on both sides is first reassigned as active, which prevents an isolated sub-threshold day from splitting one continuous period into two events. Runs of at least three consecutive active days are then counted as events. The order matters in the marginal case, so an active–inactive–active sequence bridges to a three-day run and counts as one event, whereas two consecutive inactive days terminate an event, since the bridging reads the original flag sequence and does not cascade. A bridged day counts towards $\mathrm{NUD}_{\mathrm{BR}}$ and contributes its daily-median $\bar{D}_{\mathrm{in}}$ to $\mathrm{TCU}_{\mathrm{BR}}$.

> AS SUBMITTED. "Events are sequences of at least three consecutive SST-structure-active days, with one interior inactive day permitted."

2.6 Section 2.7, comparing the two detectors

L. 190–196

The SST-structure detector and the Bakun-analogous wind detector are two classifiers of the same daily state, neither of which is the truth, so the comparison is one of agreement rather than validation. Per (UC, product, season) we report six agreement statistics for the matched pair, namely balanced accuracy, sensitivity and specificity taken in both directions, raw agreement p_o, Cohen's κ, and the Matthews correlation coefficient (MCC) (Table S5). No single statistic characterises a 2×2 agreement table completely, which is why we report the set.

Cohen's κ requires a caveat at the perennial-wind UCs. It measures agreement in excess of that expected by chance, $\kappa = (p_o - p_e)/(1 - p_e)$, and p_e approaches unity when one detector returns the same state on nearly every day. The numerator and denominator both become small, and κ is driven towards zero

even where raw agreement is high. This compression under saturated marginals is documented (Byrt et al., 1993; Chicco and Jurman, 2020), and the MCC is less sensitive to it. We therefore quote the MCC as the summary statistic and report κ alongside it, with the κ against wind-marginal relationship in Fig. S3.

AS SUBMITTED. "Per (UC, product, season), we report six agreement statistics on the matched pair, namely balanced accuracy [...], sensitivity [...], specificity (the same, complementary), Matthews correlation coefficient (MCC) (Matthews, 1975), raw agreement p_o, and Cohen's κ. MCC is the summary statistic at perennial-wind UCs where Cohen's κ is compressed by saturated marginals (Byrt et al., 1993; Chicco and Jurman, 2020)."

2.7 Section 3.2, the correlation interval and the multiple testing

L. 269–275

At Lüderitz the annual aggregate r_{50} in OSTIA is −0.16, so across the year the fitted inshore drawdown and the 50 km fixed-distance contrast vary in opposite senses. The interval on that estimate is wide. Because adjacent transects are not independent, we resample in alongshore blocks of 50 km, and the block-bootstrap 95 % interval goes from −0.21 to +0.13 (Table S9). The annual sign inversion therefore holds in the point estimate while the interval encompasses zero, and we do not treat it as established.

Seasonal aggregates resolve the pattern more sharply. Forty (UC, product, season) combinations were tested, and at that number some will return a negative r_{50} by chance, so we control the false discovery rate by the Benjamini–Hochberg procedure, which converts each p value into an adjusted q value bounding the expected proportion of false positives among the combinations declared significant. Four pass at $q < 0.001$, namely Cape Peninsula in MAM on both fine-resolution products, Lüderitz in MAM on OSPO, and Walvis Bay in SON on OSPO.

AS SUBMITTED. "The annual aggregate r_{50} at Lüderitz in OSTIA is −0.16 with an alongshore 50 km block-bootstrap 95 % interval of [−0.21, +0.13] (Table S9 in the Supplement), namely the annual sign-inversion holds in point estimate but the interval encompasses zero. [...] The seasonal r_{50} tests are Benjamini–Hochberg adjusted to control the false discovery rate (FDR) across the 40 UC-product-season tests."

Remaining passages

The same editorial work will be made to the 0.5 activity threshold that classifies a day at a cell, the block bootstrap and why blocks are needed at all, the date-shuffled null that sets the chance baseline for detector agreement, and the per-profile bootstrap and bimodality test that establish whether an individual breakpoint is well determined. We have not reproduced those here only to keep this document to a reasonable length.

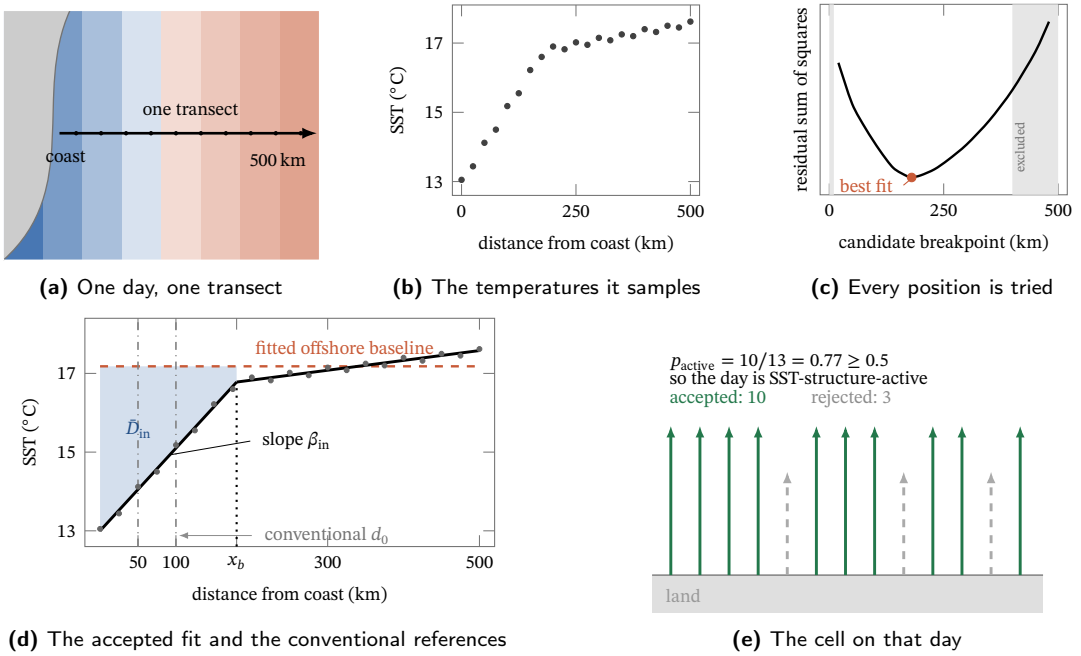
3 The proposed walk-through figure

Figure 1 is the figure the referee asked for. It follows one transect, on one day, at one cell, from the temperature field through to the classification of the day. It is drawn here as a schematic with illustrative values so that each step is legible. We propose to place it in the Methods, after the acceptance criteria, and we are equally happy to move it to the Supplement if the editor prefers a shorter main text (we prefer for it to remain in the main text, though!).

4 Point-by-point replies

Abstract

L. 6, product acronyms rather than full names. Agreed, and it also shortens a crowded abstract. We will use OSTIA, OSPO, AVHRR_OI, and MW_OI on first mention with the full names given in the Methods only. See in the revised Abstract, below.



(a) One day, one transect **(b)** The temperatures it samples **(c)** Every position is tried

(d) The accepted fit and the conventional references **(e)** The cell on that day

Figure 1: How a single daily breakpoint is calculated, from the temperature field to the classification of the day. (a) One transect of the cross-shore mesh crosses the SST field on one day, and one temperature value is taken per satellite pixel along it. (b) Those values form the cross-shore profile that is fitted. (c) Every candidate breakpoint position is tried in turn and the position giving the smallest residual sum of squares is kept; the shaded bands at either end are excluded by the edge-buffer and offshore-baseline criteria. (d) The accepted two-segment fit, with the breakpoint x_b, the inshore slope β_{in}, the fitted offshore baseline, and the shaded inshore drawdown $\bar{D}_{in}$; the conventional 50 km and 100 km reference distances are marked and fall inside the inshore segment. (e) The fit is repeated at every transect in the cell, and the fraction accepted, p_{active}, classifies the day. Values are illustrative and drawn for legibility.

L. 8, mention the 50 to 100 km convention in the first sentence. Agreed. Naming the convention in the opening sentence states the problem before the solution, which is the right order and was not what we did.

L. 10–15, the results are not followable from the abstract alone. This is the most useful comment we received on the abstract. We accept all three parts. The symbol r_{50} was never defined in the abstract, so we have removed the notation and said the thing in words. The one-fifth sentence was *really* ambiguous: it was not meant to report a poor match to an established method but to report that the two detectors describe different things on a measurable minority of days, which is a finding and not a shortfall. The closing sentence did read as a criticism of the existing indices, and that was a failure of emphasis on our part rather than our intention. The revised version closes on what the method provides.

Revised abstract.

> Satellite sea-surface-temperature (SST) upwelling indices measure the contrast between coastal water and an offshore reference point, and that reference is conventionally fixed at 50 to 100 km from the coast. The surface cool tongue, however, reaches different distances offshore at different cells and on different days, so no single fixed distance can sit beyond it everywhere. We replace the fixed reference with a thermal breakpoint fitted each day to each cross-shore SST profile by a two-segment regression, accepted only when the fit satisfies criteria on model evidence, slope contrast, offshore baseline length, distance from the ends of the profile, and residual autocorrelation. Applying the method to five Benguela upwelling cells over 2015 to 2025 on two 0.05° analyses (OSTIA and OSPO), with two 0.25° analyses (AVHRR_OI and MW_OI) as resolution comparisons, we find median fitted breakpoints of 114 to 270 km. Conventional references at 50 to 100 km therefore sample water inside

the cool tongue rather than the water beyond it, and the consequence is measurable. In OSTIA, the daily correlation between the fitted inshore temperature drawdown and the conventional fixed-distance contrast rises from 0.06 when the reference sits at 50 km to 0.93 when it sits at 500 km, so moving the reference changes what the index measures rather than simply rescaling it. Comparing the fitted SST structure against a Bakun-type offshore-Ekman-transport wind flag day by day, the two agree on most days but part company on about one fifth of days at the southern Benguela cells, where the surface structure persists while the local wind flag is absent. Fitting the reference rather than fixing it therefore yields cell-specific cross-shore length scales, an SST contrast that can be compared across cells, and an explicit four-class description of when surface structure and local wind forcing coincide.

1 Introduction

L. 36, UC used before it is defined in the Introduction. Correct, and we will give "upwelling cells (UCs)" in full at first use in the Introduction. The abbreviation was defined in the abstract only, which did not help the rest of the manuscript.

2 Methods

L. 60, a reference for the Laplace and Dirichlet construction. A fair request. The construction is standard elliptic grid generation: Laplace's equation solved on a domain with Dirichlet conditions on two facing boundaries is the classical way to build a boundary-conforming curvilinear coordinate system, described in general by Spekreijse (1995), and used in ocean modelling to generate coastline-following orthogonal curvilinear grids (Chan et al., 1994; Murray, 1996; Liu et al., 2017). We will cite these.

In ocean modelling the streamlines of the potential become model grid lines. Here we use them as analysis transects instead, so that each transect crosses the shelf approximately normal to the local coast orientation and neighbouring transects do not intersect. The underlying mathematics is long established, but the application to a cross-shore SST transect mesh is, as far as we are aware, our own, and we will present it that way rather than implying an existing precedent.

L. 93, 'worse' → 'worst'. We will keep as is. Only two models are being compared, the one-segment and the two-segment fit, and English takes the comparative for two and the superlative for three or more. That said, the construction is awkward enough to have prompted the comment, so we have removed it. The rewritten sentence (Section 2) reads "the conventional point at which the poorer of two candidate models is regarded as having considerably less support", which makes the two-way comparison explicit.

L. 131, 'austral seasons'. Agreed, and we will delete "austral" here. The sentence refers to all seasons generically, so the hemisphere qualifier carries no information.

L. 160, does one positive day, a gap, then another positive day flag as an event? Yes, it does. The submitted sentence was ambiguous about this case. The rule as implemented has two steps in order. A single inactive day flanked by active days is first converted to active. Runs of three or more active days are then counted as events. An active–inactive–active sequence therefore becomes a three-day run and does count as an event. Two consecutive inactive days never bridge, and the bridging does not cascade, so it cannot chain a long alternating sequence into one event. A bridged day also counts towards $\mathrm{NUD_{BR}}$ and contributes to $\mathrm{TCU_{BR}}$. The replacement text in Section 2 states all of this.

3 Results

L. 284, 'austral-summer'. Agreed, and we will drop "austral" throughout the Results where a named season already appears next to its month abbreviation. DJF removes any ambiguity on its own.

L. 285, figures should not be the grammatical subject. Agreed, and this is a better convention than ours. We will rewrite every such sentence so that the result leads and the figure is cited in parentheses. The three instances in the submitted manuscript become, for example, "Paired SST and wind daily classifications separate into four classes of unequal size (Fig. 4)" in place of "Paired SST–wind daily-classifications are in Figure 4", and similarly for the two sentences introducing Figures 2 and 3.

Supplement

Table text size varies with no clear reason. The referee is right, and there is no principled reason. The sizes were set one table at a time to make each fit its page width, which produced `\scriptsize` in seven tables, `\footnotesize` in Tables S2 and S6, and `\small` in Table S3. We will set every supplement table in `\footnotesize` and rotate the two widest to landscape rather than shrink them further, so that the type size is uniform and legibility does not depend on which table a reader happens to open. But we think the journal's professional typesetting will take care of this (unless it does not typeset supplementary material—we will fix, regardless).

Table 1

Units are given twice. Agreed. The final sentence of the caption, which repeats the units already stated in the column headers, will be deleted.

Figure 1

We are glad the mesh and the palette work. Thank you.

Figure 2

The legend overlaps the panel titles. Thanks. It will be fixed.

Stray commas and backslashes in the panel titles. A very good catch! The titles will read $d_0 = 50\,\mathrm{km}$ and so on in the revised figure.

Cape Frio points are overplotted. A real limitation and we thank the referee for noting that transparency is already in use. Two things can still be done. The plotting order is currently fixed, so the same cell is always drawn last and always sits on top; we will randomise the draw order so that no cell is systematically hidden. We will also reduce the marker size and add a light per-cell density contour, which recovers where the mass of the Cape Frio distribution lies even where individual points are obscured.

Figure S3

Why does the y-axis reach 0.35? Because that is where the data end. Cohen's κ can in principle take any value between -1 and 1, but the largest value anywhere in this dataset is just under 0.3, and the axis limit is simply the plotting range with a small margin. The limited range is itself the result the figure exists to show, namely that κ never becomes large in these comparisons however well the two detectors agree. We will add the observed maximum to the caption so that the reader is not left inferring the axis limit, and we are happy to extend the axis to the full -1 to 1 range if the referee prefers, though that would compress the structure the figure is meant to display.

Thank you to the referee again for the time taken and for a review that has improved the manuscript in ways we would not have noticed on our own.

New references proposed for the revision

These are the sources we propose to add in reply to the comment at L. 60. All other citations named in this document are already in the manuscript reference list.

Chan, C.-T., Cheong, H. F., and Shankar, N. J.: A boundary-fitted grid model for tidal motions: orthogonal coordinates generation in 2-D embodying Singapore coastal waters, Computers & Fluids, 23, 1994.

Liu, X., Wang, X., Xu, H., and Zhang, X.: On the generation of coastline-following grids for ocean models, trade-off between orthogonality and alignment to coastlines, Ocean Dynamics, 67, 2017.

Murray, R. J.: Explicit generation of orthogonal grids for ocean models, Journal of Computational Physics, 126, 251–273, 1996.

Spekreijse, S. P.: Elliptic grid generation based on Laplace equations and algebraic transformations, Journal of Computational Physics, 118, 38–61, 1995.