Advances in
Geosciences

Open Access

Displacement-based error metrics for morphodynamic models

J. Bosboom[1] **and A. J. H. M. Reniers**[2,1]

[1]Faculty of Civil Engineering and Geosciences, Delft University of Technology, the Netherlands
[2]Applied Marine Physics, Rosenstiel School of Marine & Atmospheric Science, University of Miami, USA

Correspondence to: J. Bosboom (j.bosboom@tudelft.nl)

Abstract. The accuracy of morphological predictions is generally measured by an overall point-wise metric, such as the mean-squared difference between pairs of predicted and observed bed levels. Unfortunately, point-wise accuracy metrics tend to favour featureless predictions over predictions whose features are (slightly) misplaced. From the perspective of a coastal morphologist, this may lead to wrong decisions as to which of two predictions is better. In order to overcome this inherent limitation of point-wise metrics, we propose a new diagnostic tool for 2-D morphological predictions, which explicitly takes (dis)agreement in spatial patterns into account. Our approach is to formulate errors based on a smooth displacement field between predictions and observations that minimizes the point-wise error. We illustrate the advantages of this approach using a variety of morphological fields, generated with Delft3D, for an idealized case of a tidal inlet developing from an initially very schematized geometry. The quantification of model performance by the new diagnostic tool is found to better reflect the qualitative judgement of experts than traditional point-wise metrics do.

1 Introduction

Quantitative validation methods for morphodynamic models are often grid-point based; they compare observations and predictions per grid-point and compute various metrics for the entire set or subset of grid-points (e.g., Sutherland et al., 2004). Unfortunately, point-wise accuracy metrics, such as the commonly used MSE (Mean-Squared Error) and RMSE (Root-Mean-Squared Error), tend to penalize, rather than reward, the model's capability to provide information on features of interest, such as scour holes, accumulation zones and migrating tidal channels. For instance, a prediction of a morphological feature that is correct in terms of timing and size,

but is misplaced in space, may not outperform even a flat bed, which is inconsistent with the common judgement of morphologists (Fig. 1). This "double penalty effect" (Bougeault, 2003), which applies in full when a feature is misplaced over a distance equal or larger than its size, makes it difficult to demonstrate the quality of a high variability prediction (Anthes, 1983; Mass et al., 2002). Clearly, a high quality validation process requires alternative validation techniques that account for the spatial structure of 2-D morphological fields.

For the verification of weather variables (e.g. precipitation), methods are being actively developed to quantify forecast performance based on spatial structure; see for instance Casati et al. (2008) and Gilleland et al. (2009) for an overview. One of the approaches in meteorology, now also pioneered in other fields (e.g., Haben et al., 2014; Ziegeler et al., 2012), is to find an optimal deformation of the predictions that minimizes the misfit with observations. This optimal deformation can be obtained by employing one of many existing image matching methods, of which optical flow techniques, designed to estimate motion, are probably most well-known in the coastal community. The result of the image matching or warping is a vector field of displacements, which can be regarded as a displacement error field. In addition, an intensity or amplitude error field may be defined as the difference between the deformed prediction and the observations (e.g., Marzban and Sandgathe, 2010), which can be seen as the point-wise error if no penalty applies for misplacements.

Existing verification methods, based on field deformation of meteorological fields, not only differ in the applied image matching method, but also in the approach to the subsequent extraction of map-mean errors. Keil and Craig (2009) determine RMS (root-mean-squared) intensity and mean displacement errors within the boundaries of precipitation features, which they then combine into a single error metric.
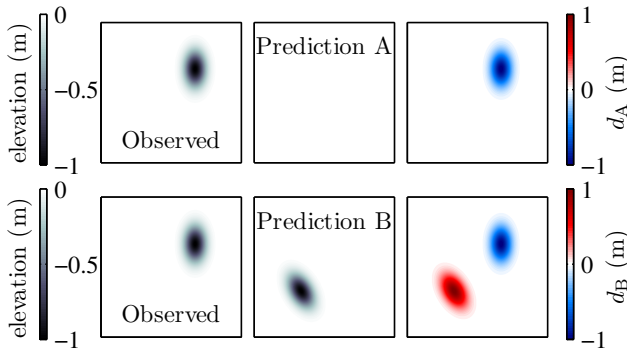
Fig. 1. The "double penalty effect". Top panels: the featureless prediction A has a non-zero difference d_A between predicted and observed depth values at the location of the observed feature only. Lower panels: prediction B, which reproduces the feature at the wrong location, is penalized twice (d_B is non-zero both where the predicted feature is and where it should be) and is thus diagnosed with a twice as large (R)MSE as prediction A.

The latter requires the normalization of the two errors to put each term on equal footing, which introduces two parameters to the formulation. In contrast, Gilleland et al. (2010) propose a combined error metric that besides the post-warp RMS intensity error and the mean displacement error also takes the original RMS intensity error into account, enabling a more fair comparison of forecast performance. Their metric, however, is not easily applicable since it requires three user-chosen weights that are dependent on the error terms themselves.

The goal of this paper is to quantify morphodynamic model performance, while taking the spatial characteristics of 2-D morphology into account. Using a field deformation technique, we have developed and tested a new diagnostic tool for the validation of 2-D morphological predictions. It includes a location (displacement) error metric and a robust and physically intuitive combined error metric that incorporates both location and intensity error. The combined metric rewards predictions to the degree that a larger error reduction can be obtained with smaller displacements. As a reference, we use the subjective but very powerful method of visual inspection of morphological patterns by experts.

Our method is outlined in Sect. 2, along with a brief description of the image warping method that we have adopted to calculate the optimal deformation. Next, in Sect. 3, we put the new diagnostic tool to the test, using morphological fields generated with Delft3D for an idealized case of a tidal inlet developing from an initially very schematized geometry. Section 4 concludes with a summary of our findings and the implications for morphodynamic model validation.

2 Method

This section outlines our two-step approach to quantify the (dis)agreement between 2-D morphological patterns. Section 2.1 describes the first step of deforming (or warping) the predicted morphology to minimize the point-wise error with observations. Next, Sect. 2.2 formulates two new error metrics, a mean location error that is distilled from the displacement vector fields and a single-number error metric that measures both the correspondence with respect to location and intensity (i.e. depth-values).

2.1 Warping method

The measure of closeness between images or spatial fields is encountered in many fields from radiography to meteorology. This has led to the development of a multitude of image matching methods that, depending on the scientific field, are also named registration or warping methods. The goal of such methods is to find the optimal transformation that maps each point of a static image to a corresponding point (with the same intensity) in the moving image. Within the context of morphodynamic model validation, the static image represents the observed depth field o and the moving image the predicted depth field p.

Of all the available techniques, the class of optical flow techniques, designed to estimate small displacements in temporal image sequences, is probably the most well-known in our field. The basic assumption of optical flow is that the intensity of a moving object does not change appreciably in the considered time interval. We employ the efficient, non-rigid (i.e. allowing for free-form deformations) registration technique named Demon's registration (Thirion, 1998), which bears similarities to optical flow, in an implementation by Kroon and Slump (2009). The Demon's approach can be considered as similar to a minimization of the sum of square image intensities between the deformed predictions and observations (Pennec et al., 1999). It is therefore consistent with our quest to find the optimal deformation of the predictions that minimizes the point-wise (R)MSE.

The estimated backward pixel displacements $\boldsymbol{B}^* = (B_x^*, B_y^*)$ that are required for a given point in a static image (the observations in our validation context) to match the corresponding point in a moving image (the predictions) is given by Thirion (1998):

$$\boldsymbol{B}^* = \frac{(I_p - I_o)\nabla I_o}{|\nabla I_o|^2 + \alpha^2(I_p - I_o)^2} \tag{1}$$

in which α is a normalization factor that is equal to 1 in the original method and I_o and I_p are the intensities of the static and moving image, respectively. The latter are taken as the observed and predicted depth fields, normalized by scaling between 0 and 1. Since Eq. (1) is based on local information, it is solved iteratively while including Gaussian smoothing as a regularization criterion. This ensures that a

realistic, smooth displacement field is found instead of an irregular field that nonetheless minimizes the sum of squares. The normalization factor is chosen as $\alpha = 2.5$ in line with Kroon and Slump (2009) and the standard deviation of the Gaussian smoothing window as $\sigma = 4$. These parameters are kept constant for all registrations presented in Sect. 3. The forward displacements $\boldsymbol{F}^* = (F_x^*, F_y^*)$ from the moving to the static image can be determined from $\boldsymbol{B}^*$ after the registration. Note that when in the following the subscript $*$ is dropped, we refer to the displacement fields transferred to a physical distance.

For the purpose of model validation, we interpret $d_0 = p_0 - o$, with p_0 the prediction prior to warp, as the total point-wise error and $d_1 = p_1 - o$, with p_1 the deformed prediction as follows from the registration, as the point-wise error if no penalty is imposed for location disagreement. Next, we use this perspective in the formulation of map-mean errors.

2.2 Formulation of new error metrics

From the Demon's registration (Sect. 2.1), we obtain the optimal displacement vector field between predictions and observations as well as the optimal deformation of the predictions. "Optimal" in this context means that the sum of squares between the deformed predictions and observations is minimized, such that $0 \leq \text{RMSE}_1 \leq \text{RMSE}_0$, where RMSE_0 and RMSE_1 are the root-mean-squared errors before and after the warp, respectively. Note that we have preferred the RMSE over the MSE, since the first is measured in the same units as the data. Out of two predictions that have the same RMSE_0, a prediction that has similar morphological features as the measurements, albeit displaced, may receive a lower RMSE_1 than a prediction that is not able to reproduce the observed morphological features at all. Thus, the RMSE_1 is expected to diagnose the agreement between morphological fields if a zero penalty is imposed for misplacements of features. However, which of the two predictions is valued the better prediction by morphologists not only depends on RMSE_0 and RMSE_1, but also on the magnitude of the displacements required to obtain the error reduction. Therefore, we expect that the similarity in both location and intensity between morphological patterns can be fully assessed using three error metrics in concert: RMSE_0, RMSE_1 and a mean location error $\bar{D}$ that we will formulate next from the displacement vector fields.

It is tempting to define $\bar{D}$ as the arithmetic mean of $D = \sqrt{(B_x{}^2 + B_y{}^2)}$, the field of displacement magnitudes. However, it should be realised that the optical flow problem is underconstrained; for a single grid-point, we only have information on the displacements normal to the contour lines, whereas along the contour lines the displacements are ambiguous (the so-called aperture problem). In the Demon's approach, the Gaussian smoothing acts as the necessary additional constraint, requiring that nearby grid-points have similar displacements. As a consequence, non-zero displace-

ments may be found along depth contours in morphologically inactive regions (see Sect. 3), whereas these displacements do not improve the match between the deformed prediction and the observations. Therefore, we propose a weighted mean location error that weights the local backward displacement magnitudes D with their effect on the reduction of the local squared error. In this way, displacements are only taken into account to the extent that they contribute to the minimization of the sum of squares. This yields:

$$\bar{D} = \frac{\sum_{i=1}^{n} w_i D_i}{\sum_{i=1}^{n} w_i}; \quad w_i = \frac{\text{SE}_{0,i} - \text{SE}_{1,i}}{\sum_{i=1}^{n} (\text{SE}_{0,i} - \text{SE}_{1,i})} \qquad (2)$$

Here $\text{SE}_0 = (p_0 - o)^2$ and $\text{SE}_1 = (p_1 - o)^2$ are the local squared errors before and after the warp, respectively, n is the number of equidistant points in the spatial domain and $\sum_{i=1}^{n} w_i = 1$. Note that $\text{RMSE}_j = \sqrt{n^{-1} \sum_{i=1}^{n} \text{SE}_{j,i}}$, with $j = [0, 1]$.

Whereas model performance is usually diagnosed based on RMSE_0 only, we now have two additional metrics RMSE_1 and $\bar{D}$. In Sect. 3, it is demonstrated that considering these three metrics in concert allows a full assessment of model quality, avoiding the double penalty effect for misplaced features. In practice, guidance may be required on how to weight these three metrics. Besides, the morphologist may sometimes desire a single-number summary of model performance, especially if automated calibration routines are used. To serve these needs, we propose an adjusted RMS error measure, RMSE_w, that is computed from a field of weighted squared errors SE_w. The latter are determined by locally weighting SE_0 and SE_1. The purpose of the weighting procedure is to locally relax the requirement of an exact match to an extent determined by the local displacement magnitude. Figure 2 illuminates the weighting procedure for the ith gridpoint; an error reduction is awarded that is a fraction $1 - \delta_i$ of the full error reduction potential $(\text{SE}_{0,i} - \text{SE}_{1,i})$. Here, $\delta_i = D_i / D_\text{max}$ and D_max is a maximum displacement length above which no relaxation is allowed. A larger fraction $1 - \delta_i$ is allowed for smaller displacement magnitudes D_i, with a maximum of $1 - \delta_i = 1$ and thus $\text{SE}_{\text{w},i} = \text{SE}_{1,i}$ for $D_i = 0$ m. For $D_i \geq D_\text{max}$, we have $1 - \delta_i = 0$ and thus $\text{SE}_{\text{w},i} = \text{SE}_{0,i}$. Note that D_max is a user-defined, physically intuitive parameter that is dependent on the prediction situation and the goal of the simulation. It can be seen as the maximum distance over which morphological features may be displaced for the prediction to still get (some) credit for predicting these features. We now have for RMSE_w:

$$\text{RMSE}_\text{w} = \sqrt{\frac{\sum_{i=1}^{n} \text{SE}_{\text{w},i}}{n}} \qquad (3)$$

where

$$\text{SE}_\text{w} = \text{SE}_1 + \delta (\text{SE}_0 - \text{SE}_1) \qquad (4)$$

$$\delta_i = \frac{D_i}{D_\text{max}} \text{ for } D_i \leq D_\text{max}; \quad \delta_i = 1 \text{ for } D_i > D_\text{max} \qquad (5)$$
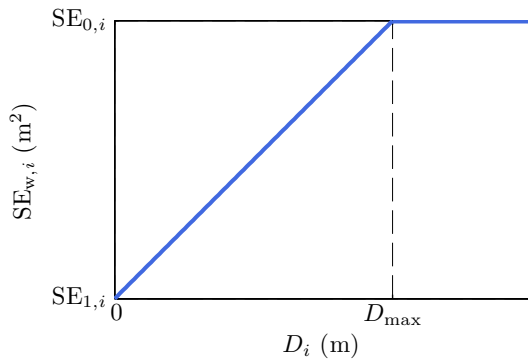
Fig. 2. Weighted squared error for the ith gridpoint $SE_{w,i}$, which is the sum of the local squared error after the warp $SE_{1,i}$ and a penalty for misplacements $\delta(SE_{0,i} - SE_{1,i})$ with $\delta = D_i/D_{max}$. The penalty ranges from 0 for $D_i \to 0$ to $(SE_{0,i} - SE_{1,i})$ for $D_i = D_{max}$, a user-defined maximum displacement length. For $D_i \geq D_{max}$ the full point-wise error applies and $SE_{w,i} = SE_{0,i}$.

In conclusion, $RMSE_w$ as an error metric rewards forecasts to the degree that a larger error reduction can be obtained by smaller displacements. By definition, $RMSE_1 \leq RMSE_w \leq RMSE_0$. If the error reduction due to the image deformation is negligible or can only be obtained with displacements equal to or larger than D_{max}, the diagnosed error is equal to the original error prior to the deformation $RMSE_0$. If, on the other hand, the displacements required to minimize the point-wise error are very small relative to D_{max}, we have $RMSE_w \approx RMSE_1$. The justification for this approach lies in the tendency of coastal morphologists to credit a prediction for the reproduction of features, albeit displaced, while imposing a relatively small penalty for misplacement. The intuitive weighting of these two aspects is mimicked by the user-defined parameter D_{max}.

3 Application

Below, the new error metrics are used to diagnose the correspondence between model-generated pairs of morphological patterns for an idealized tidal inlet as well as the relative ranking between the pairs. The fields have been generated for the idealized case of a tidal inlet developing from an initially very schematized geometry (Roelvink, 2006). First, Sect. 3.1 demonstrates that the location error $\bar{D}$ is able to capture the overall misplacement of the morphological patterns. Next, in Sect. 3.2, the combined error metric $RMSE_w$ is put to the test. Two examples are shown where the $RMSE_w$ makes the right the decision as to which of two predictions is the better prediction while the conventional, purely point-wise $RMSE_0$ fails to do so.
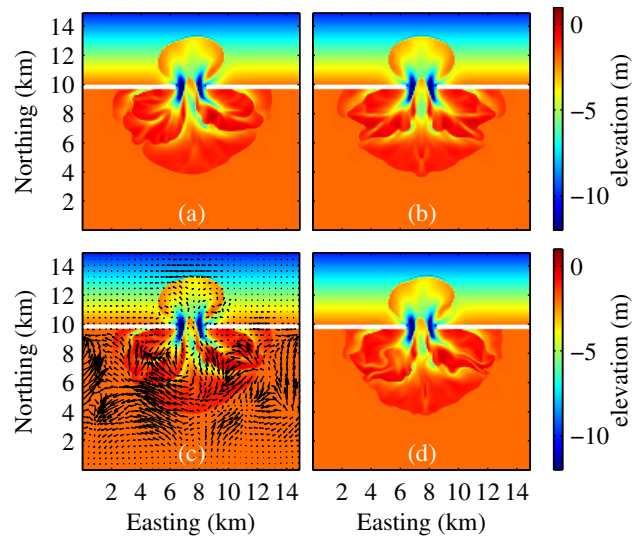

Fig. 3. Example of the image warp: **(a)** the "observations", calculated using Delft3D with Coriolis at 53° N; **(b)** the predictions, calculated at 0°; **(c)** the backward displacement vector field $\boldsymbol{B}$ of the observations towards the predictions, shown on top of the observations; and **(d)** the predictions deformed to more closely match the observations.
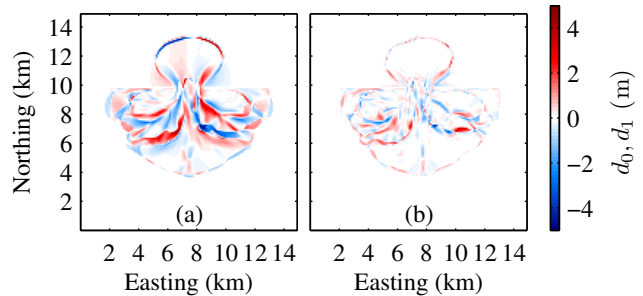

Fig. 4. Point-wise error fields for the predicted depth field at 0°: **(a)** the total error $d_0 = p_0 - o$ before the warp; **(b)** the error $d_1 = p_1 - o$ after the warp, to be regarded as the remaining point-wise error if no penalty applies for location disagreement.

3.1 Location error

In this subsection, we consider a subset of the model-generated depth fields which only differ with respect to the latitude, and hence Coriolis parameter, used in the model. Of four depth fields, we label the field generated at 53° N as the "observations" (Fig. 3a) and consider the other fields, for latitudes 90° N, 0° and 90° S, as three competing predictions. Even though the predictions are not shown here, it will not come as a surprise that the point-wise error $RMSE_0$ is smallest for 90° N and largest for 90° S (Table 1).

In order to determine $RMSE_1$ and $\bar{D}$, the image warping method is applied, following the procedure outlined in Sect. 2, and illustrated here for the prediction at 0° (Fig. 3b). The deformed prediction that matches the observations most

Table 1. Errors for competing predictions that differ with respect to the latitude, and thus the Coriolis parameter, used in Delft3D. The model results for 53° N are regarded as the "observations".

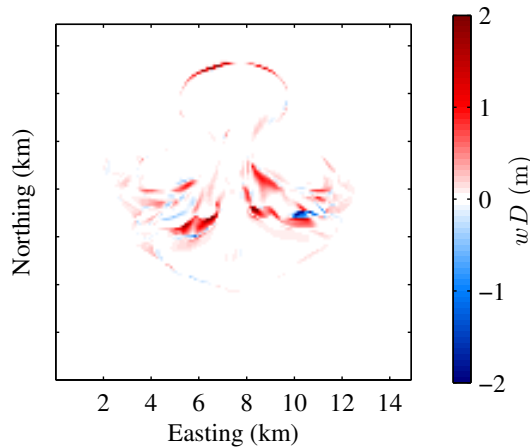| Latitude	RMSE_0 (m)	RMSE_1 (m)	$\bar{D}$ (m)
90° N	0.29	0.12	180
0°	0.52	0.26	350
90° S	0.73	0.35	710

Fig. 5. Weighted displacements wD for the prediction at 0°. Here $D = \sqrt{(B_x{}^2 + B_y{}^2)}$ is the field of displacement magnitudes computed from the backward displacement vector field $\boldsymbol{B}$ (see Eq. 1), and w is determined according to Eq. (2).

closely is shown in Fig. 3d and the corresponding backward vector displacement field $\boldsymbol{B}$ in Fig. 3c. As explained in Sect. 2.1, in the inactive outer regions, physically unrealistic displacements are found along depth contours, since no penalty is imposed in the minimization for displacements along depth contours. As will be illustrated next, this is solved for in the formulation of $\bar{D}$ (Eq. 2).

The difference d_0 between the predictions prior to the warp and the observations is shown in Fig. 4a, whereas Fig. 4b shows the difference d_1 after the warp. Note that taking the root-mean-square of d_0 and d_1 yields RMSE_0 and RMSE_1, respectively. From d_0, the double penalty problem is clearly observed; for instance at the edges of the ebb-tidal delta, an error is diagnosed both where the delta is present in the observations but absent from the predictions and vice versa. After the warp, both errors have practically disappeared, such that they will not count towards RMSE_1, demonstrating again that RMSE_1 should be regarded as the point-wise error if no penalty for misplacement is taken into account. For the prediction at 0°, $\mathrm{RMSE}_1/\mathrm{RMSE}_0 = 0.5$, and slightly smaller ratios are found for the other two predictions (Table 1).

The weighted dispacements wD, with $D = \sqrt{(B_x{}^2 + B_y{}^2)}$ and w according to Eq. (2), are shown in Fig. 5. Inherent to the use of the squared error to determine w is that larger error reductions are heavily weighted. Here, we have never-

Table 2. Subjective ranking (with 1 being the best prediction) and errors for competing predictions, generated with Delft3D for various boundary conditions. The "observations" are taken as the model outcome at 0° (cf. Sect. 3.1). The values for RMSE_w hold for $D_{\max} = 3000\,\mathrm{m}$.

Predic-tion	Rank-ing	RMSE_0 (m)	RMSE_1 (m)	$\bar{D}$ (m)	RMSE_w (m)
A	1	0.78	0.38	610	0.49
B	2	0.77	0.53	770	0.60
C	3	1.16	0.56	860	0.78
D	4	0.96	0.77	1230	0.84

theless chosen this weighting since squared errors are consistent with the minimization as performed by the registration method as well as with the use of the (R)MSE as the point-wise metric, which is common in morphodynamic model validation. Note that for the computation of $\bar{D}$ (Eq. 2), we require the backward (from the observations to the predictions) rather than the forward displacements; for each point in the observational domain, these provide the distance at which the point in the predictions is located that is shifted to the considered location in the observations. Summing wD for the entire domain yields a location error $\bar{D} = 350\,\mathrm{m}$ at 0° (Table 1).

The values for $\bar{D}$ for the three predictions demonstrate a qualitative behaviour consistent with the error in latitude and hence Coriolis effect in the various predictions (Table 1). In fact, all three error metrics, RMSE_0, RMSE_1 and $\bar{D}$ diagnose the predictions for 90° N and 90° S as the best and worst predictions, respectively. Next, we will consider situations in which a ranking consistent with expert judgement is only obtained by considering these three metrics in concert, using an appropriate weighting, or from RMSE_w.

3.2 Ranking according to the combined error metric

In this subsection, we present an example, again using depth fields generated with the Delft3D model of the schematized tidal inlet, that demonstrates that RMSE_w outperforms the traditional score RMSE_0. Now, the model results at a latitude of 0° (see Sect. 3.1) are assumed to be the "truth". Four competing predictions are considered that are generated at 0° with various changes to the model boundary conditions (w.r.t. tidal amplitude and flow direction). Figure 6 shows the four predictions, the "observations" and the deformed predictions that minimize the point-wise error.

We have labelled the predictions according to a subjective ranking based on visual inspection, with A the prediction with the closest match with the observations and D, the worst prediction. We have a slight preference for prediction B over C, but it is possible that other morphologists would tend to regard C as the better prediction. Not surprisingly, the relative ranking as diagnosed by RMSE_0 deviates from the expert ranking (Table 2); based on RMSE_0 one would wrongfully
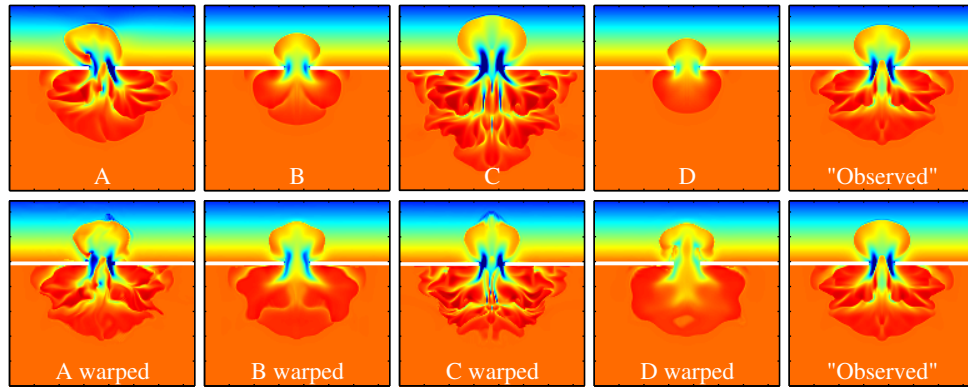
Fig. 6. Predictions A, B, C and D, the "observations" (taken as the model results for 0°) and the corresponding deformed predictions that minimize the point-wise mismatch between predictions and observations. The labels are chosen such that the lower the label in the alphabet, the higher the quality that the prediction is probably diagnosed with upon visual inspection. The axes are as in Fig. 3.
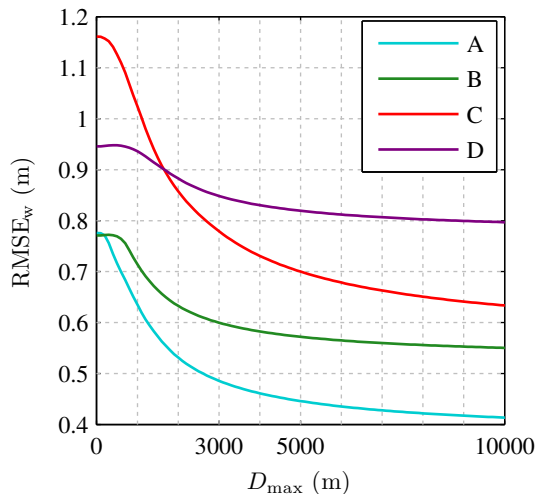

Fig. 7. The combined error metric $RMSE_w$ as a function of D_{max} for predictions A, B, C and D. The larger D_{max}, the more the requirement of an exact match is relaxed; for $D_{max} \to 0$, we have $RMSE_w \to RMSE_0$ and for $D_{max} \to \infty$, we have $RMSE_w \to RMSE_1$.

conclude that predictions A and B perform equally well and that prediction D outperforms prediction C.

The values of $RMSE_1$, $\bar{D}$ and $RMSE_w$ for the respective predictions provide the necessary additional information on model performance (Table 2). The smaller $RMSE_1$ for prediction A than for prediction B shows that if no penalty is imposed for misplacements, prediction A receives a lower error than B. Moreover, a smaller average displacement $\bar{D}$ is required to minimize the point-wise error. Thus, even though no distinction can be made based on $RMSE_0$, we can conclude that pattern A more closely corresponds to the observations than pattern B. Clearly, considering the values of $RMSE_0$, $RMSE_1$ and $\bar{D}$ in concert leads to a diagnosis of

relative model performance of A and B in line with visual inspection.

To determine $RMSE_w$, a value for D_{max} must be chosen. A defendable choice would be to limit D_{max} to the scale of the morphological features of interest. For this particular case, $D_{max} = 3000$ m is considered appropriate, being in the order of magnitude of the seaward extent of the ebb-tidal delta. In general, of course, D_{max} must be chosen in accordance with the goal of the simulation.

Figure 7 shows that with $D_{max} = 3000$ m, $RMSE_w$ reports a higher quality for prediction A than for prediction B, regardless of the exact choice for D_{max}. Only if one decides to not allow any relaxation of the requirement of an exact match ($D_{max} = 0$ m), $RMSE_w$ is identical to the full point-wise error $RMSE_0$ and no distinction can be made between A and B. If one wishes to allow the full error reduction potential ($D_{max} \to \infty$), we have $RMSE_w = RMSE_1$.

Table 2 illuminates that prediction C, the prediction with the largest $RMSE_0$, has a much larger potential for error reduction by warping than prediction D; notwithstanding the larger $RMSE_0$, $RMSE_1$ is smaller for prediction C than for D and at a smaller mean displacement $\bar{D}$. The relatively small error reduction potential for D is a result of the fact that features not present in the predictions remain absent after the warping procedure, as evident in the deformed predictions in Fig. 6. As a result, $RMSE_1$ remains relatively high for D, rightfully penalizing the prediction for the absence of the observed features. A conclusive answer as to whether C or D is the better prediction now requires a (subjective) weighting of $RMSE_0$, $RMSE_1$ and $\bar{D}$. Conveniently, the weighting between location errors, pre-warp and post-warp intensity errors is already provided by the formulation of $RMSE_w$, allowing a quantitative single-number comparison between predictions C and D. For $D_{max} = 3000$ m, the values for $RMSE_w$ indicate that prediction C outperforms D (Fig. 7), consistent with the ranking based on visual

inspection. Naturally, the occurence of this ranking reversal, as compared to the ranking based on RMSE_0, depends on the chosen value of $D_{\max}$.

4 Conclusions

We have developed a new diagnostic tool for morphodynamic model validation. It employs an image warping method that finds the smooth displacement field between predictions and observations that minimizes the point-wise error. Two new metrics are proposed: (1) a location error $\bar{D}$ that is determined as a weighted mean distance between morphological fields; and (2) a combined error metric RMSE_w that takes both location and intensity errors into account.

A full appreciation of the quality of a prediction can be obtained when considering $\bar{D}$ in concert with both the original point-wise error RMSE_0 and the point-wise error of the deformed predictions, RMSE_1. In order to quantify the relative performance between predictions, a (subjective) weighting of these three metrics must be carried out. Alternatively, the weighting is already provided by RMSE_w that combines all relevant information on location errors and pre- and post-warp intensity errors.

The combined error metric credits predictions to the degree that a larger error reduction can be obtained with smaller displacements. It reduces to RMSE_0 if all displacements are larger than a user-defined $D_{\max}$ and to RMSE_1 for displacements that are negligible relative to $D_{\max}$. The latter can be seen as the maximum distance over which morphological features may be displaced for the prediction to still get (some) credit for predicting these features. The appropriate choice for $D_{\max}$ depends on the prediction situation and the goal of the simulation. Since it only requires a single, physically intuitive parameter, RMSE_w provides a robust basis for comparison.

An example of a schematized tidal inlet has demonstrated that RMSE_w outperforms the conventional validation approach based on a strictly point-wise metric such as RMSE_0. In situations where morphological features are misplaced, point-wise accuracy metrics tend to favour predictions that underestimate variability. For the schematized tidal inlet, it was shown that, as opposed to RMSE_0, the new combined error metric RMSE_w makes choices as to which of two predictions is better, which are consistent with visual validation by experts.

References

Anthes, R. A.: Regional models of the atmosphere in middle latitudes, Monthly weather review, 111, 1306–1335, doi:10.1175/1520-0493(1983)111<1306:RMOTAI>2.0.CO;2, 1983.

Bougeault, P.: The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events, CAS/JSC WGNE Report, 18, WMO/TD-NO.1173, Appendix C, 1–11, http://www.wcrp-climate.org/documents/wgne18rpt.pdf, 2003.

Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E., Brown, B., and Mason, S.: Forecast verification: current status and future directions, Meteorol. Appl., 15, 3–18, doi:10.1002/met.52, 2008.

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E.: Intercomparison of spatial forecast verification methods, Weather Forecast., 24, 1416–1430, doi:10.1175/2009WAF2222269.1, 2009.

Gilleland, E., Lindström, J., and Lindgren, F.: Analyzing the image warp forecast verification method on precipitation fields from the ICP, Weather Forecast., 25, 1249–1262, doi:10.1175/2010WAF2222365.1, 2010.

Haben, S., Ward, J. A., Vukadinovic Greetham, D., Singleton, C., and Grindrod, P.: A new error measure for forecasts of household-level, high resolution electrical energy consumption, Int. J. Forecast., 30, 246–256, doi:10.1016/j.ijforecast.2013.08.002, 2014.

Keil, C. and Craig, G. C.: A displacement and amplitude score employing an optical flow technique, Weather Forecast., 24, 1297–1308, doi:10.1175/2009WAF2222247.1, 2009.

Kroon, D.-J. and Slump, C. H.: MRI modality transformation in demon registration, in: From Nano to Macro, 2009, ISBI'09, IEEE International Symposium on Biomedical Imaging, 963–966, IEEE, doi:10.1109/ISBI.2009.5193214, 2009.

Marzban, C. and Sandgathe, S.: Optical flow for verification, Weather Forecast., 25, 1479–1494, doi:10.1175/2010WAF2222351.1, 2010.

Mass, C. F., Ovens, D., Westrick, K., and Colle, B. A.: Does increasing horizontal resolution produce more skillful forecasts?, B. Am. Meteorol. Soc., 83, 407–430, doi:10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2, 2002.

Pennec, X., Cachier, P., and Ayache, N.: Understanding the "demons algorithm": 3D non-rigid registration by gradient descent, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI'99, 597–605, Springer, doi:10.1007/10704282_64, 1999.

Roelvink, J. A.: Coastal morphodynamic evolution techniques, Coastal Engineering, 53, 277–287, doi:10.1016/j.coastaleng.2005.10.015, 2006.

Sutherland, J., Peet, A., and Soulsby, R.: Evaluating the performance of morphological models, Coastal Engineering, 51, 917–939, doi:10.1016/j.coastaleng.2004.07.015, 2004.

Thirion, J.-P.: Image matching as a diffusion process: an analogy with Maxwell's demons, Medical Image Analysis, 2, 243–260, doi:10.1016/S1361-8415(98)80022-4, 1998.

Ziegeler, S. B., Dykes, J. D., and Shriver, J. F.: Spatial error metrics for oceanographic model verification, J. Atmos. Ocean. Tech., 29, 260–266, doi:10.1175/JTECH-D-11-00109.1, 2012.