A global delta dataset and the environmental variables that predict delta formation

Rebecca L. Caldwell[1,2], Douglas A. Edmonds[1], Sarah Baumgardner[2,3], Chris Paola[3], Samapriya Roy[4], Jaap H. Nienhuis[5]

[1] Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, 47401, USA
[2] Chevron Energy Technology Company, Chevron Corporation, Houston, TX, 77005, USA
[3] Department of Earth Sciences, University of Minnesota, Minneapolis, MN, 55455, USA
[4] Department of Geography, Indiana University, Bloomington, IN, 47401, USA
[5] Department of Physical Geography, Utrecht University, Utrecht, NL

Correspondence to: Rebecca Caldwell rcaldwell93@gmail.com

Abstract. River deltas are sites of sediment accumulation along the coastline that form critical biological habitats, host megacities, and contain significant quantities of hydrocarbons. Despite their importance, we do not know which factors most significantly promote sediment accumulation and dominate delta formation. To investigate this issue, we present a global dataset of 5,399 coastal rivers and data on eight environmental variables. Of these rivers, 40% (n = 2,174 deltas) have geomorphic deltas, defined either by a protrusion from the regional shoreline, a distributary channel network, or both. Globally, coastlines average one delta for every ~300 km of shoreline, but there are hotspots of delta formation, for example in Southeast Asia there is one delta per 100 km of shoreline. Our analysis shows that the likelihood of a river to form a delta increases with increasing water discharge, sediment discharge, and drainage basin area. On the other hand, delta likelihood decreases with increasing wave height and tidal range. Delta likelihood has a non-monotonic relationship with receiving basin slope: it decreases with steeper slopes but increases for slopes > 0.006. This reflects different controls on delta formation on active versus passive margins. Sediment concentration and recent sea-level change do not affect delta likelihood. A logistic regression shows that water discharge, sediment discharge, wave height, and tidal range are most important for delta formation. The logistic regression correctly predicts delta formation 75% of the time. Our global analysis illustrates that delta formation and morphology represent a balance between constructive and destructive forces, and this framework may help predict tipping points where deltas rapidly shift morphologies.

1. Introduction

Deltas provide a variety of ecosystem services, such as carbon sequestration and nitrate removal (Rovai et al., 2018; Twilley et al., 2018), and they provide home to close to half a billion people (Syvitski and Saito, 2007) living within large

agricultural and urban centers (Woodroffe et al., 2006). Deltas form at river mouths where fluvial sediment accumulates nearshore long enough for the deposit to become subaerial. This simple view of delta formation is a statement of sediment mass balance and understanding where deltas form requires knowing how and why sediment accumulates. Sediment accumulates provided it is supplied and deposited at the coast faster than it is removed. Sediment supply and removal are chiefly determined by the river, waves, tides, rate of relative sea-level change, and offshore bathymetry. To complicate matters, most of these variables can be both sources and sinks, and their exact roles in the deltaic sediment mass balance remains uncertain. Previous research suggests that rivers are almost always sources (Bates, 1953; Coleman, 1976; Wright, 1977; Syvitski et al., 2005; Syvitski and Saito, 2007), whereas the roles of waves and tides are ambiguous (Nienhuis et al., 2015; Hoitink et al., 2017; Lentsch et al., 2018). The conditions that lead to delta formation are not completely known, but we know those conditions are not easily met—pick nearly any oceanic shoreline on earth and there will be several river mouths that intersect the coast, but only some of these rivers will have a delta. Previous studies on delta formation (Wright et al., 1974; Audley-Charles et al., 1977; Milliman and Syvitski, 1992; Syvitski and Saito, 2007; Nyberg and Howell, 2016) focused on large-scale patterns and concluded that major modern delta locations are influenced largely by tectonic margin type and drainage patterns. While useful, these datasets were biased towards the largest and most populated deltas. Expanding the prediction effort to deltas of all sizes is a logical next step, especially because smaller deltas are thought to be more resilient to rising sea levels (Giosan et al., 2014).

In addition to expanding the range of delta sizes, to understand the controls on delta formation we need to consider cases where delta formation is suppressed. In this paper we present a global delta dataset and use it to investigate why some rivers form deltas and others do not. Understanding conditions for modern delta formation should also help exploration for ancient deltaic deposits, which requires predicting where deltas might form under past environmental conditions (Nyberg and Howell, 2016). Similarly, as research moves towards delta risk assessment due to global environmental change (Tessler et al., 2015) and improving efforts to build new deltaic land (Kim et al., 2009), we must understand how different environmental variables govern delta formation. For example, understanding the conditions for delta formation would help restoration efforts that seek to build new deltaic land in places like the Mississippi River Delta (Paola et al., 2011; Edmonds, 2012; Twilley et al., 2016).

To achieve these goals, we developed a global dataset that includes the locations of 5,399 coastal rivers, information on whether they form deltas or not, and the related environmental variables important for delta formation. We use global datasets of coastlines (Dürr et al., 2011; Nyberg and Howell, 2016), sediment and water (Syvitski and Milliman, 2007; Milliman and Farnsworth, 2011), wave climate hindcasts (Tolman, 2009; Chawla et al., 2013), a tidal inversion model (Egbert and Erofeeva, 2002), ocean bathymetry data (Amante and Eakins, 2009), and rate of sea-level change (https://www.aviso.altimetry.fr). Of the 5,399 included rivers, 2,174 form geomorphic deltas that are visible in aerial imagery, defined either by a protrusion from the regional shoreline, a distributary channel network, or both. We use statistical relationships between independent environmental variables and the presence or absence of a delta to determine what controls the likelihood of a river to form a delta.

2. A global coastal and river delta dataset

2.1 Identifying river deltas

River deltas are fundamentally systems of sediment accumulation and distribution at the coastline. Accordingly, we identify coastal deltas by distinguishing geomorphic expressions of sediment accumulation and distribution at locations where rivers meet the coast. We consider a river to have formed a delta at the coastline if the river-mouth area contains an active or relict distributary network (Fig. 1e), ends in a subaerial depositional protrusion from the lateral shoreline (Fig. 1d), or does both (Fig. 1c). Distributary networks are an expression of sediment deposition and distribution (Edmonds et al., 2011) and we identify them by the presence of one or more channels that bifurcate and intersect the coast at different locations. We include relict channels, where they are clearly visible in imagery and connect to the main channel, because they are evidence of sediment distribution and accumulation through avulsion (Slingerland and Smith, 2004). We do not include channels that bifurcate solely around non-deltaic topographic highs. Our second criterion is oceanward-directed shoreline protrusions. We classify a protrusion as deltaic if it has a relatively smooth depositional shoreline, as opposed to rough shorelines associated with rocky coasts (Limber et al., 2014), and if it extends more than ~5 channel widths oceanward relative to the position of the regional shoreline. We only map protrusions that are associated with the river, ignoring protrusions that may exist near the channel mouth that we judge to be pre-existing undulations in the shoreline. Examples of this include promontories associated

with preexisting geology or depositional protrusions created by other processes, such as wave-driven sediment transport (Ashton et al., 2001).

Our delta identification method does not account for deltaic deposition with no geomorphic signature, such as a single-channel delta infilling a drowned valley that produces no protrusion from the regional shoreline. Although such features may be considered deltaic, we cannot unambiguously identify them as deltas based on aerial imagery alone and we do not include them in the dataset.

We applied the preceding criteria to a scan of oceanic coastlines using Google Earth. First, we identified all rivers reaching the coast that are connected to an upstream catchment (Figs. 1a, 1c–1e). Channels not clearly connected to an upstream catchment, such as tidal channels, were not included in the dataset (Fig. 1b). This was done to restrict the study to coastal depositional landforms that represent the interaction of upstream and downstream environmental variables. We selected rivers at least 50 m in width because they have corresponding data, such as basin area, that can be reliably determined on coarser resolution elevation models. This width designation was applied to the rivers' bankfull widths, and thus includes any visible mid-channel bars. Channel widths on rivers without a delta were measured at the shoreline or upstream from visible marine influence such as significant tidal widening (Nienhuis et al., 2018). If a river empties into a gradually widening estuary or embayment, we measured the channel width where it is representative of the river devoid of significant downstream widening. Channel widths on rivers that have deltas were measured immediately upstream of the delta node, which we define as the location of the most upstream bifurcation, or if no bifurcation occurs, we use the intersection of the main channel with the regional shoreline (e.g., Fig. 1c and 1d, blue dot). In all cases, channel widths were not measured in areas of clear human influence. This includes, for example, man-made levees that can cause artificial widening or narrowing of channels.

We mapped rivers and deltas on the coastlines of Earth's continents and large islands (Fig. 2). We exclude small islands where rivers large enough for inclusion are rare and it is difficult to obtain environmental data. Thus large islands, such as Papua New Guinea and Fiji, were included but not all the associated smaller islands. Coastlines dominated by fjords (as determined using *Dürr et al. (2011)*) were not included because offshore glacial over deepening and protection from coastal waves and tides make their comparison to most of the world's coastal deltas difficult. Ephemeral rivers in arid regions were included in the dataset, though the rivers in these regions are often difficult to identify due to poor imagery and difficulty

distinguishing the channel banks when they are dry. If a clear distinction was not possible, the river was not included in the dataset. Thus, the total count of rivers and deltas in arid regions should be considered a minimum. Finally, we did not include river channels that do not clearly reach the coast to avoid conflating alluvial fans with deltas.

For each river we marked the latitude and longitude of the main river mouth (Figure 1, *RM*) (Supplemental Table 1). For rivers without a delta, this is the location where the river meets the coastline (Fig. 1a), and for rivers with deltas, this is the location of the widest river mouth in the distributary network (Fig. 1c–1e). For rivers sheltered by barrier islands or rocky islands, we mark the river mouth landward of those obstructions. How do you do this for the Lena delta?

2.2. Environmental variables

To determine controls on delta formation we also compiled data on eight environmental variables (Table 1). We classify the environmental variables into two groups: (1) upstream variables include water and sediment supply from the river, sediment concentration, and the drainage basin area; and (2) downstream variables include wave heights, tidal ranges, bathymetric slopes immediately offshore of the river mouth, and the rate of sea-level change.

Notably absent in the collected environmental variables are tectonic data. At present, there are no globally available measurements of tectonic activity (e.g., uplift). However, we consider some of the variables to be reasonable proxies for tectonics. For instance, models predicting sediment flux to the ocean represent tectonics in the form of basin area (Syvitski and Morehead, 1999; Syvitski and Milliman, 2007). We also include bathymetric slope, which is a rough proxy for tectonics because, on average, tectonically active margins have steeper slopes than passive margins (Pratson et al., 2007).

2.2.1. Upstream variables

We compiled the four upstream variables from the global river dataset of Milliman and Farnsworth (2011) (hereafter referred to as MF2011). We matched rivers in this dataset with entries in MF2011 based on geographic proximity or by the river name. If neither matching method yielded a confident result, the MF2011 data were not included in this study. If two or more rivers in the MF2011 dataset combine to make one river in this study's dataset, the data from all relevant MF2011 rivers are included. In cases where matches were found, we included the river ID(s) from MF2011 in our dataset (Supplemental Table 1). Our dataset includes 1,217 MF2011 rivers, representing 1,158 entries in our dataset (54 entries are made from 2 or

more MF2011 rivers). There are 314 MF2011 rivers not included in this dataset because they are too small (< 50 m wide), exist on coastlines not included in our dataset, or could not be matched.

Water discharge (Q_w, expressed as mean annual volumetric flux, m^3s^{-1}) data come from the MF2011 dataset. The Q_w measurements are compiled from various sources of reported gauging station measurements, where the downstream-most gauging station data is used. As MF2011 note, water discharge values may be over- or under-estimated due to distance upstream of the river mouth. In many regions, additional water input downstream of the gauging station increases the true Q_w value reaching the river mouth. However, in arid regions, water volume may be lost due to evapotranspiration, groundwater recharge, or irrigation water removal. In total, 17% of rivers (n = 943) in this dataset have Q_w data.

Sediment discharge (Q_s, expressed as mean annual volumetric flux, m^3s^{-1}) data come from the MF2011 dataset of annual sediment load measurements and are converted to m^3s^{-1} assuming a density of 2650 kg m^{-3}. The Q_s data are compiled from various sources of reported loads and most often represent suspended load measurements rather than total load. Bedload is assumed to represent only 10% of total load (Milliman and Meade, 1983), but this estimation may be less valid for small mountainous rivers where relative proportion of bedload can be greater (Amante and Eakins, 2009). Like the Q_w data, many of these measurements may have been made upstream of the actual river mouth, and thus actual Q_s values that reach the river mouth likely vary (e.g., due to fluvial plain deposition downstream of measurement location). Finally, extrapolation of measurements taken over varying lengths of time to represent annual sediment loads is potentially risky (e.g., when considering the significance of event-driven discharge events). In total, 11% (n = 600) of all rivers in this dataset have Q_s data. Sediment concentration (Q_s/Q_w) is calculated from the sediment and water discharge data and 11% (n = 571) of all rivers have Q_s/Q_w data.

We also include upstream drainage basin area (A_b, km^2) in our dataset because it partly sets the magnitude of Q_w and Q_s (Syvitski et al., 2003; Syvitski and Milliman, 2007) and compensates for the relatively small number of rivers with water and sediment data. A_b data come from the MF2011 dataset. Although these values are often well documented for larger river systems, they may sometimes represent the total drainage area upstream of a hydrologic station, which would be a smaller value than total drainage area upstream of the river mouth. Given the potential error, A_b values should be considered a minimum.

2.2.2. Downstream variables

Four downstream variables are included in this dataset. Annual significant wave heights (H_w, m) were calculated using the NOAA WAVEWATCH III 30-year Hindcast Phase 2 for 1979–2009 (Tolman, 2009; Chawla et al., 2013). The model outputs 30 years of hourly significant wave height data on five different ocean grids with varying resolution, and the final product is interpolated to a global 0.5-decimal degree grid. We ran a nearest-neighbor search from each *RM* location to the nearest grid cell with wave data that is within one grid cell diagonally, which is equivalent to 0.7071 decimal degrees, or ~80 km at the equator. Because some coasts are missing wave data not all 5,399 rivers have corresponding wave data. For each calendar year, we calculate the annual mean of the top 1/3 largest wave heights. The resulting 30 years of annual significant (top 1/3 largest) wave height data are representative of the strongest wave action that occurs at each location within a year, or representative of a stormy season for areas with strongly seasonal wave climates. The mean of these 30 annual values is the mean annual significant wave height (H_w).

Median tidal ranges (H_t, m) were calculated using the previously published Oregon State University TOPEX/Poseidon Global Inverse Solution TPXO model results (Egbert and Erofeeva, 2002). The model outputs tidal harmonics component data on a 0.25-decimal degree resolution grid derived from a barotropic inverse solution. Following Baumgardner (2015), we use the main tidal components, the lunar semidiurnal and the lunar diurnal, to calculate mean tidal range by building a composite tidal sine wave and calculating the average range. We ran a nearest neighbor search from each *RM* location to all grid cells with tidal data that are within the same distance used for the wave search. The median of the tidal range values within this search radius is used to represent each river mouth's tidal range.

Receiving-basin bathymetry is an important attribute of delta formation because it sets the size and shape of the volume to be filled from a mass balance perspective. The size of the basin could be characterized by the average depth whereas the shape is most simply characterized by the bathymetric slope. In most cases, we do not know basin depth prior to delta formation, and current depths offshore deltaic river mouths will be deeper than the initial depths if the basin has offshore-dipping bathymetric slopes. Thus, instead of using depth, we characterize the receiving basin with bathymetric slopes. Bathymetric slopes (S_b) are calculated from ETOPO1 bathymetric data (Amante and Eakins, 2009) and *RM* locations. ETOPO1 is a global surface elevation model with 1 arc-minute resolution (1/60 decimal degree, or ~1,800 m at the equator). For each

river, we collect all bathymetric elevations within a 20-km radius from the *RM* location. We calculate linear slopes between each point and the *RM* (assumed elevation = 0 m), and take S_b as the 75th percentile of all slopes. We purposefully search far away from the shoreline because we want to characterize the offshore depths not affected by sediment deposition from the river. Is there some scaling length?

Rate of sea-level change is calculated from AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data, https://www.aviso.altimetry.fr). The AVISO dataset combines sea-level change from different satellite altimetry missions from 1992-2018 using the delayed time Ssalto/Duacs multi-mission altimeter data processing system, which corrects biases among instruments and applies inter-calibration to the record. Rates of sea-level change are calculated for every 0.25° x 0.25° cell by finding the best fit to the data over 26 years. The data we use are not corrected for glacial isostatic adjustments. These rates are decidedly modern and that makes it difficult to compare with deltas, many of which formed 1000s of years ago as sea-level rise started slowing following deglaciation (Stanley and Warne, 1994). It would be ideal to compare delta formation to sea-level change data averaged over their lifespans, but those data do not exist.

3. Results

Our mapping reveals there are 5,399 coastal rivers with widths greater than 50 m, and 2,174 of those rivers (~40%) have a geomorphic delta. Herein, we refer to all 5,399 coastal rivers as "rivers", the 3,225 that do not have deltas as "river mouths," and the 2,174 with deltas as "deltas." These terms are not completely accurate because, for example, an individual "river" that is considered a "delta" rather than a "river mouth" still has at least one main river mouth (*RM*) and may have additional river mouths for each distributary channel.

3.1. Global distribution of rivers and deltas

River deltas are not distributed evenly on coastlines and there are locations on the world's coastlines where deltas are unusually common (Fig. 2). These "delta hotspots" occur primarily in Southeast Asia (dashed box Fig. 2b). Notably, these areas are also densely populated with rivers (Fig. 2a), though river abundance does not always equate to delta abundance. For example, East Asia has high river density but low delta density (black box, Fig. 2b). Similarly, along the west coasts of Central and southern North America (from 5°N to 45°N) the coast is densely populated with rivers, but the northern portion is delta-

poor compared to the southern portion. There are also a surprising number of deltas in arid environments. For instance, there is high delta density in the Red Sea and on Baja California. This arises because the alluvial fans coming off the mountains reach the coastline and satisfy our definition of a delta.

Binning these data by latitude reveals preferential locations of rivers and deltas (Fig. 3). The largest numbers of rivers and deltas occur roughly from -12° to 45°, and 66° to 72° (Fig. 3a). This unequal distribution is partly explained by the unequal latitudinal distribution of global shoreline length (Wessel and Smith, 1996) (Fig. 3b). River density, or rivers per shoreline kilometer, shows that globally there is one river for every 230 km of coastline and one delta for every 333 km of coastline. Coastlines within the -6° to -3° bin have the highest density of deltas with roughly one delta per 100 km of shoreline (Figure 3c). River density is above average from -45° to 45° (Fig. 3c, white bars). Delta density, however, is above average over a smaller range from -21° to 30° (Fig. 3c, solid black line).

To determine which environments promote delta formation, it is perhaps most instructive to observe locations where the likelihood for rivers to create deltas is highest. Delta likelihood (L_d) is defined as the number of deltas relative to the total number of rivers for a given set of samples (Fig. 3d, solid black line). For the entire dataset 40% of rivers form deltas, and thus the global L_d is 0.40 (Fig. 3d, dashed black line). Regions where L_d is higher than the global mean exist from -27° to 30° and 60° to 72°, whereas rivers located from -57° to -25° and 30° to 60° are least likely to form a delta (Fig. 3d).

These latitudinal zones where rivers are more likely to create deltas coincide with peaks in environmental variables that influence delta formation. Both Q_w and Q_s have notable peaks from -9° to 30° and 60° to 75° (Fig 4a, b), which are similar in location to L_d peaks. A_b has the high latitude peak, but is missing the equatorial peak (Fig. 4d) probably reflecting the importance of small mountainous rivers in those locations (Milliman and Syvitski, 1992). On the other hand, delta formation is infrequent where H_w and H_t are high, namely -57° to -27° and 42° to 60° (Fig. 4e, f). There are no latitudinal changes in Q_s/Q_w, S_b, or H_s that are easily relatable to delta formation (Fig. 4c, g, h).

3.2. Relationships between environmental variables and delta formation

We explore controls on delta formation by analyzing how the likelihood of a river creating a delta varies with each environmental variable. River mouths and deltas have statistically different population distributions for seven of the eight

environmental variables (all but Q_s/Q_w) (Table 2), suggesting that deltas form under certain ranges of environmental variables. To determine this, we used the Kolmogorov-Smirnov test, which is a non-parametric, distribution-free test that uses the cumulative distribution functions of the two populations to estimate statistical difference. Although a few variable pairs show some correlation, such as Q_w and A_b, none have a strong statistical correlation (Pearson correlation coefficient > 0.9), suggesting they exert largely independent controls on delta formation.

Delta likelihood (L_d) generally increases as the upstream environmental variables increase (Fig. 5). Increasing Q_w, Q_s, and A_b causes a linear increase in semi-log space in L_d (Figs. 5a-b, d). Deltas have characteristic Q_w, Q_s, and A_b values that are an order of magnitude larger than those of river mouths (statistically significant, $p < 0.05$) (Table 2). These data suggest that rivers with small water and sediment discharge and/or that come from small drainage basins rarely form deltas, whereas rivers with larger values of the upstream variables frequently create deltas. Sediment concentration (Q_s/Q_w) exerts no clear control on L_d (Fig. 5c), and there is no statistical difference between the mean and median Q_s/Q_w values for rivers mouths versus those for deltas (Table 2).

Rivers are less likely to create deltas where H_w and H_t are large. L_d shows a clear linear decrease as H_w increases (Fig. 6a). Rivers that experience little wave energy at the coast ($H_w < 1$ m) create a delta more than half of the time ($L_d \approx 0.5$-0.6), but delta formation becomes nearly impossible for larger wave heights. L_d also seems to show a linear decrease with H_t (Fig. 6b), but this relationship shows significantly more scatter than that with H_w. If the long tail of the distribution is eliminated, where the sample size is small ($H_t > 8$ m), the relationship is clearer. The population of river mouths has higher mean and median H_w and H_t than rivers with deltas (statistically significant, p-value < 0.05) (Table 2).

S_b displays a non-monotonic relationship where L_d decreases then increases across the range (Fig. 6c). S_b data are bimodally distributed for the rivers in our dataset, suggesting rivers empty into two types of receiving basins (separated by the dashed gray line). Based on visual observation, the shallowly-dipping basin types reflect passive margins, and the steeply-dipping basins active margins, though we did not pursue a more robust confirmation. If these basin types are separated, there is a clearer relationship between S_b and L_d. For shallowly-dipping basins ($S_b < 0.006$), there is a negative relationship between L_d and S_b (Fig. 6c, left of dashed gray line), and delta likelihood increases as slope decreases. In steeply-dipping basins ($S_b > 0.006$), L_d is approximately constant to slightly increasing as slopes steepen (Fig. 6c). There is no clear relationship between

sea-level change (H_s) and L_d (Fig. 6d), which is somewhat surprising given that river mouths and deltas have statistically different mean and median H_s values (Table 2).

To quantify the relative importance of the environmental variables for delta formation, we develop an empirically-derived logistic regression. The result of a logistic regression is a statistical model that predicts a dichotomous outcome (in this case, a river creates a delta, or it does not) based on multiple independent variables. This dataset contains 8 total independent variables collected on all rivers, where four are upstream variables (Q_w, Q_s, Q_s:Q_w, A_b), and four are downstream variables (H_w, H_t, H_d, S_b). Of the 5,399 rivers in this dataset, 490 of them (9.1%) have data available for all eight independent variables.

The data meet the assumptions of binary logistic regression because the dependent variable has two mutually exclusive outcomes, and the sample size is large (45 samples or more per independent variable). Additional assumptions that the data must meet include having little to no multicollinearity and no outliers. We tested for multicollinearity by calculating the Pearson correlation coefficients (R) between all continuous independent variables and no variables exhibited $R > 0.9$. We also remove 14 rivers that have outliers in any of the independent variables based on a modified z-score, where an absolute value modified z-score > 3.5 is considered an outlier (Iglewicz and Hoaglin, 1993). The final subset of data used for the regression has n = 476 rivers (248 rivers without deltas, 228 rivers with deltas). The samples were randomly separated into training (2/3 of the samples) and validation (1/3 of the samples) subsets, each of which represented similar distributions of independent and dependent variables. We do this to see how well the logistic regression can predict delta formation on river mouths not used in the original equation.

The binary logistic regression between the probability that a river will create a delta and the eight environmental variables yields the following log odds relationship:

$$\ln(\pi_{delta}/(1-\pi_{delta})) = 1.38 + 0.000524Q_w + .4.77Q_s - 0.952H_w - 0.175H_t \qquad (1)$$

where π_{delta} is the probability that a river will form a delta, and ranges from 0 (river is unlikely to form a delta) to 1 (river is most likely to form a delta). This is different from the L_d values presented earlier only because it is predicted whereas L_d was measured. Environmental variables with $p > 0.05$ (Q_s/Q_w, A_b, S_b, and H_s) are not included in the final empirical relationship,

because any control these variables exert on delta formation is minimal (e.g., variations in Q_s/Q_w have no clear relationship with L_d, Fig. 5d) or related to variations in the other important variables (e.g., A_b influences Q_w and Q_s).

Thus, the combination of environmental variables that comprises the right side of equation (1) predicts the log odds that a river will form a delta. When tested using the validation subset, equation (1) has a 75% success rate at predicting delta presence (Fig. 7), where $\pi_{delta} > 0.5$ is considered a prediction that a delta exists, and $\pi_{delta} < 0.5$ is considered a prediction that no delta exists.

This empirically-derived relationship can be used to calculate the probability that a certain combination of the most important environmental variables will form a delta. For example, using environmental variable values for the Godavari River in the right-hand side (*RHS*) of equation (1) results in $RHS = 3.93$. The probability that the Godavari River should form a delta is $\pi_{delta} = \frac{e^{RHS}}{1+e^{RHS}} = 0.98$. Thus, the environmental variables that conspire to form the Godavari River are very likely to form a delta, which is not surprising given the existence of the large Godavari River delta.

4. Discussion

4.1. Which environmental variables most strongly control delta formation?

We have considered eight environmental variables and determining which ones matter the most for delta formation is not straightforward. After all, most combinations of environmental variables that exist globally completely suppress delta formation (60% of the rivers included in this dataset do not have a delta). Our likelihood analysis shows that deltas are more likely to form at river mouths with large water discharge Q_w (Fig. 5a), sediment discharge Q_s (Fig. 5b), and drainage basin area (Fig. 5c), and with small significant wave heights H_w (Fig. 6a), and tidal ranges H_t (Fig. 6b). Increasing upstream variables (Q_w, Q_s, A_b) across their value range accounts for the full range of delta likelihood—that is, the smallest Q_w, Q_s, and A_b values have $L_d \approx 0$, and largest Q_w, Q_s, and A_b largest have $L_d \approx 1$ (Figure 5). In contrast, increasing the downstream variables (H_w, H_t) decreases the likelihood that a river forms a delta, but does not produce the full range of possible L_d values. At the lowest values of H_w and H_t delta likelihood is still 0.5. The relationship with H_w is more significant, it has a steeper slope and less scatter compared to H_t. In fact, downstream variables seem to be of secondary importance for forming deltas. When we remove H_w and H_t from equation (1) the prediction success rate decreases by only 3%, from 75% to 72%.

These controls on delta formation explain first-order latitudinal variations observed in Figures 3 and 4. For example, the peaks in water and sediment discharge values from -9° to 30° and 60° to 75° (Fig. 4) likely explain the similarly located peaks in delta formation (Fig. 3). The suppressing effects of waves and tides can also be seen at a global scale. Low delta formation rates from -57° to -27° and 30° to 60° are likely due to large H_w and H_t values in these regions, where Q_w and Q_s are low (Figs. 3, 4). Moreover, the zone from 60° to 75° that has increased Q_w and Q_s values (Fig. 3) also has some of the lowest H_w and H_t values (Fig. 4). Thus, while high Q_w and Q_s values in this region promote delta formation, the decreased H_w and H_t values also allow delta formation to occur.

Downstream bathymetric slope (S_b) displays a complex relationship with delta likelihood. At slopes < 0.006, delta likelihood decreases with increasing slope (Fig. 6c), because all else being equal, deeper areas should take longer to fill with sediment and they are also less effective at damping incoming waves and tides. But, interestingly at slopes > 0.006, delta likelihood increases with steeper slopes, which is more difficult to explain. If these steeper margins relate to active margins, then larger sediment sizes and higher supply on active margins may explain this difference (Audley-Charles et al., 1977; Orton and Reading, 1993; Milliman and Farnsworth, 2011). After all, the supply of coarser sediment to the coast is more easily retained nearshore (Caldwell and Edmonds, 2014), thereby increasing the likelihood of delta formation.

4.2. The roles of rivers, waves, and tides in delta formation

Our data suggest that deltas are fundamentally created by water and sediment discharge, whereas waves, and possibly tides, suppress delta formation. This perspective stands in contrast to existing thoughts on delta formation. The *Galloway* (1975) diagram is the foundational study on delta morphology and formation. Galloway's diagram implies that deltaic formation and morphology is the result of the interplay of river, waves, and tides. But, Galloway's diagram remains largely qualitative and it is not clear how the forces of rivers, waves, and tides are quantified, nor it is clear what kinds of predictions the diagram makes. In fact, our data offer a different view of deltaic formation than the one proposed by Galloway. Our data suggest that delta formation is the result of constructive upstream forces set by the river, and destructive downstream marine forces. Consider the case of a purely wave dominated delta. Galloway's diagram would predict a cuspate delta, but our data clearly show that the most wave-dominated delta is no delta at all, consistent with the work of *Nienhuis et al.*, (2013). This

suggests to us that the concept of delta formation and morphology might be better cast as a balance between constructive and destructive forces.

From this perspective new questions emerge: How do wave and tidal processes change the ability of fluvial processes to construct deltas? How stable is the balance between a given set of constructive and destructive forces? With regard to the last question, there are examples of rapid changes in delta morphology through time, which suggests that the balance can be precarious. The Rhône River clearly shifted in morphology from channel network dominated in the 16th century to its more familiar wave-smoothed shape today, as floods and sediment loads declined during the Little Ice Age (14th-19th centuries) (Provansal et al., 2015). The Po River delta in Italy showed three morphological transitions each time the balance between river and waves changed over the last 4000 years (Anthony et al., 2014). Future work would benefit from linking our empirically derived delta likelihood predictor with metrics of delta morphology to understand when morphological shifts might occur.

5. Implications

River deltas are the final filters of water and sediment before they are discharged to the global ocean (Sawyer et al., 2015). As we have shown here, certain environmental variables promote sediment accumulation and delta formation. This accumulation results in the storage of sediment, yet all existing efforts to calculate sediment flux to the global ocean ignore sediment deposited in deltas (Milliman and Farnsworth, 2011). In an analogy with blue carbon, we define the volume of sediment deposited on the coastline, in deltas, or just offshore, as "blue sediment." Our results suggest that the amount of blue sediment stored in river deltas at yearly to millennial timescales could be significant. Based on our results, we find that 5.9 Bt/yr, or 85%, of the measured global sediment flux (Milliman and Farnsworth, 2011), moves through a river delta before being discharged into the ocean. This is important because deltas are exceptionally good at impounding sediment because their extensive channel networks self-organize to evenly cover the topset, so that during flood all areas are nourished with sediment (Edmonds et al., 2011; Tejedor et al., 2016; Tejedor et al., 2017). Limited calculations suggest deltas retain 30% of the sediment supplied (Goodbred and Kuehl, 2000; Syvitski and Saito, 2007; Kim et al., 2009), in which case deltas may be an important, and presently unaccounted for, sink in the global sediment cycle.

what does ancient record tell us?

We also think our data has important implications for resource exploration and coastal restoration. Although using equation (1) to predict delta formation for modern rivers is somewhat redundant, it may prove useful for predicting past or future delta existence. Ancient deltaic deposits comprise significant hydrocarbon reservoirs, and provided this analysis holds through geologic time, equation (1) could predict the presence of deltaic deposits in the rock record if Q_w, Q_s, or A_b can be estimated via other geologic methods.

Looking forward, this relationship can be used to predict future deltaic formation. Global environmental change will continue to put coastal environments at risk, largely by land loss due to accelerated sea-level rise and decreased sediment delivery to the coast. Coastal restoration and hazard-mitigation techniques often involve the creation of new deltaic land via controlled river diversions (e.g., Kim et al. (2009)), though it can be difficult to predict the risk related to such projects. Predictions made using equation (1) can help the decision-making process concerning setting controllable environmental variables, such as water discharge. For example, in a hypothetical environment where a river diversion is being considered, and the current set of environmental variables yields RHS = -0.2005 (which suggests the probability of delta formation is π_{delta} = 0.45), a 600 $m^3 s^{-1}$ increase in Q_w alone will increase the probability of delta formation 8% (from 0.45 to 0.53) (assuming the increased Q_w has no effect on other variables).

6. Conclusions

Based on analysis of a new data set comprising 5,399 coastal rivers that are 50 m wide, along with eight environmental variables, we find that only 40% (2,174) of coastal rivers have deltas, and these are unevenly distributed geographically, with delta formation being more likely in latitudes -27° to 30° and 60° to 72°. Likelihood of delta formation increases with increasing sediment flux, water discharge, and basin area, whereas likelihood decreases with increasing tidal range and significant wave height. Receiving-basin bathymetry has a bimodal effect on likelihood of delta formation. At slopes less than 0.006 delta formation decreases with increasing slope, but the trend is reversed at slopes greater than 0.006. Recent sea-level change and sediment concentration have no clear effect on delta formation. Finally, we derive a logistic regression that predicts probability of delta formation with an accuracy of 75%. Together our results suggest that delta formation is a balance between constructive forces, such as water and sediment, and destructive forces, such as waves and tides.

7. Acknowledgements

DAE would like to acknowledge funding and support from National Science Foundation grants 1135427, 1426997, and 1812019. Sea-level products were processed by SSALTO/DUACS and distributed by AVISO+ (https://www.aviso.altimetry.fr) with support from CNES. We thank M. Domaracki, I. Thomas, K. Rhodes, A. Whaling and S. Adams for helping with data collection and organization.

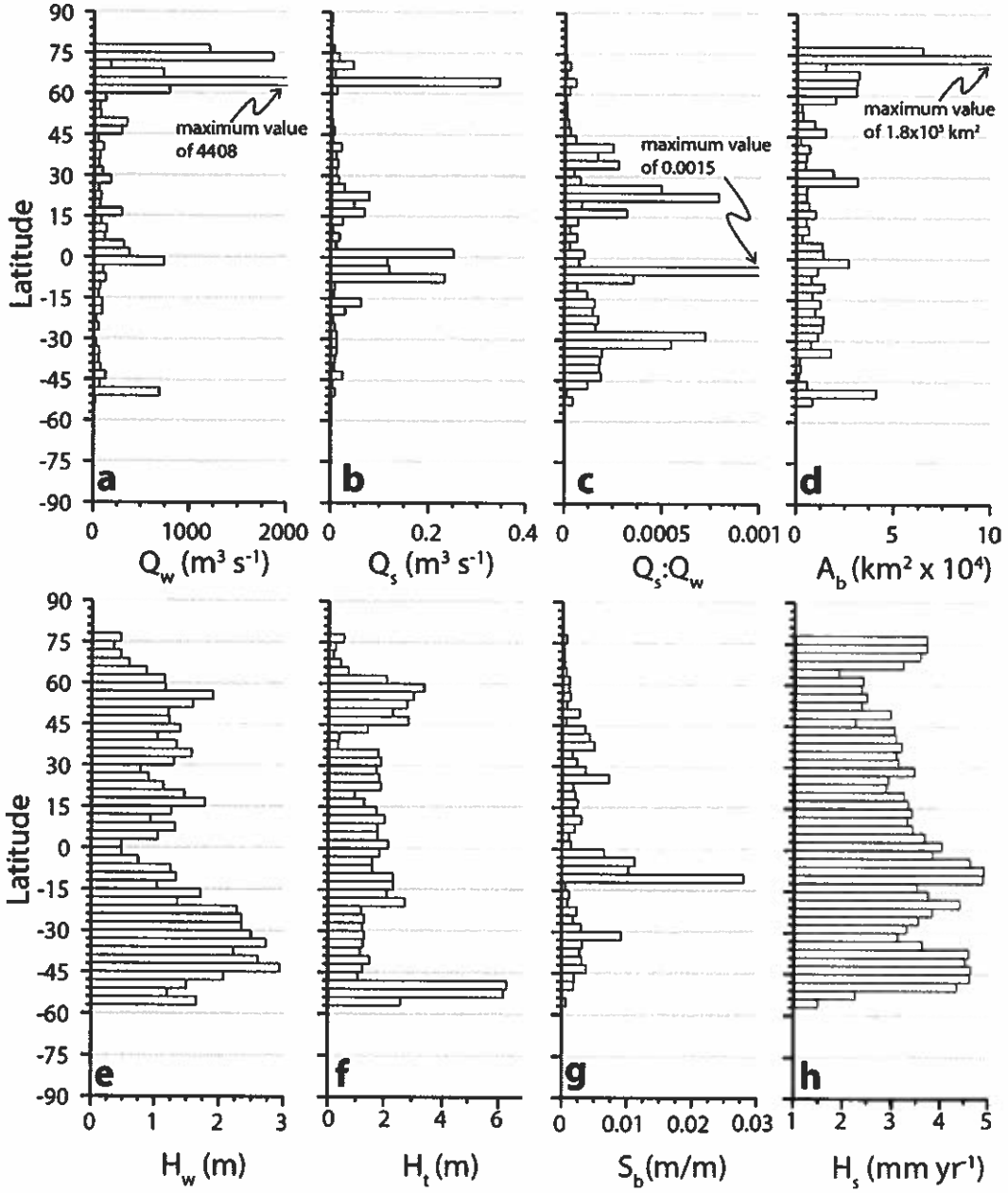
Tables

Table 1. Independent variables: upstream and downstream environmental variables
[1]Sources: 1. *Milliman and Farnsworth* [2011], 2. WaveWatch, 3. OSU TOPEX/Poseidon Global Inverse Solution TPXO, 4. ETOPO1. 5. AVISO (https://www.aviso.altimetry.fr)

UPSTREAM VARIABLES			
Quantity	**Notation (units)**	**# of Rivers with Available Data**	**Source[1]**
Water discharge	Q_w ($m^3 s^{-1}$)	943	1
Sediment discharge	Q_s ($m^3 s^{-1}$)	600	1
Sediment concentration	Q_s:Q_w (–)	571	1
Drainage basin area	A_b (km^2)	1,143	1
DOWNSTREAM VARIABLES			
Quantity	**Notation**	**# of Rivers with Available Data**	**Source[1]**
Wave height	H_w (m)	5,209	2
Tidal range	H_t (m)	5,259	3
Bathymetric slope	$S_{b,r}$ (–)	5,358	4
Sea level change rate	H_s (m)	5,172	5

Table 2: Statistical differences between rivers with no deltas and rivers with deltas. Percentages are calculated relative to the total number of rivers with no deltas (3,225) and with deltas (2,174). *D* is the two-sample Kolmogorov-Smirnov test statistic and is equal to the maximum variance between the cumulative distribution functions of the two populations tested. *p* is the *p*-value at the 5% significance level. *h* is the test decision, where 1 rejects the null hypothesis that the distributions are from the same population and 0 accepts the null hypothesis.

	All Data				*River Mouths (no deltas)*				*Deltas*				*KS test results*		
Variable	**min**	**max**	**mean**	**median**	**n**	**(%)**	**mean**	**median**	**n**	**(%)**	**mean**	**median**	**h**	**D**	**p**
Q_w	0.03	2.1×10^{5}	880	95.13	487	15	171	57.08	456	21	1636	178	1	0.31	5.7×10^{-21}
Q_s	4.8×10^{-5}	14.37	0.13	0.012	312	10	0.024	0.006	288	13	0.25	0.03	1	0.34	1.4×10^{-15}
$Q_s:Q_w$	9.8×10^{-8}	0.078	6.7×10^{-4}	8.5×10^{-5}	293	9	8.9×10^{-3}	9.6×10^{-5}	278	13	0.0004	$8.\times10^{-5}$	0	0.09	0.17
A_b	180	6.3×10^{6}	6.9×10^{4}	8011	587	15	2.3×10^{4}	5100	556	26	1.19×10^{5}	1.3×10^{4}	1	0.25	7.9×10^{-16}
H_w	0.067	5.18	1.33	1.16	3118	97	1.53	1.40	2091	96	1.03	0.87	1	0.28	8.5×10^{-87}
H_t	0.003	15.44	1.74	1.57	3152	98	1.79	1.57	2107	97	1.67	1.57	1	0.06	3.6×10^{-4}
S_b	5.0×10^{-5}	0.25	0.011	0.0021	3211	99	0.011	0.0023	2147	99	0.012	0.0017	1	0.08	5.5×10^{-8}
H_s	-12.45	11.86	3.41	3.36	3121	97	3.35	3.28	2051	94	3.50	3.47	1	0.10	1.4×10^{-11}



why pick this vs. median mean?

Figure 4. Latitudinal variation of the independent variables used in this study. All panels show the median value for 3° bins. (a) water discharge, Q_w; (b) sediment discharge, Q_s; (c) sediment concentration, Q_s/Q_w; (d) drainage basin area, A_b; (e) mean annual significant wave height, H_w; (f) median tidal range, H_t; (g) bathymetric slope, S_b; (h) rate of sea-level change, H_s. For a, c, and d the outliers have been cut off for viewing purposes.

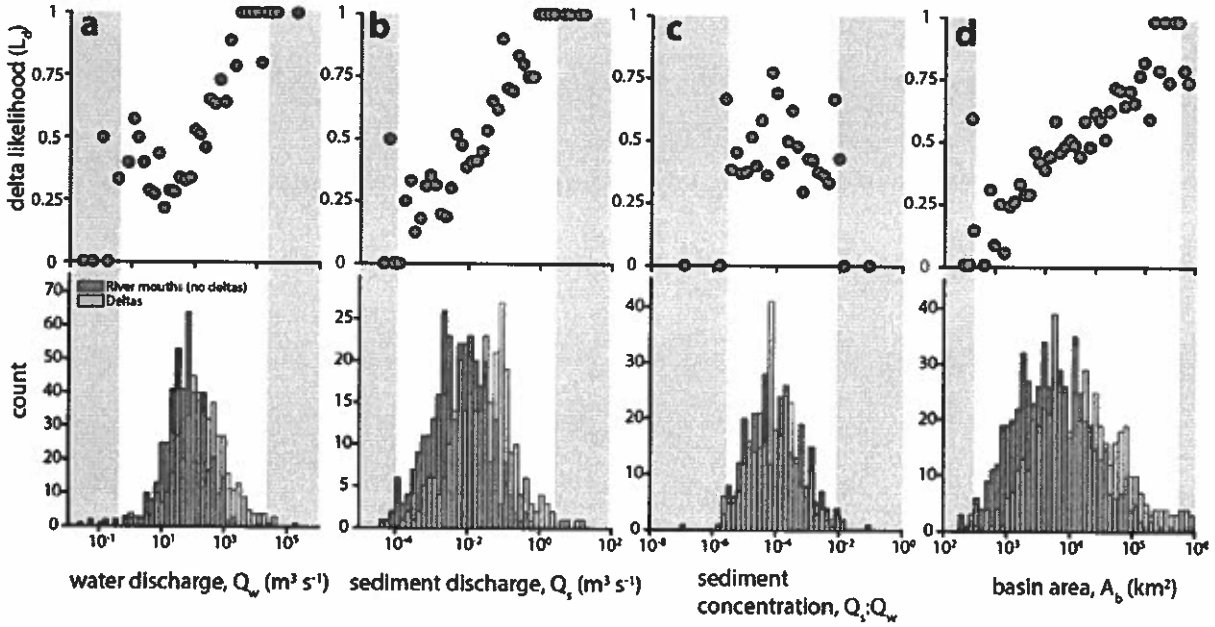
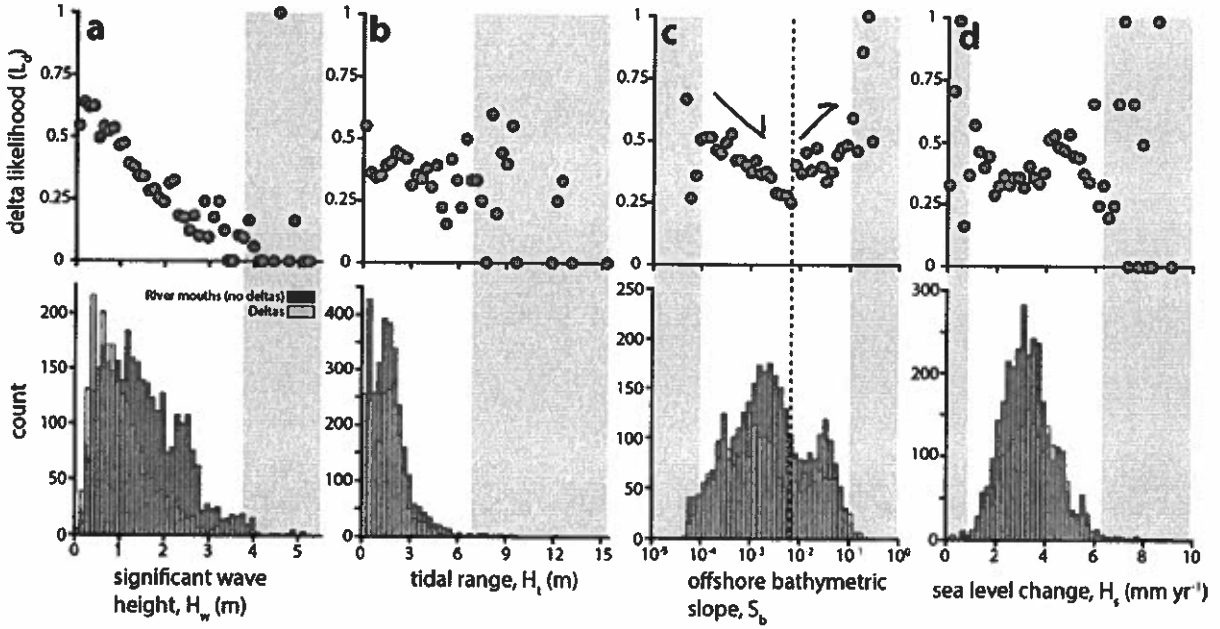


Figure 5. Differences in upstream environmental variables for rivers with and without deltas. (top panel) Scatter plots of delta likelihood, defined as number of rivers with a delta relative to total number of rivers in that interval. (bottom panel) Histograms binned into equal log-spaced intervals. Gray boxes outline ranges represented by 1% or less of total sample number.



5