A Validated Tropical-Extratropical Flood Hazard Assessment for New York Harbor

P.M. Orton[1*], T.M. Hall[2], S.A. Talke[3], A.F. Blumberg[1], N. Georgas[1], S. Vinogradov[1,4]

Accepted to Journal of Geophysical Research, 10/24/2016 – in press

[1] Stevens Institute of Technology, Castle Point on Hudson, Davidson Laboratory, Hoboken, NJ, 07030, USA

[2] NASA Goddard Institute for Space Studies, New York, NY, 10025, USA

[3] Portland State University, Department of Civil and Environmental Engineering, Post Office Box 751, Portland, OR, 97207, USA

[4] Now at Earth Resources Technology, Inc., NOAA Coast Survey Development Lab, 1315 East West Highway, Silver Spring, MD, 20910, USA

* Author to whom correspondence should be addressed; E-Mail: philip.orton@stevens.edu; Tel.: +1-201-216-8095

Three Key Points

1) A new approach for coastal flood hazard assessment is presented, using synthetic tropical cyclones

2) The return period for Hurricane Sandy's flood is 260 years; it was the largest back to at least 1700

3) Differing flood exceedance curves for extratropical and tropical cyclones require separate analysis

Running title: **Flood Assessment for New York Harbor**

Abstract

Recent studies of flood risk at New York Harbor (NYH) have shown disparate results for the 100-year storm tide, providing an uncertain foundation for the flood mitigation response after Hurricane Sandy. Here, we present a flood hazard assessment that improves confidence in our understanding of the region's present-day potential for flooding, by separately including the contribution of tropical cyclones (TCs) and extratropical cyclones (ETCs), and validating our modeling study at multiple stages against historical observations. The TC assessment is based on a climatology of 606 synthetic storms developed from a statistical-stochastic model of North Atlantic TCs. The ETC assessment is based on simulations of historical storms with many random tide scenarios. Synthetic TC landfall rates and the final TC and ETC flood exceedance curves are all shown to be consistent with curves computed using historical data, within 95% confidence ranges. Combining the ETC and TC results together, the 100-year return period storm tide at NYH is 2.70 m (2.51-2.92 at 95% confidence), and Hurricane Sandy's storm tide of 3.38 m was a 260-year (170-420) storm tide. Deeper analyses of historical flood reports from estimated Category-3 hurricanes in 1788 and 1821 lead to new estimates and reduced uncertainties for their floods, and show that Sandy's storm tide was the largest at NYH back to at least 1700. The flood exceedance curves for ETCs and TCs have sharply different slopes due to their differing meteorology and frequency, warranting separate treatment in hazard assessments.

1.0 Introduction

New York Harbor (NYH) and its connected tidal waterways lie at the apex of the New York Bight, and are surrounded by over 21 million people in cities such as New York, Newark, Jersey City and Hoboken. Hurricane Sandy in 2012 was a harsh reminder of the area's vulnerability to hurricane storm surges. Much of the land immediately surrounding NYH is filled-in former waterways and marshes, or flat coastal plains. The region was recently ranked second in the United States for present-day monetary risk from storm surge [*Botts et al.*, 2013]. Unfortunately, because the last flood comparable to Sandy occurred during the pre-instrumental era, in 1821 (**Figure 1**) [*Brandon et al.*, 2014; *Coch*, 1994; *Scileppi and Donnelly*, 2007; *Swiss Re*, 2014; *Talke et al.*, 2014], Sandy's flood elevation and area came as a surprise to many. Now, scientists, insurance companies, decision-makers and citizens across the region are re-assessing their flood risk.

There is strong momentum across the U.S. Northeast toward building flood mitigation projects, which require an understanding of coastal flood hazards for determining the design level of protection. New York City developed a $20 billion plan for coastal adaptation under the Special Initiative on Rebuilding and Resilience [*City of New York*, 2013]. This included hundreds of proposed actions to improve resilience, many of which are currently underway. Over $1 billion of funding has gone to city and state governments implement the winning projects of the "Rebuild By Design" competition, organized by the federal Department of Housing and Urban Development [e.g., *Orff et al.*, 2014]. Most recently, the US Army Corps of Engineers laid out a set of potential adaptation strategies in the North Atlantic Coast Comprehensive Study (NACCS)

67 report released in early 2015 [*USACE*, 2015]. Many of these plans seek to provide protection
68 against a 100-year flood or worse.
69
70 Researchers have made major advancements in coastal flood hazard assessment methodologies
71 in recent years. Lin et al. [2012] innovated within this area of hazard science by using a
72 dynamical-statistical approach for simulating TCs that can be run with climate model simulations
73 to account for future storm climatology changes. This work showed that future changes to storms
74 could have a similar effect as sea level rise, for worsening flood risk. One of FEMA's [2014b]
75 innovations has been to separately simulate and statistically evaluate TC and ETC flood
76 probabilities in an assessment of total storm tide risk, as the US Mid-Atlantic and New York
77 Bights are known to have severe floods arising from both storm types [*Colle et al.*, 2008; *Colle*
78 *et al.*, 2010; *Dolan and Davis*, 1992; *Orton et al.*, 2012]. A subsequent study similarly analyzed
79 flood risk for cold- and warm-season storms, using data from an innovative Monte Carlo
80 approach to create flood hazard statistics and to demonstrate the substantial seasonal differences
81 in flood risk [*Lopeman et al.*, 2015].
82
83 However, the most prominent recent assessments of the present-day flood hazard at New York
84 Harbor (NYH; the Battery tide gauge) disagree in their results, likely in part due to insufficient
85 model validation against observations. Studies that have examined all types of storms have found
86 the 100-year flood to be 2.44 m (2.19-2.93 m at 95% confidence) [*Zervas*, 2013], and 3.50 m
87 (with no confidence intervals provided) [*FEMA*, 2014b]. The Lin et al. [2012] study looked only
88 at TCs and found a present-day 100-year storm tide of 2.03 m (1.95-2.11 m at 90% confidence),
89 which is substantially different from the same metric in the FEMA study, at 2.75 m. The

associated return period estimates for Hurricane Sandy's storm tide range from 100 years [*FEMA*, 2014b] to ~1500 years [*Zervas*, 2013]. These substantial differences in both the TC-only assessments and the complete flood assessments (TC plus ETC) may arise due to insufficient validation of flood modeling for specific historical events – validation of storm tide modeling in the Lin et al. [2012] study was simply referenced to prior studies by other groups [e.g., *Colle et al.*, 2008] and not performed with the specific simplified meteorological modeling methods used in their study. Validation of the FEMA study included evaluations of tides and 7 historical storms, but did not include the largest storm surge event in the study period of 1938-2009, a 1950 ETC [*FEMA*, 2014b].

Several challenges complicate any assessment of the storm tide hazard for the region. Compared with the more heavily studied United States Gulf Coast, the larger tides in NYH increase the potential height of storm tides and also increase the magnitude of non-linear interaction between tides and surge. ETCs are the most common cause of flooding [*Colle et al.*, 2010], a cold-season contribution to flood risk that is often neglected for warmer climates and even for NYH [e.g., *Aerts et al.*, 2014]. TCs produce extremely rare, "long-tail" probability distribution events [*Lin et al.*, 2010], that are difficult to predict. The converging New York Bight coastline also adds the potential for amplification of rare events where strong winds blow from the southeast quadrant. However, TC storm tracks typically veer eastward as they approach, so the area has a relative scarcity of historical TC floods upon which to base a hazard assessment. As a result, a hazard assessment method is needed that is not overly reliant on local historical storm tracks, such as is the case with the widely-used Joint Probability Method [e.g., *Toro et al.*, 2010].

In this paper, we present a thorough modeling study that quantifies the "present day" NYH flood hazard. We address the difficulties illuminated above, by separately assessing TCs and ETCs and including detailed validation of modeling and assessment results, ensuring they are consistent with historical observations and assessments, within uncertainty. The present-day storm climatology is based on recent historical ETCs and synthetic TCs. We utilize a new statistical-stochastic TC simulation method that utilizes historical storm data from the entire North Atlantic basin (1950-2013) to create a set of synthetic storms that could affect that region. We also use an ETC assessment method where several historical storm events (1950-2009) are simulated with 50 repetitions of random tides, capturing the important contribution of tides to flood levels. **Section 2** summarizes methods including storm climatology development, atmosphere and ocean modeling, statistics, and uncertainty quantification. **Section 3** is a reassessment of the 1821 hurricane storm tide, as well as a similar but less-studied event in 1788, both useful for model validation and context for Sandy. **Section 4** presents study results including validations and exceedance curves (storm tide versus return period), **Section 5** is a discussion of results that addresses differences between ours and other studies, TC and ETC storm tides, looks at the role of tide-surge interaction, and discusses the implications of the findings regarding the 1788 and 1821 hurricanes, and **Section 6** lays out the primary conclusions of the study.

2.0 Methods

A detailed operational ocean forecasting system was leveraged to run three-dimensional ocean simulations for a flood hazard assessment that accounts for a full climatology of TCs and ETCs,

136 seasonal cycles of sea level, stratification and air temperature, and a complete range of the
137 region's possible tides. The goal is to accurately estimate storm tides at New York Harbor for
138 return periods from 5 to 10000 years. Storm tides are defined here as the water elevation above a
139 year's mean sea level. The hazard assessment separately integrates storm tide data over a
140 climatology of TCs and ETCs, then merges the exceedance probabilities to quantify the total
141 flood hazard. This is done because TCs occur less frequently and represent the worst flood
142 events in NYH history, whereas ETCs occur more frequently (many times per year) and
143 represent the more common, less-extreme flood events (**Figure 1**).
144
145 Methods for developing sets of TCs and ETCs representing their present-day storm climatology
146 are described in detail below in **Sections 2.1** and **2.2**, respectively. The historical record may not
147 adequately represent the large TC storm tides, as there have been only a small number of events
148 (**Figures 1** and **2a**). So, we use an ensemble approach where TCs are represented by a large
149 number of synthetic storms created with a statistical model. This model leverages historical TC
150 data from the entire North Atlantic Ocean from 1950-2013 to create a 1 million year (1 My)
151 record of TC activity. A subset of 606 TCs that spans the range of the TC flood hazard was
152 selected from the full storm set, and simple parametric equations are used to represent each TC's
153 wind and pressure forcing for the ocean model. We have no similar synthetic ETC model, but
154 nearly 200 years of historical ETCs have a well-defined range of storm tide with no significant
155 outliers (**Figure 1**) as long as we consider the powerful hybrid Sandy as part of the TC storm set.
156 This separation enables us to use an existing set of 30 of the worst historical ETC storm surge
157 events from 1950-2013 to represent the present-day climatology, as was done by FEMA [2014c].
158 The worst ETCs are simulated with 50 tide scenarios to capture a near-complete range of

possible storm tides for each storm, because of the high importance of tides for ETC flooding in this region (see **Section 2.4**). Meteorological data from atmospheric reanalyses is used for the ETCs as ocean model forcing.

The flood hazard statistics are based on hydrodynamic modeling, described below in **Section 2.3**, instead of historical water levels, for three primary reasons – (1) to merge tides and storm surge dynamically, as described in **Section 2.4**, as well as potentially other factors such as rainfall, (2) to estimate the storm tide over an entire region, not just at tide gauges, thus overcoming a limitation of tide-gauge based assessments; and (3) to account for realistic storm events and tide/storm combinations that have not occurred in the limited historical record. Synthetic events provide improved estimation of low probability events such as the 1% annual chance (100-year) storm tide, and even enable estimation of the 0.01% annual chance (10000-year) storm tide.

A heavy emphasis is placed on model validation using historical events, to leverage historical data and avoid bias. Validation should address the specific coupled ocean-atmosphere simulation methods, as error may exist in the ocean, atmosphere, and their coupling. Rare severe storm tide events are outside the normal range of variability, and thus forecast-oriented ocean models may be tuned toward normal conditions or annually-recurring storm conditions. In addition, atmospheric forcing data is typically a strong determinant of storm tide modeling accuracy [e.g., *Colle et al.*, 2015; *Orton et al.*, 2012]. Historical meteorological reanalyses are available for the ETCs [*FEMA*, 2014a], but simplified wind and pressure fields are used for the TCs, raising the possibility that oversimplification would lead to biases in the storm tide simulations. Historical storm validations for each storm type are given in **Section 4**, and include 42 historical events.

The storm tide error resulting from coupled ocean-atmospheric modeling is then incorporated in uncertainty analyses (**Section 2.5**).

2.1 TC modeling and reduced storm set

TC data are produced using a statistical-stochastic model of the complete life cycle of North Atlantic TCs, initially developed by Hall and Jewson [2007], and extended in Hall and Yonekura [2013]. The model consists of four components: genesis, propagation, and intensity (V_{max}) and size (R_{max}). Genesis, propagation, and V_{max} are documented in detail in Hall and Yonekura [2013]. Genesis and propagation are based on optimized local regression of HURDAT data [*Landsea et al.*, 2004] on repeating annual-cycle and interannually-varying climate covariates (500 mbar winds, North Atlantic subtropical SST with respect to the global subtropical mean, and ENSO state), while V_{max} is based on a random sampling scheme with perturbation. A space-date kernel-density PDF modulates the genesis rates to enforce seasonality. R_{max} is modeled by log regression with V_{max} and latitude covariates [e.g., *Vickery et al.*, 2000], using extended best-track data, 1998-2010 [*Demuth et al.*, 2006]. The error term is treated as lag-one autoregression.

The TC model is developed with historical HURDAT TC data from the entire North Atlantic Ocean from 1950-2013 to create a synthetic million-year record of TC activity. The 1950 cutoff represents a compromise between the desire to maximize data quantity and data quality. TC aircraft reconnaissance only began in the late 1940s, and there is evidence for undercounts of non-landfalling TCs in the earlier part of the HURDAT record [*Vecchi and Knutson*, 2011]. A cutoff of 1950 should largely avoid these early quality issues.

205 The probability of a storm's temporal maximum storm tide (η_{max}) exceeding a given arbitrary
206 level (η) is the multidimensional integral of the hurricane incidence probability times the
207 conditional probability of inundation:

208 $$P(\eta_{\max} > \eta) = \int_{\mathbf{x}} f(\mathbf{x}) P(\eta_{\max}(\mathbf{x}) > \eta) d\mathbf{x} \quad (1)$$

209 Here, **x** is the vector of hurricane hazard variables (e.g. landfall intensity, size, location, bearing,
210 and propagation speed), f(**x**) is the annual occurrence frequency of hazard **x**, and P(η(**x**)>η) is
211 the conditional probability that a storm of these characteristics will cause a storm tide in excess
212 of an arbitrary value η. The probability of hurricane event *f* at **x** is determined by the frequency
213 of its set of TC parameters in the large stochastic hurricane set. As it stands, the integral in Eq. 1
214 suffers from the "curse of dimensionality." It is too large to solve directly; for example, if **x**
215 represents 10 hazard variables, each sampled at 5 values, then there are nearly 10 million (5^{10})
216 combinations, or "nodes," at which $P(h|\mathbf{x})$ must be evaluated using the hydrodynamic model. As
217 noted in prior studies [e.g., *Condon and Sheng*, 2012; *Toro et al.*, 2010], approximation methods
218 must be utilized.
219
220 Our approach is to limit the number of hazard variables to five, with each coarsely sampled: five
221 landfall gates (**Figure 2b**), three landfall angles, six categories of intensity (wind speed), three
222 storm sizes, and three storm translation speeds, the latter three variables evaluated at the last 6-
223 hour time step prior to landfall. The intensity categories are 0 for tropical depressions and
224 tropical storms, and 1-5 for the Saffir-Simpson categories. Landfall angles are defined counter-
225 clockwise from gate parallel, and are 0-45, 45-90, and 90-180 degrees. These are not evenly
226 spaced, but have finer divisions for storms heading NW-to-N across gates 0 and 1, and NE-to-E
227 across gates 2-4, due to storms having preferred tracks in these directions, versus the larger

228 impact angles. The gates are chosen to capture all possible TC events with surges over 1.25 m
229 (e.g. **Figure 2a**), capable of causing storm tides exceeding ~2 m. The potential for low-bias in
230 our flood assessment, due to missing storms outside these gates, is evaluated at the start of
231 **Section 5.2**. This method is different from studies that use optimal sampling [e.g. Toro et al.
232 2010], as the storm set is not optimized and leaves us with some storms with very low annual
233 rates. However, our hydrodynamic/wave model is relatively fast compared with the models and
234 resolutions used in that study, and we did not see a need for optimization to reduce the number of
235 storms. Other studies use runs for all storms over a predefined period with a coarse, fast model
236 (SLOSH), then run the most important (e.g. largest) ones using a detailed model [Lin et al. 2010;
237 2012].
238
239 We refer to the specific combination **x** of hazard variables as the TC "flavor", and the
240 permutations of the five parameters give a total set of 810 TC flavors. The number of landfalls in
241 each flavor in the million-year dataset defines the annual landfall rate simply as the number of
242 events divided by 1,000,000 y. There are 204 storm flavors that are not observed to occur in the
243 dataset, so the total storm set of probable storms comprises only 606 storm flavors. For each
244 storm flavor, one storm is randomly selected as a representative event, and several of these tracks
245 are shown in **Figure 2**. Compared with the total number of storms crossing the five gates in the
246 million-year period, 290182 TCs, this set of 606 is much more tractable for numerical
247 simulation. To quantify the typical variability of storm tides within a flavor, we separately
248 simulated 50 events for 12 of the TC flavors, discussed below in **Section 2.5**.
249

Data needed for creating wind, tides and other forcing data are provided from the Hall TC model, including date and time (for tides and seasonality of temperature and salinity), year (for tides, between 1950 and 2013), latitude, longitude, radius of maximum winds (R_{max}) and speed of maximum winds (V_{max}). Additional parameters needed are the central pressure ($P_{central}$), pressure drop ΔP (or equivalently, the far-field pressure), and the Holland-B parameter, as described below. The Hall TC data are modeled on the 6-hour time step that is native to HURDAT, so we interpolate the data from 6 h to 60 min (SNAP grid; see **Section 2.3** on ocean model grids) and 15 min (NYHOPS grid) prior to projection onto the ocean model grids as forcing. The 6-hour time step where a TC makes landfall should not be interpolated, as this would lead to a continuous weakening over a 6 h period, whereas storms weaken more abruptly after making landfall. To avoid interpolation and excessive storm weakening before landfall, wind speeds in the last 6-hour time step before landfall are projected forward without change up to the first 6-hour time step after landfall.

2.1.1 Synthetic TC pressure forcing

The Hall TC model has not included central pressure ($P_{central}$) in prior publications, and a model was required here, due to the importance of atmospheric pressure for storm surges. $P_{central}$ and V_{max} are tightly, but imperfectly, correlated measures of TC intensity. A variety of regression relationships were tested to relate the central pressure deficit ($\Delta P = P_{env} - P_{central}$) to V_{max}. A successful candidate was:

$$\Delta P = a_0 + a_1 V_{max} + a_2 V_{max}^2 + a_3 V_{max}^3 + a_4 \sin(lat) + a_5 V_{max} \sin(lat) + \varepsilon \qquad (2)$$

The fitting was performed using HURDAT data from 1900-2007 for TCs, and using climatological pressure for the ambient pressure, P_{env}. However, data prior to the 1970s plays

almost no role, as there are few $P_{central}$ data. To test whether the resulting regression coefficients are significantly different from zero, a generalized jackknife test was employed, in which the model is refit multiple times with random 80% subsets of the data. We found that the linear term (a_1) is not significantly different than zero, but the remaining terms are. Thus, the model we use is:

$$\Delta P = a_0 + a_2 V_{max}^2 + a_3 V_{max}^3 + a_4 \sin(lat) + a_5 V_{max} \sin(lat) + \varepsilon \quad (3)$$

Here, $a_0 = -1.627$, $a_2 = 0.0197$, $a_3 = -7.448e\text{-}5$, $a_4 = 6.587$, and $a_5 = 0.542$ for units V_{max} m s^{-1} and $P_{central}$ in hPa. The error (ε) term constitutes the stochastic component of the model. It is $\varepsilon = \sigma x(t)$, where σ is the RMS variance of the residuals, the anomaly x(t) is lag-one autocorrelation (AR(1)): $x(t) = c*x(t-1)+sqrt(1-c^2)\ \varepsilon$, c is the autocorrelation coefficient (ACC), and ε the standard normal forcing. We found the best results by making σ depend on V_{max}: $\sigma = b_0 + b_1 V_{max}$. The error-model coefficients are fit by an iterative process to minimize overall model variance from observations, and are $b_0 = 0.944$, $b_1 = 0.161$, and ACC = 0.655.

The model we use for radial profiles of pressure, P(r), is Holland [*Holland*, 1980].

$$P(r) = P_{central} + \Delta P \exp\left[-\left(\frac{R_{max}}{r}\right)^B\right] \quad (4)$$

Here, it is assumed that the exponential pressure profile radius is equal to the radius of maximum winds (R_{max}), though for moderate-sized and large-sized storms, the pressure radius becomes larger. The differences are small relative to the natural scatter in both variables [*FEMA*, 2014c]. Parameter B controls the profile shape, and is related to the maximum wind speed and pressure deficit [*Holland*, 1980]:

$$B = V_{max}^2\ \rho\ exp(1) / \Delta P \quad (5)$$

Here, ρ is air density, computed from seasonally-varying air temperature.

296

297

298 ***2.1.2 Synthetic TC wind forcing***

299 The Sea, Lake, and Overland Surges from Hurricanes (SLOSH) wind profile model [*Jelesnianski*

300 *et al.*, 1992] can be expressed as:

301 $$V_{vortex}(r) = V_{max,sym}\frac{2R_{max}r}{{R_{max}}^2+r^2} \quad (6)$$

302 Here, V_{vortex} is the vortex wind speed, r is the radial location, $V_{max,sym}$ is the symmetrical

303 maximum wind speed, and R_{max} is the radius of maximum wind. The relationship between the

304 symmetrical maximum wind speed and the maximum wind speed is:

305 $$V_{max,sym} = V_{max} - \alpha V_{storm} \quad (7)$$

306 Here, α is the background wind modification factor, and V_{storm} is the storm translation speed.

307 Departing from normal NWS SLOSH wind modeling, for the contribution of the background

308 winds, we add the background wind velocity to the vortex wind velocity after using an empirical

309 estimate of inflow angle of 20 degrees, and a modification factor α of 0.55, shown to be more

310 accurate representations of modeled wind and storm surge [*Lin and Chavas*, 2012].

311

312 An empirical model of wind inflow angle is applied to the TC vortex winds [*Bretschneider*,

313 1972; *Phadke et al.*, 2003]. Lastly, the TC vector wind velocity is the vector sum of the vortex

314 wind velocity ($\mathbf{V_{vortex}}$) and background wind velocity ($\mathbf{V_{storm}}$) [*Lin and Chavas*, 2012]:

315 $$\boldsymbol{V_{TC}} = \boldsymbol{V_{vortex}} + \alpha \boldsymbol{V_{storm}} \quad (8)$$

316

317 TC maximum wind speed in the HURDAT database and from the Hall model is the 1-minute

318 maximum sustained wind, yet the ocean model requires 10-minute averages, so winds were

319 reduced by a 0.93 factor to account for gustiness [*Harper et al.*, 2010]. This is done because the
320 1-minute maximum sustained wind is actually a gust, and because the model's wind stress is
321 computed using a drag law that was developed using 10-minute average winds. An example of
322 synthetic TC wind forcing applied to a historic storm, the 1821 hurricane, is shown in **Figure 3**,
323 and validation of historical TCs with these methods is covered in **Section 4**.
324
325 **2.2 ETC modeling and storm set**
326 A set of 30 regionally severe ETC storm surge events from 1950-2009 was recently created for
327 the FEMA [2014b] study, with reanalyses of wind and surface pressure created by
328 Oceanweather, Inc. The reanalyses were constructed using the Interactive Objective Kinematic
329 Analysis system and a wide variety of data (e.g. buoy, satellite), as described in Cox et al.
330 [1995], and summarized in the recent FEMA [2014a] report. These 30 ETCs were selected from
331 a 60 year period (1950-2009), so each storm is given an annual rate of 1/60 y^{-1} in the assessment,
332 following a peaks-over-threshold type assessment approach [*Pickands*, 1975]. We use hindcasts
333 with historical tides for validation, but for the hazard assessment we simulate each storm with
334 random tides and the 19 largest ETC surge events are simulated 50 times with different random
335 tides, as explained in **Section 2.4**.
336
337 The FEMA ETC storm set we use here includes 8 of the top 10, and 19 of the top 30 ETC surge
338 events at the NYH Battery tide gauge in the period 1950-2009 (**Table 1**). Some events are
339 missing due to unavailability of good marine meteorological data for the earlier periods (1950s
340 and 1960s), and some are missing because the FEMA study was a regional study with the goal of
341 creating a storm set for the entire study area of New Jersey and the New York City region

342 including western Long Island Sound, not only for NYH. To account for the two missing storms
343 in the top 10, the annual occurrence rates of two storms with similar surges were simply doubled
344 in the statistical analysis. The surge ranked #5 for 1950-2009 was missing (19531107; 1.53m
345 surge) and was replaced by doubling the rate of #4 (19840329; 1.57m); #7 was missing
346 (19801025; 1.47m) and replaced by doubling the rate of #6 (19681112; 1.49m). Similarly, 9
347 storms ranked in the #11-30 range were missing, so nine storms' rates were doubled for the
348 statistical analysis (see **Table 1**). Lower surge events (below the top 30 in the 60 year period) are
349 not fully represented in the statistical analysis, but appear to have little effect on the tail of the
350 distribution which is of interest for 5-year and longer return periods.
351
352 Hybrid TC/ETC storms are classified differently in our study depending on how they are
353 represented in the HURDAT database. While Sandy officially became an ETC just prior to
354 landfall [*Blake et al.*, 2013], we include Sandy in the assessment as a TC because Sandy is in the
355 HURDAT database and its entire life cycle until landfall is part of the data that informs our
356 statistical TC model. On the other hand, the 19911031 ("The Perfect Storm") is in the ETC set,
357 though it was associated with a hurricane. According to HURDAT, the storm was purely an ETC
358 up to the time of the peak storm surge at NYH, and only later did it transition to a hybrid storm.
359 This storm had only a moderate surge at NYH of 1.41 m (ETC rank: #8; **Table 1**). Its data does
360 not inform the statistical TC model because by the HURDAT database it began as an ETC and
361 later became a TC. The impact of this separation is addressed in **Section 5.3**.
362
363 **2.3 Detailed ocean modeling using the NYHOPS system**

364 The Stevens ECOM (sECOM) three-dimensional hydrodynamic model [*Blumberg et al.*, 1999;
365 *Georgas and Blumberg*, 2009; *Georgas et al.*, 2014; *Orton et al.*, 2012] has been providing
366 highly accurate storm surge forecasts on its NYHOPS grid/domain (http://stevens.edu/
367 NYHOPS) for over a decade with mean water level errors of 0.10 m [*Georgas and Blumberg*,
368 2009]. The NYHOPS domain includes the Mid-Atlantic and Northeastern U.S. coastline from
369 Maryland to Rhode Island (e.g. **Figure 3**). The NYHOPS sECOM domain includes measured or
370 modeled freshwater and heat inputs from 544 sources including tributaries, power plants,
371 wastewater treatment plants, and sewer inputs [*Georgas and Blumberg*, 2009; *Georgas*, 2010].
372
373 A major uncertainty in storm surge modeling is the sea surface drag coefficient parameterization
374 [*Cardone and Cox*, 2009; *Lin and Chavas*, 2012; *Resio and Westerink*, 2008]. Hindcast studies
375 of two recent TCs with sECOM on the NYHOPS grid have found that replacing the wind drag
376 parameterization used by the operational system [*Large and Pond*, 1981] with a wave-slope
377 sensitive parameterization [*Taylor and Yelland*, 2001] gives improved results and RMS errors of
378 0.15 m for Tropical Storm Irene [*Orton et al.*, 2012], and 0.17 m for Hurricane Sandy [*Georgas*
379 *et al.*, 2014], and the latter is used here. Recent studies have presented evidence that the drag
380 coefficient hits a ceiling at high wind speeds [e.g., *Powell et al.*, 2003], but this saturation has not
381 been demonstrated for the shallow coastal ocean. Cardone and Cox [2009] point out that, due to
382 uncertainties in wind and air-sea interaction physics, it is common to tune the drag coefficient
383 ceiling based on model validations. Here, we find that the best fit for historical TCs (measured
384 with RMS error) is attained with a drag coefficient ceiling of 0.0035. We use the same ceiling
385 value for ETCs in this study, although the choice of ceiling value has a much smaller effect on
386 ETC storm tides because the ETC wind drag here almost never exceeds 0.0025.

Modeling of each storm with sECOM is first performed on the Stevens Northwest Atlantic Predictions (SNAP) grid that extends to the Gulf of St. Lawrence and Cape Hatteras and ~2000 km offshore. Having a domain of this large extent can be crucial for remote sea level forcing for large TCs and for all ETCs (**Figure 4**). For the SNAP grid, sECOM is run with wind and pressure forcing (no tides or streamflows), with constant ocean temperature and salinity. Water elevation boundary conditions are based on a local inverse barometer calculation using atmospheric pressure.

The modeling on the nested NYHOPS grid is set up with clamped offshore water elevation boundary conditions that are a sum of timeseries of tides, surge from the SNAP domain, and mean observed seasonal variations in mean sea level, which vary from -10 cm (January) to +7 cm (September). All model runs for TCs are 6 days, allowing for 3.5 days spin-up time prior to storm landfall. Model runs for ETCs are 10 days, allowing for similar spin-up time but longer storm durations. Vertically- and spatially-varying ocean initial conditions (salinity and temperature) and a spatially constant air temperature (important to wind stress) are set as seasonally-varying climatological values. Freshwater inputs have only a very small effect on temporal-maximum NYH storm tides, on the order of a few centimeters [*Orton et al.*, 2012], but modeled (TCs) and historical values (ETCs) were utilized to avoid any bias.

sECOM is coupled with a modified form of the Donelan [1977] wave model [*Georgas et al.*, 2007; *Georgas and Blumberg*, 2009; *Orton et al.*, 2012], including wave-slope sensitive wind stress as mentioned above [*Taylor and Yelland*, 2001] and wave-current bed stress. Bed stress is

410 adjusted due to wave-current interaction based on the theory of Grant and Madsen [*Grant and*
411 *Madsen*, 1979], simplified by assuming co-linear waves and currents [*Signell et al.*, 1990].
412 However, wave set-up of sea level due to breaking waves is not included. Inside a harbor like
413 NYH with a wide deep entrance, wave setup is not expected to be large and one prior published
414 study found a contribution to storm tide that averaged below 1.5% [*Lin et al.*, 2012].
415 Nevertheless, this constitutes a low-bias, and its elimination is discussed in **Section 5**.
416
417 **2.4 Tide forcing**
418 Tides were included in hydrodynamic simulations for each storm, to fully account for their well-
419 known nonlinear interaction with storm surge [e.g., *Horsburgh and Wilson*, 2007; *Pugh*, 1987;
420 *Tang et al.*, 1996]. The tides are modeled as normally is the case with the NYHOPS forecasting
421 system, using nine constituents, equilibrium arguments and nodal factors to accurately cover
422 semi-diurnal, diurnal, spring-neap, monthly, and interannual variations [*Georgas and Blumberg*,
423 2009]. Tide timeseries are imposed at the NYHOPS grid's offshore boundary and based on the
424 EC2001 tide database [*Mukai et al.*, 2002].
425
426 For ETCs, to include a wide variety of random tide-surge phasing and build a smoother rate
427 distribution for the statistical analysis, the 19 most intense storms (largest historical surge) were
428 each simulated 50 times with different random tides from any date in this same year range
429 (**Table 1**). Each of these storms' annual rates was correspondingly divided by 50 in the statistical
430 analysis. The remaining 12 storms are simulated only once and kept in the storm set for
431 validation and for future studies of broader regional interest (e.g. the New Jersey coast), but have

only a small influence on the area of the storm tide distribution that is relevant for 5-year and longer return period events.

With TCs, where we have a much larger storm set (606 storms), we use a computationally inexpensive approach of simulating only one random tide, but smoothing the probability distribution to account for tide variability. We randomly select a tide timeseries within hurricane season from 1950-2013. The distribution smoothing is done within the same framework as used for random storm selection (**Section 2.1**), by including random tides in the simulations of the 50 events for 5 of the TC flavors, discussed further in **Section 2.5**. This simplified smoothing approach for tides may lead to greater uncertainty, but on average the tides account for 36% of the peak ETC water levels in **Figure 1**, relative to only 10% for the TCs, and a prior study has noted a decreasing importance of tides for increasingly large TC storm tide events [Lin et al. 2012].

2.5 Statistical analysis and assessment uncertainty

Temporal maximum storm tides (η_{max}) were compiled for each model run's 10-minute outputs datasets for analysis using exceedance statistics. Rate distributions (similar to probability distributions) were formed from these data and each storm's annual rate. Similar to Niedoroda et al. [2010], we do not use fitted distribution types in place of the resulting empirical rate distributions of storm tides, because we do not want to enforce a theoretical shape on the distribution when the results of the physical models can define it instead. Moreover, we are representing the tails smoothly, in the case of ETCs by using a large number of random tide simulations, and in the case of TCs with a large number of storm flavors, and smoothing,

455 described below. The inverse cumulative rate distribution is equivalent to the exceedance
456 probability curve, and these are presented in **Section 4.3** with return periods (inverse
457 probabilities), as flood exceedance curves.
458
459 Each of the 606 TC flavors was represented using only one TC and one random tide event, as
460 explained above. Each modeled storm tide result represents $\eta_{max}(\mathbf{x})$ in **Eq. 1** as a single value, yet
461 one should ideally represent all the possible tides and TCs that occur within a flavor, and
462 represent $\eta_{max}(\mathbf{x})$ as a distribution of values instead. To study the distributions for a range of TC
463 storm tides, we separately simulated 50 events for 12 of the TC flavors, with four ~1.5 m, four
464 ~2.1 m, and four ~2.7 m TC flavors (average storm tide within the flavor). The chosen flavors
465 were chosen to span both the New Jersey and Long Island coasts and have relatively high rates
466 of occurrence (above 0.0001 y^{-1}). The storm flavors were segregated into these three categories,
467 and data for each flavor were differenced with their mean and merged to view the aggregate
468 distribution of TC storm tides for each of the three categories. These aggregated distributions
469 were fit with Generalized Extreme Value (GEV) distributions, shown in **Figure 5**, and show that
470 the larger events were approximately Gaussian, the smaller events exhibited a positive skew, and
471 the moderate events had a positive skew with a long-tail. The standard deviations of the small,
472 moderate, and large events were 0.29, 0.58 and 0.67 m, respectively. These results show a clear
473 'floor' on storm tides at about 1.0 m, which is due to tidal variations; high tide is typically 0.5-
474 1.0 m above mean sea level, and a surge event leads to higher values. The raw TC rate
475 distribution is smoothed using these three GEV distributions, interpolated to cover the entire
476 range of storm tide – GEV distribution parameters for the 1.5 m case were used for smoothing

477 the TC storm tide distribution from 0 to 1.5 m, interpolated between 1.5 m and 2.7 m, and
478 parameters for the 2.7 m case were used for smoothing the distribution at 2.7 m and above.
479
480 Uncertainty in the final TC flood exceedance curve was quantified as the spread above and
481 below the median using 1000 Monte Carlo simulations, where in each case random Gaussian
482 noise with a standard error equal to the total uncertainty was added. This uncertainty is the
483 combination of RMS error in the validation, in the annual rates of each TC flavor, and in the use
484 of using a single random tide and single storm per TC flavor. Regarding the annual TC rates,
485 based on the TC-model jackknife uncertainty analysis in Hall and Sobel [2013] we use a ±20%
486 (±2 sigma) uncertainty at the 0.01 y^{-1} rate, linearly increasing to ±80% at 0.00001 y^{-1}. Regarding
487 the use of a single tide and storm per TC flavor, we use standard deviations described above and
488 illustrated in **Figure 5**.
489
490 Uncertainty in the final ETC flood exceedance curves was also quantified using 1000 Monte
491 Carlo simulations, including a bootstrap re-sampling of historical events, of which there were
492 effectively 41 after the doubling of rates for 11 storms. Random Gaussian noise was first added
493 to results of each of the simulated storms, with a standard deviation equal to the RMS error in the
494 validation (0.19 m). By running 50 iterations with random tides for all the important ETCs, it is
495 assumed that uncertainty due to the contribution of random tide assignment is negligible.
496
497
498 **3.0 Reassessment of extreme NYH storm tides back to 1700**

Multiple hurricanes flood events in the 18th and 19th centuries [*Brandon et al.*, 2014; *Scileppi and Donnelly*, 2007] suggest that large, possibly Hurricane Sandy magnitude events may occur once every order-of-100 years, so a better understanding of these rare extreme events is needed to understand flood risk in NYH. Qualitative historical storm and flood reports for NYH are available from newspapers and historical archives, and have been extensively reviewed [*Boose et al.*, 2001; *Kussman*, 1957; *Ludlum*, 1963]. Estimates of wind velocity are also available for all events affecting the New England and NYC area back to the 1600s [*Boose et al.*, 2001; *Boose et al.*, 2003]. These data suggest that the estimated Category-3 hurricanes of 1788 and 1821 exceeded 2.5 m and possibly even 3.0 m storm tides [*Kussman*, 1957; *Ludlum*, 1963; *Scileppi and Donnelly*, 2007].

To better understand and constrain the contribution of the 1788 and 1821 events to flood hazard, we reassess available archival data below (**Section 3.1)** and then simulate these events using best available data to check archival estimates (**Section 4.2**). To ensure against model bias, we apply the same modeling approach to reassess an August 1893 Category-1 hurricane whose storm track passed directly over NYH, as well as several other TC events. Such validation is necessary because archival estimates based on contemporary accounts can be mistaken; for example, the 1893 event was long thought to have caused a storm tide of 3.0 m +/- 0.5 m at NYH [*Scileppi and Donnelly*, 2007], but was recently shown to have registered a peak storm tide of only 1.6 m at NYH [*Talke et al.*, 2014].

3.1 The 1821 Norfolk and Long Island Hurricane

Records suggest that the 1821 hurricane struck NYH at low water, with peak winds occurring between 6 pm and 7:30 pm [*Espy*, 1839; *Redfield*, 1831]. Data for the storm such as wind magnitude, direction, and track were also compiled by Redfield [1831], [*Espy*, 1839], and modern sources [*Boose et al.*, 2001; *Boose et al.*, 2003; *Kussman*, 1957; *Ludlum*, 1963; *Swiss Re*, 2014], and suggest a fast moving storm eye (~50 km h^{-1}) with a large storm surge or storm tide from Norfolk (3 m) to New York. Using 19th century hourly tide data from Governors Island (NYH) to compute the tidal constituents [see *Talke et al.*, 2014], we hindcast that a low tide of ca. -0.4 m below sea-level would have occurred at approximately 18:45 local time (EST) on Sept. 3, 1821.

Wind speed measurements were not available during the 1821 storm, but dozens of observers along the U.S. East Coast made detailed observations of wind direction (eight points of the compass), qualitative reports of speeds, and detailed reports of damages [*Ludlum*, 1963]. These have recently been summarized [*Boose et al.*, 2001; *Elsner*, 2006], and two estimated timeseries of storm location, maximum wind speed, and radius-to-maximum-wind have been derived based on this information [*Boose et al.*, 2003]. Upon arrival at NYC, Boose et al. [2003] estimate a R_{max} of 50 km, a V_{max} of 51.4 m s^{-1} (a weak category 3 hurricane), and a storm translation speed of 12 m s^{-1}. Separately, central pressure, R_{max}, and translation speed near landfall have been estimated to be 965 mbar, 56 km and 15 m s^{-1} [*Elsner*, 2006].

A recurrent misunderstanding about the extreme water levels at NYH produced by the 1821 event stems from Redfield [1831], who states that “a 13 foot storm tide occurred from low water”. Subsequent investigations have interpreted *low water* as the predicted low tide of -0.4 m

545 on September 3rd, which implies that a storm tide of 13 ft (3.96 m) reached a height of
546 approximately 3.2 m above MSL1821 [*Ludlum*, 1963; *Scileppi and Donnelly*, 2007]. However,
547 contemporary newspapers upon which Redfield [1831] is based reported an exact surge of 13 ft,
548 4 in (4.0 m) in the East River relative to *low water mark* [The Gazette and General Advertiser,
549 Sept. 7 1821; see *Boose et al.*, 2003; *Kussman*, 1957]. Here, as suggested also in Kussman
550 [1957], we assert that the *low water mark* (LWM) was most likely an extreme low water datum,
551 rather than the predicted low water for the evening of Sept. 3rd, 1821.
552
553 The inference of an LWM datum is supported by nautical maps such as Blunt's American Coast
554 Pilot of 1827, or the US Coastal Survey map of 1844. These early maps, which referenced their
555 depths to tidal datums using tide measurements, base their topography on an extreme low water
556 datum called '*low water'*. The 1844 Hassler map datum explicitly uses the measured spring low
557 water on Oct. 23rd and 24th, 1835, which harmonic analysis suggests was approximately -1.0 m
558 below MSL1835. Considering typical variation between measured and predicted tides, we
559 suggest that the *LWM* datum used pre-1850 was approximately -1.0 +/- 0.2 m below MSL1821.
560 We further assume that the 1821 measurement of 4.0 m carries a precision of +/- 0.2 m (we
561 consider it probable that the storm tide estimate was made after the fact based on water-line or
562 debris-line measurements, rather than direct observation, given the severity of the event and the 4
563 day time lag before the number was published).
564
565 Considering these factors, we estimate that the storm tide of Sept. 3, 1821 was approximately 3.0
566 m (2.5-3.3 m at 95% confidence) (**Table 2**), and represents a storm surge of 3.4 m, or 0.6 m
567 larger than the hurricane Sandy storm surge. Our storm tide estimate approximately agrees with

568 the Kussman [1957] storm-tide estimate of 2.7 to 3.0 m relative to MSL1821, based on
569 contemporary flooding accounts of structures, and is within the error bounds of the 3.2 +/- 0.5 m
570 estimate of Scileppi & Donnelly [2007]. Because mean sea level at the Battery was 0.50 m lower
571 in 1821 than 2012 [*Kemp and Horton*, 2013], and the storm tide was 0.4 m smaller (**Table 2**), the
572 1821 peak water elevation was 0.9 m lower than Sandy in absolute terms.
573
574 **3.2 The 1788 hurricane**
575 News accounts and data compiled by Boose et al. [2001] and Ludlum [*Ludlum*, 1963] describe a
576 small and fast hurricane that made landfall in Delaware Bay and traveled overland to NYH,
577 parallel to the coast. "The storm that began on Monday, continued on Tuesday, and augmented to
578 a greater degree of violence, wind at SE and ESE. At half past 11 increased to a perfect gale,
579 which raged near half an hour with irresistible fury. Shipping would have been greatly damaged
580 had not the wind shifted suddenly to E.N. and NW about mid-day." (The Providence Gazette and
581 Country Journal, Tues. 8/30/1788). "For upwards of 23 minutes it blew with incredible fury; and
582 had the wind not suddenly shifted to the N every vessel in the harbour must have been drove on
583 shore." (The Independent Journal (NYC), Wed. 8/20/1788).
584
585 Upon arrival at NYC, Boose et al. [2001] estimate a R_{max} of 50 km, a V_{max} of 51.4 m s^{-1} (a weak
586 category 3 hurricane), and storm track and timing data that imply a storm translation speed of
587 17.5 m s^{-1}. The Boose storm track estimate suggests a fast storm translation speed of 17.5 m s^{-1}
588 [*Boose et al.*, 2003]. The Boose paper estimates storm sizes in categories of 50 km, 75 km and
589 100 km, but the news accounts and debates about whether it was a squall or a hurricane [*Ludlum*,
590 1963] suggest it was smaller than the 50 km minimum scale allowed in the Boose study. Also,

591 the extremely brief period of high winds has been used to infer that the 1788 hurricane was a
592 small, fast-moving storm [*Ludlum*, 1963]. If we take 11:30 EST to be the time of maximum
593 wind, 11:53 EST ("23 minutes" later) to be the time the eye began passing over NYC, and 12:00
594 EST ("about mid-day") to be time of the directional switch, then the storm center passed at about
595 ~11:57. A 27-minute time with a storm traveling ~63.5 km h^{-1} leads to an estimate of 29 km for
596 the radius of maximum winds. We estimate with tide-only modeling that the high tide on 19
597 August 1788 at 09:30 EST would have been +0.78 m, and the peak water level of 2.9 m would
598 have occurred 2.75 h after high tide, at 12:15 EST, when the tide was at -0.1 m MSL. This
599 contrasts with 1821, where the storm surge occurred at low tide.
600
601 Flood-driven damages of the storm in NYH included destruction of the Battery seawalls, many
602 wharves, several bridges, and a filling in of many cellars on Lower Manhattan streets such as
603 Front, Water and to a lesser extent Queen (now called Pearl Street) [*Boose et al.*, 2003], similar
604 to the 1821 TC. "It was full sea at 6 minutes past 11 A.M. but it still flowed until 12, when there
605 was a rise of water of at least 5 feet higher than usual, which filled all the cellars in Front-Street,
606 besides many in Water Queen-Streets, and on the North-River [Hudson River]." (The Providence
607 Gazette and Country Journal, Tues. 8/30/1788). Wind-driven damages from the roughly 50 m s^{-1}
608 winds were also severe, with chimney and roof damage to many homes on Manhattan [*Boose et*
609 *al.*, 2003].
610
611 The 1788 hurricane's storm tide has previously been estimated as 3.0 m at NYH (error bars from
612 2.5 to 3.5 m), based on the flooding of basements on Front and Water Streets and their present-
613 day elevation [*Scileppi and Donnelly*, 2007], and multiple studies state or imply that the storm

614 tide was lower than that of the 1821 TC in the region [*Donnelly et al.*, 2001; *Scileppi and*
615 *Donnelly*, 2007]. We interpret "… at least 5 ft higher than usual" as meaning 5 ft above spring
616 high tide, meaning the 'usual variability of water', not mean high water [Again, the pre-1850
617 convention was to use spring tides when discussing a tidal datum; see e.g.,*Shalowitz*, 1964]. By
618 that interpretation, the storm was at least 1.52 m (5 ft) above the spring-high water line of 1.0 m,
619 or, in other words, a storm tide of "at least" 2.5 m. Based on the water levels and damages
620 described above, it is clear that it was nearly as high a storm tide as the 1821 hurricane, giving a
621 likely range from 2.5 m up to the 1821 high-end estimate of 3.3 m.
622
623 Our final quantitative estimate of the observed 1788 storm tide comes from newspaper reports in
624 September 1821, which state that the high water line for the 1821 event exceeded the previous
625 high at a tidal mill by 0.10 m ["4 inches"; *Kussman*, 1957, page 47]. No other known hurricane
626 storm tide (e.g. 1805,1815) came close to the 1788 event, according to news quotes from Boose
627 et al. [2003], so the 'previous high' appears to have been from 1788. Along with our 1821
628 estimate of 3.0 m, we use this to finalize our estimate as 2.9 m (2.3 to 3.3 at 95% confidence) for
629 the 1788 TC storm tide. We use a larger low-end uncertainty for both the 1788 and 1821 events
630 to account for our observation that contemporary accounts (e.g. newspapers, diaries) often over-
631 estimate, but seldom underestimate, storm tide magnitudes (e.g. the 1893 event mentioned at the
632 top of Section 3). The resulting temporal-maximum storm surge estimate, differencing this
633 estimated storm tide and the modeled tide, is 3.0 m (**Table 2**). Because mean sea level at the
634 Battery was 0.56 m lower in 1788 than it was in 2012 [*Kemp and Horton*, 2013], the 1788 peak
635 water elevation was about 1.05 m lower than Sandy in absolute terms.
636

4.0 Results

Before presenting results for the flood hazard posed to NYH by TCs and ETCs (section **4.3** below), we first show that our model reproduces historical events (section **4.2** below) and that our modeled TC storm climatology approach (described in section **2.1**) produces landfall rates that are consistent with those estimated from a historical observation database (section **4.1** below)

4.1 Tropical cyclone climatology and landfall return period validation

The modeled annual rates, intensities, storm sizes and storm tides of each of the 606 TC flavors are summarized in **Figure 6**. The TCs are organized in order of the gate crossed, from 0 to 4 as defined in **Figure 2**. Rates of landfall of any TC through gates 0 through 4 are 0.058, 0.012, 0.035, 0.065, and 0.121, respectively. Gate 4 is the most likely to be crossed by a TC in any given year, due to being a relatively long gate and being in line with the more common offshore storm tracks. While storms crossing Gate 0 are half as likely as Gate 4, they typically cause a higher storm tide because the largest observed winds are to the right of the storm track in the northern hemisphere (**Figure 6b**). Broken down by intensity, categories 0 through 5 occur at annual rates of 0.205, 0.0529, 0.0200, 0.00977, 0.00294 and 0.000025, or return periods of 4.9, 19, 50, 100, 340 and 40000 y. Integrated over all events, there is between a 1/3 and 1/4 chance of a TC occurring in any given year (probability = 0.29). Qualitatively, these return periods match with historical experience; for example, 3 Cat-1 or higher TCs are believed to have struck the

660 NYH area (gate 2) in the last 240 years (1788, 1821 and 1893), or once every 80 years, giving an
661 observed probability similar to the 0.009 y^{-1} return period we estimate.
662
663 The modeled probability of a TC passing any of our gates (**Figure 6**) is validated by comparing
664 against the measured HURDAT database from 1950 to 2013 (**Figure 7**). The modeled
665 probability (red line) of an event with a certain sustained wind speed follows historical
666 observations (blue line) well. Both the modeled and measured curve are concave down, such that
667 larger wind speed events have a sharply lower probability of occurrence in this area. Model
668 landfall is unbiased, or consistent with historical landfall rates, to the extent that the observed
669 return period curve falls inside the 95% confidence range of model-based return periods (orange
670 lines), which is the case.
671
672 **4.2 Storm tide modeling validations**
673 We next validate that our modeling approach by comparing measured and simulated storm tides
674 for ETCs (**Figure 8**). Modeled results show a small mean bias of -0.03 m with an RMSE of 0.19
675 m. This mean bias is subtracted from all modeled ETC storm tides before use in the statistical
676 analysis. After bias correction, only four events have an error on the peak storm tide of greater
677 than 0.30 m, the 19610413, 19660123, 19911031 and 19921211 storms, at 0.34, 0.37, 0.41 and -
678 0.31 m, respectively.
679 A close look at **Figure 8** shows that the 1950 ETC model result (storm #1) is nearly identical to
680 the observed historical values. The 1950 ETC simulation was initially found to substantially
681 overestimate the observed storm tide by 1.2 m (58%). This was not likely a problem with the
682 sECOM model; we also performed simulations with the same 1950 meteorological reanalysis

683 data using the model ADCIRC in unpublished work related to a separate study [*Orton et al.*,
684 2015a], and found the same amount of overestimation. This strong overestimation, coupled with
685 this event being easily the largest storm surge in the ETC storm set, together motivated a
686 reduction of wind speed to create a model simulation that better represents the historical event. A
687 wind reduction factor of 0.83 was found to create a near-perfect fit to the event's peak storm tide,
688 as well as an RMSE of 0.16 m and mean bias of -0.03 m for the storm tide timeseries. The
689 justification for this adjustment is discussed in **Section 5.3**. Due to this tuning of wind speed, this
690 storm's results are not included in the statistics shown in **Figure 8**.
691
692 A set of 12 historical events (**Table 2**) was selected for validation of the TC atmosphere-ocean
693 simulation methods, including five large surge events from 1.7-3.4 m (1788, 1821, 2012 Sandy,
694 1960 Gloria, 1985 Donna) and another seven that are in the more common 1.0-1.3 m surge
695 range. The TC validation data is built by comparing peak storm tide (η_{max}) model results to
696 observations (**Figure 9**). For the 1788 and 1821 hurricanes, this is the maximum model result in
697 10-minute output timeseries, as the observation is a high water mark. For the other storms, it is
698 the maximum model result in decimated output timeseries, as we use the hourly historical
699 observation timeseries from the Battery [*NOAA*, 2015] or Fort Hamilton, two measurement
700 stations in NYH [*Talke et al.*, 2014]. To emulate the methods used for the synthetic TCs, the
701 wind and pressure for historical TCs were modeled using methods summarized in **Section 2.1**,
702 with data at 6-hourly intervals.
703
704 Parameters for validation storms are from the HURDAT database [*NOAA-NHC-HRD*, 2015],
705 Extended HURDAT [*Demuth et al.*, 2006], the recent FEMA study [*FEMA*, 2014a], and for the

706 oldest storms Boose et al [*Boose et al.*, 2003]. The 1788 storm simulation, however, uses an R_{max}
707 of 29 km, as estimated in the prior section, due to the strong evidence for it being a smaller storm
708 than the minimum in the Boose study. Pressure timeseries for the 1788, 1821 and 1893 storms
709 are from **Eq. 3**, with central pressures around the time of passage at NYH of 952, 952 and 976
710 mbar, respectively.
711
712 The model mean bias for the TC hindcasts using parametric forcing is below 1 cm and root-
713 mean-square error (RMSE) is 0.32 m (**Figure 9**). Hurricane Sandy's peak storm tide is
714 overestimated by 0.71 m, whereas a few other storms are underestimated by 0.35-0.46 m (1944,
715 1960 and 1821). However, a majority of the storms are simulated very accurately, within 0.20 m
716 (1788, 1893, 1954, 1976, 1985, 1999, and 2011). The 1788 modeled storm tide of 2.95 m was
717 0.05 m higher than our archival estimate, and the 1821 modeled storm tide of 2.66 m was 0.34 m
718 lower the archival estimate. The 1821 TC meteorology is highly uncertain and it is not surprising
719 that the storm tide modeling error would be large. Sandy was a hybrid ETC/TC before landfall,
720 and therefore simplified parametric methods may not accurately represent the event physics; in
721 fact, modeled wind speed peak was far too high (29%) and RMS error was also high (5.6 m s^{-1})
722 in New York Bight, compared with observations. The large overestimation of the Sandy storm
723 tide can be explained largely by these excessive wind speeds, as sECOM runs with an accurate
724 meteorological reanalysis have lower wind speeds and give very accurate results [*Orton et al.*,
725 2015b]. This is likely due to Sandy's being a hybrid storm; in looking at other recent TCs (e.g.
726 Gloria, Irene) the parametric wind modeling is much more accurate (wind speed RMS errors of
727 2.8 and 3.5 m s^{-1}, respectively). These considerations show that the hydrodynamic model is
728 working well, and suggest that though individual events may be over or underestimated with a

parametric approach, an ensemble with a sufficient number of events will yield unbiased estimates of storm tide hazard.

4.3 Annual rate distributions and flood exceedance curves

Annual rate distributions for ETCs and TCs are shown in **Figures 10-11**, and show a strong contrast in distribution shapes. The annual rate distribution for ETCs has a short tail, with the highest ETC storm tide at 2.67 m. The 1950 ETC is an important storm, due to its largest observed ETC storm surge (2.41 m), and its 50 simulated events with random tides have a spread from 2.00 to 2.67 m. The annual rate distribution for TCs (**Figure 11**) and the smoothed version (explained in **Section 2.5**) are both very long-tailed, with low-rate events stretching out to the highest value of 5.77 m (a 0.000077 annual rate). Many storms represent the tail of the distribution, with 127 storm tide events above 3 m, and 12 events above 5 m.

Storm tide exceedance curves with 95% uncertainty shading are shown for ETCs and TCs in **Figures 12-13**. The results show 100-year storm tides of 2.28 m (1.88-2.65) for ETCs, and 2.65 m (2.34-2.98 within 95% uncertainty) for TCs. The uncertainty for the ETC flood exceedance curve is slightly larger due to the reliance on a small number of historical events over a 60-year period, and is quantified with the bootstrap re-sampling of these events in the Monte Carlo uncertainty analysis.

The model-based assessments for ETCs and TCs are broadly consistent with tide gauge observations, based on comparisons with observation-based return period analyses and empirical point-by-point return periods (**Figures 12-13**). The ETC observational data are annual maximum

752 storm tides for events that were not matched up to a HURDAT historical TC, and TC
753 observational data are annual maxima for events that are in HURDAT. Annual mean sea levels
754 have been removed to obtain the observed storm tides. Both the ETC and TC observation-based
755 assessments use a Generalized Pareto Distribution (GPD) distribution fit of data from 1950-
756 2013.
757
758 The ETC, TC, and combined storm tide exceedance curve are all shown in **Figure 14**. The
759 results of the combined assessment show 10-, 100-, 1000-, and 10000-year return period events
760 of 1.94 m (1.78-2.14 at 95% confidence), 2.70 m (2.51-2.92), 4.35 m (4.11-4.65), and 5.77 m
761 (5.37-6.37) respectively. The storm tide hazard for shorter return periods is produced primarily
762 by ETCs, while the risk of larger events with 100-year return periods or longer is almost
763 exclusively driven by hurricanes.
764
765
766 **5.0 Discussion**
767 It is imperative that model-based flood hazard assessments have rigorous model validation,
768 together with a verification that final flood exceedance curve results are not inconsistent with
769 observation-based assessments. Our TC and ETC flood exceedance curves are consistent with
770 historical data, to the extent that our curves fall within the uncertainty of empirical return periods
771 and observation-based GPD curves (**Figures 12-13**). The ETC flood hazard at NYH is relatively
772 well-constrained, with a 95% confidence range for the 100-year storm tide from 1.90-2.35 m,
773 and our curve and uncertainty overlap. Our TC assessment results are also consistent with the
774 observations, though the observation-based uncertainties grow larger with higher return periods,

775 making it easier to achieve "consistency" between the observation- and model-based curves. As
776 such, this type of comparison is a more useful evaluation or validation for shorter return periods,
777 but less useful, and not a powerful validation for longer return periods greater than the length of
778 the observational record. Moreover, cases can arise where the observational record is short and
779 lacks any extreme surge events, resulting in a lack of consistency with the model-based
780 assessment. It may be necessary in these cases to use news reports (e.g. the 1788 and 1821 storm
781 in this study) or paleorecords to better characterize the historical assessment of the flood hazard
782 [*Lin et al.*, 2014].
783
784 The observational and model-based results also demonstrate the importance of separately
785 evaluating TCs and ETCs, for locations where both influence flooding at the return periods of
786 interest. Flood return period curves for ETCs and TCs at this location have sharply different
787 slopes, due to their differing meteorology. One primary inflection point is visible in the modeled
788 flood exceedance curve (**Figure 14**). The lower return periods, up to a return period of about 100
789 years, have a concave shape, due to the switchover from ETC to TC dominance. The higher
790 return periods above 100 years have a convex appearance. As a result, flood hazard assessments
791 must treat the storm types separately in validation and return period computations. Fitting a
792 single probability distribution to a combined storm tide dataset with this inflection would lead to
793 a poor representation of the flood hazard.
794
795 Physically, the convex shape of the flood exceedance curve at high return periods is likely a
796 direct result of the convex shape of the TC wind speed exceedance curve (**Figure 7**). The latter is
797 an expected result of TCs traveling northward into the mid-latitudes and encountering declining

798 sea surface temperatures, leading to declining intensity and a lower probability of the strongest
799 TCs [*Emanuel*, 2000]. The convex TC flood exceedance curve at high return periods is also
800 likely a result of negative feedbacks on storm tide, such as wind drag saturation, tide-surge
801 interaction, and oceanic mixing of cold subsurface water, which can both weaken storms and
802 reduce wind stress. However, extratropical transition is only coarsely captured in our assessment
803 through the parametric TC variables, for example the increase in TC size with latitude. Thus,
804 uncertainty in our results for longer return periods is likely underestimated. The concave shape
805 of the observation-based flood exceedance curve in Figures 13 and 14 leads to higher storm tides
806 at long return periods than the model-based curve. This is at least partially a result of including
807 Sandy in a GPD analysis of only a 64-year record (1950-2013; in **Figure 13**), whereas we know
808 it was the highest event in over 314 years.
809
810 Results clearly demonstrate that ETCs dominate relatively short return period storm tides which
811 occur every 5-10 years, while TCs drive risk for the 100-year and larger storms tides. The cross-
812 over return period for the TC and ETC curves is at 60 years (**Figure 14**); both storms have a 1/60
813 annual probability of producing a 2.3 m storm tide. This means that a 30-year storm tide (2/60
814 annual chance) is equally likely to be caused by a TC as an ETC (uncertainty from 10 to 91 years
815 at the 95% level, sampling from the Monte Carlo analysis results). This result contrasts with the
816 FEMA [2014b] study, where results show ETCs driving the flood level for all return periods at
817 NYH.
818
819 Other recent studies have found substantially different exceedance curves at NYH, as highlighted
820 in **Section 1**. The 100-year storm tide at NYH has had estimates ranging from 2.44 m [*Zervas*,

821 2013] to 3.50 m [*FEMA*, 2014b]. Our result of 2.70 m falls between these studies and is within
822 the 95% confidence range of NOAA's historical-data based estimate (2.19-2.93 m) [*Zervas*,
823 2013] and our own historical-data based estimate (**Figure 14**). Another more recent study found
824 the 100-year storm tide to be 3.38 m (2.53-5.78) after removing their 2013 mean sea level offset
825 [*Lopeman et al.*, 2015]. Our model-based 10-year return period storm tide of 1.94 m (1.78-2.14
826 at 95% confidence) is within our observation-based range (1.78-2.03 m) and fairly central in the
827 range of most prior studies, comparing well to the Talke et al. [2014] result of 1.91 m, the
828 Lopeman et al. [2015] result of 1.99 m (1.89-2.13), and the NOAA result of 1.85 (1.77-1.96)
829 [*Zervas*, 2013]. An outlier is the FEMA 10-year storm tide [*FEMA*, 2014b] result of 2.19 m.
830 Comparing longer return period events, only estimated in the Lin et al. [2012] TC study, our
831 results of 4.35 m (4.11-4.65) and 5.77 m (5.37-6.37) for the 1000-year and 10000-year storm
832 tides are larger than their results of 3.55 and 4.75 m, respectively. The return period for a
833 Hurricane Sandy level storm tide of 3.38 m in our assessment is 260 years (170-420 at 95%
834 confidence), in comparison to values that can be inferred from other studies' as ~100 years
835 [*FEMA*, 2014b; *Lopeman et al.*, 2015], ~850 years [*Lin et al.*, 2012], and ~1500 years [*Sweet et*
836 *al.*, 2013; *Zervas*, 2013].
837
838 **5.1 Sensitivity of results to hydrodynamic model physics**
839 Differences between our study and the other recent studies are discussed separately for TCs and
840 ETCs below, but hydrodynamic modeling differences used for both type of storm may be
841 partially responsible. A major difference distinguishing factor between studies in this region is
842 the sea surface drag coefficient parameterization and its ceiling (**Section 2.3**). We use a wave-
843 slope sensitive parameterization [*Taylor and Yelland*, 2001], and find that the best fit ceiling

844 value is 0.0035, for fitting historical TC storm tides (measured with RMS error) with our
845 simplified wind fields. Prior studies in the area have used the [*Garratt*, 1977] parameterization
846 with ceilings of 0.0025 [*Lin et al.*, 2012], 0.0030 [*Wang et al.*, 2014], 0.0035 [*FEMA*, 2014b],
847 and with no ceiling [*Colle et al.*, 2015]. Misalignment of wind and swell can be a source of
848 higher drag [e.g., *Holthuijsen et al.*, 2012], but these processes are not well understood, and other
849 factors such as storm translation speed are important [e.g., *Reichl et al.*, 2014] and not
850 represented by existing wind speed-based drag parameterizations. In our study region of the
851 funnel-shaped coastline of New York Bight, this saturation ceiling could often be higher than
852 straight coastlines due to misalignment. This can arise due to wave refraction bending waves
853 shoreward on the broad continental shelf, or due to the fact that the winds that drive storm surge
854 into the apex of the Bight are often not from the southeast (they are from the east or northeast).
855
856 A recent study separated surge, tide and sea level rise, computed statistics and used data from a
857 Monte Carlo simulation to create hazard statistics as the sum of the three components, and found
858 a 100-year flood level of 3.38 m, greater than our storm tide estimate by 0.67 m [*Lopeman et al.*,
859 2015]. We quantify the possible bias here from omission of tide-surge interactions in the storm
860 tide hazard, by repeating our assessment with sums of randomly-shuffled modeled tide
861 timeseries plus modeled surge timeseries, to emulate their use of a sum of observed storm surge
862 and astronomical tides. The aforementioned “modeled surge timeseries” is computed as the
863 modeled storm tide minus the modeled tide-only timeseries, emulating how their observations of
864 surge are computed. This tide plus surge result is 0.25 m higher for the 100-year TC storm tide,
865 is 0.23 m higher for the 100-year ETC storm tide, and is 0.30 m higher for the final combined
866 100-year storm tide. This result suggests that the omission of tide-surge interaction in the

867 Lopeman et al. [2015] paper would lead to a high-bias in their results of about 0.30 m at the
868 Battery, or 10%. If one subtracts this from their result, the result is a 100-year storm tide of 3.08
869 m, only 0.38 m above ours, a much smaller discrepancy.
870
871 Different hydrodynamic models include different processes related to waves, though impacts on
872 water elevation are likely small in deep-water harbor regions like NYH. Our modeling omits
873 wave radiation stress and resulting wave setup, which was recently shown to increase NYH
874 storm tides by a 1.5% on average [*Lin et al.*, 2012]. This low bias may be counteracted in our
875 modeling through other factors, such as a possible high-bias in wind stress (e.g., due to the
876 choice of drag coefficient parameterization). Two-dimensional ADCIRC and SLOSH modeling
877 used in the prior studies, however, neglect the contribution of wave orbital velocities to bed
878 stress [*Grant and Madsen*, 1979], which we include in our three-dimensional modeling. In
879 conclusion, a detailed study of the role these wave-related processes play in the modeling of
880 storm tides is warranted.
881
882 A final way that our approach differs from prior NY region studies is that it leverages an existing
883 ocean forecast system, and thus includes many more subtle details that are often treated as
884 negligible in coastal flood hazard assessments. This includes incorporation of the seasonality of
885 sea level and air temperature, the latter of which impacts air density and therefore wind stress.
886 The former has an impact from -10 to +7 cm, and the latter +/- ~5% on storm tides for individual
887 storms in this region, perhaps most importantly impacting the difference between warm-season
888 TC storm tide and cool-season ETC storm tides. These seasonally-varying variables act in

889 opposition and will to some extent offset one another, but were simple to include in the
890 assessment because of the availability of codes and datasets from the operational forecast system.
891
892 **5.2 The TC storm tide hazard**
893 The results of our TC assessment give a 100-year TC storm tide of 2.64 m (2.45-2.87 95%
894 confidence interval), relative to recent studies results of 2.03 m [*Lin et al.*, 2012] and 2.75 m
895 [*FEMA*, 2014b]. The former study simulated TCs within the NCEP/NCAR reanalysis-based
896 climate conditions of 1981-2000, and does not account for extratropical transition effects, which
897 may account for the lower result. Our study also differs from these two studies in its statistical
898 TC modeling method, as well as its methods for estimating TC wind and pressure fields. The TC
899 meteorology used in our study is simplistic, and we have observed it to give inaccurate winds for
900 Hurricane Sandy, but generally good agreement with observations for a few other recent storms
901 (**Section 4.2**). However, the validation of the model results with historic storm tides
902 demonstrates the results are reasonable.
903
904 One difference in the TC storm climatology is clear in **Figure 2**, because we have left-hand turn
905 storms that were not included in storm sets of the FEMA [2014] study. In our study, even though
906 the westward-directed storms are rare, they do occur and tend to cause the largest storm tides. By
907 contrast, while Lin et al. [2010; 2012] also have storms that head westward in this region, their
908 largest storm tides were caused by storms moving northward. It would be useful to have a
909 detailed inter-comparison study among TC hazard models for the region to elucidate the reasons
910 for these differences.
911

912 Care must be taken with our method of building the TC storm subset, choosing landfall gates that
913 capture all storms which are important for the combined flood hazard at NYH. The gates are
914 chosen to capture all possible TC events with surges over 1.25 m (e.g. **Figure 2a**), capable of
915 causing storm tides exceeding ~2 m. Sensitivity experiments were used to evaluate the
916 importance of passing inland storms, storms that make landfall to the east of Long Island, and
917 storms that never make landfall but pass within $2R_{max}$ of NYH. The combined impact on the
918 100-year storm tide of using three more gates to account for these TCs is an increase of 1 cm, as
919 inland storms are too weak to cause large surges and offshore storms cause a wind direction that
920 is unable to cause a large surge for NYH. Inclusion of this broader set of TCs would moderately
921 increase the 5-year and 10-year TC flood hazard, which tails off below the historically-based
922 estimates at these low return periods (**Figure 13**). However, our final results do not use these
923 TCs, because this region of the curve is not important for the total storm tide hazard (below a 1
924 cm change); ETCs completely dominate the 5-year and 10-year storm tides.
925
926 It is not surprising that our study and the FEMA study, which capped the sea surface drag
927 coefficient (C_D) at 0.0035, found larger TC storm tides than the Lin et al. [2010; 2012] studies,
928 which capped C_D at 0.0025. We found the cap of 0.0035 gave us the best fit to historical events
929 (see **Section 2.3**), but this doesn't mean it is the "correct" drag coefficient cap; it simply gives us
930 the best results in the context of our coupled ocean-wave modeling and simplified wind and
931 pressure fields. Nevertheless, TC assessments might collapse toward closer agreement if the
932 same drag coefficient methods were utilized. This underscores the importance of coupled ocean-
933 atmosphere validation and possible model tuning using historical events, as was advocated for by
934 Cardone and Cox [2009].

935

936 **5.3 The ETC storm tide hazard**

937 The results of our ETC assessment give a 100-year ETC storm tide of 2.38 m (2.07-2.75 95%
938 confidence range), relative to the FEMA study result of 3.23 m [*FEMA*, 2014b] and our
939 observation-based (GPD) assessment of 2.22 m (1.95-2.72). Our model-based and observation-
940 based estimates are in general agreement, within uncertainty, and both are substantially lower
941 than the FEMA result. Our observation-based estimate is lower than our model-based estimate
942 because the largest historical ETC storm surge (19501125) peaked at low tide, giving a storm
943 tide of 2.12 m. The modeled 19501125 ETC events occur at 50 random tide phases, giving storm
944 tides from 2.07 to 2.71 m, accounting for a wider range of possible floods.

945

946 The difference in ETC results between our study and the FEMA study are the main reason for
947 the differences between the studies' combined flood exceedance curves, considering that the
948 estimates of the 100-year TC storm tide are nearly the same. Our study differs from that study in
949 the wind speed factors used, discussed below, and in our scaling up of 11 storm annual rates to
950 make up for specific missing top-30 storm surge events at NYH (**Section 2.2**). The impact of
951 scaling up storm rates for the missing storms is an increase from 1.79 to 1.86 m for the 10-year
952 storm tide, and a small decrease from 2.29 to 2.27 m for the 100-year storm tide. A goal here is
953 to accurately capture storm tides at short return periods down to 5 y, so the scale-ups appear
954 useful.

955

956 The FEMA [2014b] study multiplied the reanalysis wind speeds for all storms by 1.04, whereas
957 we keep the winds unmodified except for the 1950 storm, where we lower them by 17% to

958 accurately fit the hindcast event (**Section 4.2**). We suspect the need for downscaling of wind
959 speed for this one storm either results from errors in the storm's wind speed estimates due to the
960 absence of measurements over the ocean for that storm, or from error in our estimated drag
961 coefficient ceiling value of 0.0035. On the latter possibility, this was a rare "southeaster" storm
962 where winds were blowing steadily at 25-30 m s^{-1} over a large fetch (**Figure 4**) and in the same
963 direction that the wave field was traveling (toward the west-northwest); recent work has shown
964 that the drag coefficient cap can be much lower for this scenario [*Holthuijsen et al.*, 2012]. Our
965 wave modeling results in New York Bight show a mean wave-wind misalignment of only 5
966 degrees for the 1950 ETC, blowing in the direction of wave propagation, in comparison to 95
967 degrees for Hurricane Sandy (average for 28-29 October 2012), blowing along the wave crests.
968
969 The FEMA study storm set was completed in 2011 and only considered ETCs from 1950-2009,
970 but our study has the benefit of a wealth of data that has become available since Hurricane Sandy
971 in 2012 [e.g., *Talke et al.*, 2014]. We can look at a broader time period to evaluate our result, and
972 there appear to have been no ETC storm tides over 2.2 m at least back to 1821, indicating that
973 their result of 3.23 m is very likely too high. Moreover, their study simulated only two random
974 tide scenarios, but our 50 random tide scenario runs for the most intense storms shows ETC
975 storm tides never reaching above 2.7 m.
976
977 Our designation of hybrid storms that begin like Hurricane Sandy as TCs is important here, as
978 moving these storms into the ETC category could change our interpretation of the historical data
979 and our comparison to our model-based results. However, it makes the most sense to divide them
980 in this way because ETCs of non-tropical cyclone origin in this coastal region have historically

never shown a peak sustained wind speed more than 35 m s^{-1} at 10 m above sea level [e.g., *Cardone et al.*, 1996; *Dolan and Davis*, 1992], whereas the HURDAT database [*Landsea et al.*, 2004] reveals that North Atlantic TCs have historically had a top wind speed of 85 m s^{-1}. The dynamical differences between TCs/hybrids and ETCs further justifies separation into two statistical distributions and extreme value statistical analyses. Because of the rare, but possible, chance of such extreme wind speeds, TC and hybrid storm surges exhibit a "long-tailed" probability distribution, and the worst-case flood scenario for a TC/hybrid is much larger than for an ETC.

5.4 Impacts of the 1788 and 1821 hurricanes at New York City

Understanding what occurred during the 1788 and 1821 hurricanes is highly valuable because they are the two largest known prior storm tide events prior to Sandy and also exhibited much stronger wind speeds that would cause severe damage and airborne debris. Our parametric modeling showed excellent agreement with our archival estimates for the 1788 event, but underestimated the 1821 estimates by 0.34 m (**Figure 9**). The mixed results are unsurprising, considering that the meteorology for both storms consists of simplified wind and pressure fields based on land-damage from wind [Boose et al. 2001], and is therefore only approximate.

Our archival research-based 1821 storm tide estimate of approximately 3.0 m (and storm surge of 3.4 m) is a downgrade by 0.2 m from the widely-cited Scileppi and Donnelly [2007] paper. Our model result for the storm tide, however, is only 2.66 m. One possible explanation for this is inaccurate wind speed estimates; a prior recent study using SLOSH found a higher wind speed is needed, relative to the Boose et al. [2001] study estimate of 51.4 m s^{-1} at landfall, to fit a

1004 presumed (inaccurate by +0.2 m) storm surge of 3.6 m, and the best fit came when maximum
1005 sustained wind speeds of 58 m s^{-1} [*Brandon et al.*, 2014]. These results suggest that the Boose et
1006 al. [2001] wind speeds at sea for that storm may be too low. In a separate paper on nature-based
1007 flood protection, we used a peak wind speed of 58 m s^{-1}, and obtained a storm tide peak of 2.95
1008 m at NYH, much closer to the estimated observed value [*Orton et al.*, 2015b].
1009
1010 One widely-cited [*Brandon et al.*, 2014; *Coch*, 1994; *Redfield*, 1831; *Scileppi and Donnelly*,
1011 2007] and impressive factor with the 1821 storm tide in NYH was that the water rose 13 feet in
1012 one hour, based on a newspaper quote. The *New Bedford Mercury* quote read "In one hour
1013 during the [hurricane] of Monday evening, the water was forced into the East River 13 feet 4
1014 inches [4.0 m] above low water mark." [*Boose et al.*, 2003]. The quote is important because such
1015 a rapid rise in one hour is extreme and would require high water velocities and likely lead to
1016 substantial erosion and possible undermining of infrastructure. The maximum water rise rate in
1017 NYH during Hurricane Sandy, for contrast, was only 0.7 m hr^{-1}.
1018
1019 Our historical reassessment shows that the 4 m h^{-1} rise rate for the 1821 TC is unlikely to have
1020 occurred, in part because the water only rose 3.4 m total, but also because our modeling suggests
1021 this rise rate was improbable. Our most accurate 1821 hurricane storm tide simulation [*Orton et*
1022 *al.*, 2015b] shows a maximum rise rate of 1.8 m h^{-1} for the storm tide, and our simulation of the
1023 smaller and faster 1788 hurricane shows a faster rise rate of 2.4 m h^{-1}. More broadly, the fastest
1024 water rise rate for 2.5-3.5 m TC storm tide events in our modeling is 2.4 m h^{-1} and the average is
1025 0.8 m h^{-1}.
1026

A recent study of lagoonal sediment cores at New York City's Staten Island showed that the 1788 event had a much thinner and finer sand layer than the 1821 event [*Brandon et al.*, 2014]. This same observation was made by Scileppi and Donnelly [2007], and they surmised that the smaller storm size explained the less prominent and sometimes missing sand overwash layer on Long Island. The maximum storm surge at NYH for the 1788 hurricane was only slightly smaller (2.9 m) than the 1821 storm tide (3.0 m), but the storm was of a smaller size (wind fetch), faster speed, and thus a shorter duration [*Ludlum*, 1963], likely resulting in smaller waves. Our model results show peak significant wave heights and wave periods in New York Bight's apex (NDBC Buoy 44065) for 1788 of 9.4 m with period 14.5 s, whereas they are 13.5 m and 18.1 s for 1821 (versus observed values of 9.9 m and 16.0 s for Hurricane Sandy). These factors likely explain why the 1788 sediment layers were smaller, with finer sediments.

The 1788 hurricane storm tide estimate is lowered by 0.1 m relative to the recent Scileppi and Donnelly [2007] estimate, and uncertainty is also reduced due to the fact that multiple sources of the estimate now exist and are in approximate agreement – (1) the elevation over street level at Front and Water streets of lower Manhattan, because most cellars ended up filled with water, (2) a quantitative comparison of different storm tides at one tidal mill location (**Section 3.2**), and (3) our model result of 3.0 m. These all support our estimate of a storm tide of 2.9 m, at 95% confidence.

As recently noted by Swiss Re [2014], the impacts of these storms went beyond flooding, as they were Category-3 hurricanes that passed over NYC. Sustained maximum wind speeds at elevation over the city were roughly 50 m s^{-1} (not at 10 m level, which is more complicated and landscape-

1050 dependent). If a hurricane of this intensity were to strike the region today, damages would likely
1051 be even greater than those from Sandy and would come from both winds and flooding [*Swiss Re*,
1052 2014]. Our data suggest that the likelihood of a Category-3 or stronger hurricane passing near
1053 NYC (over landfall gate 1, 2 or 3) and causing a 3 m storm tide or higher is ~0.003 y^{-1}. Similarly,
1054 the probability of any combination of wind speed and flood level can be analyzed using the data
1055 produced in our assessment.
1056
1057
1058 **6.0 Summary and Conclusions**
1059 In this paper, we have reported on methods and results for a model-based flood hazard
1060 assessment we have conducted for New York Harbor, a region where prior assessments have
1061 differed widely. The methods are novel and provide a new and important contribution to the
1062 science of coastal flood hazard assessment. The assessment separately includes the contribution
1063 of TCs (including hybrids) and ETCs to the flood hazard, separating them because they represent
1064 storm tide events with different physical limits controlled by wind speed. The TC storm set
1065 includes results from a statistical TC model that has diverse storms including ones that curve
1066 westward like Hurricane Sandy. Resulting flood exceedance curves for any type of storm tide
1067 (TC and ETC) show 10-, 100-, 1000-, and 10000-year return period events of 1.94, 2.70, 4.35,
1068 and 5.77 m, respectively.
1069
1070 Our study is the most deeply validated study of storm tides for the region performed to date, and
1071 lays out a framework for conducting similar studies for other areas. The storm sets are
1072 demonstrated to be consistent with historical data, and our coupled ocean-atmosphere modeling

1073 is accurate when validated against 42 historical events. The resulting TC and ETC flood
1074 exceedance curve results are consistent with an observation-based assessment for recent decades,
1075 as well as the empirical data points that include the worst events in the 1700s and 1800s that
1076 came prior to the tide gauge era.
1077
1078 A reassessment of the 1821 and 1788 hurricane storm tides at NYH was also performed, due to
1079 the fact that these are believed to have been Category-3 hurricanes with similar storm tides to
1080 Hurricane Sandy. The 1821 storm tide is estimated to be 3.0 m (2.5- 3.3 at 95% confidence), 0.2
1081 m lower than a prior recent estimate [*Scileppi and Donnelly*, 2007]. The 1788 storm tide is
1082 estimated to be 2.9 m (2.3-3.3). The downgrade and reductions in uncertainty allow for the
1083 conclusion that the storm tides were lower than Sandy, and a resulting conclusion of this work is
1084 that Hurricane Sandy's storm tide at NYH was the largest at least as far back as 1700.
1085
1086 More research is needed to ensure we can both account for the large variety of storms, and also
1087 accurately capture the atmospheric and oceanic physics of these extreme storms, to reduce this
1088 uncertainty and the differences between studies. Hybrids are a rare but highly impactful type of
1089 storm which needs further study, as do extratropical cyclones and their sources of energy that can
1090 lead to hurricane strength winds (e.g. latent heat). The coupling between atmosphere and ocean
1091 through wind stress and sea surface drag coefficient is a particularly important area requiring
1092 study for New York Bight, where winds and waves are often misaligned during storms due to the
1093 coastal geometry.
1094

1095 The assessment framework is useful for other purposes beyond estimation of the present-day
1096 flood hazard, including sea level rise and climate change impacts on flooding. The framework
1097 and models are also useful for performing adaptation quantification studies where the model is
1098 manipulated to emulate structural or natural flood protection efforts that are happening rapidly
1099 after Hurricane Sandy. Funded projects are currently addressing both needs, using the data for
1100 studies of the Hudson River and Jamaica Bay, New York City, and will be summarized in future
1101 publications.
1102
1103
1104 **Acknowledgements**
1105 Work by PO, AB, NG and SV was funded by the NASA Centers call for support of the National
1106 Climate Assessment (Hall, PI; Agreements NNX12AI28G and NNX15AD61G), NASA's
1107 Research Opportunities in Space and Earth Science ROSES-2012 (grant NNX14AD48G), and
1108 NOAA's Regional Integrated Sciences and Assessments (RISA) program (Award
1109 NA10OAR4310212). Work by TH was also funded by the NASA projects listed above. Work by
1110 ST was funded by the U.S. Army Corps of Engineers (award W1927N-14-2-0015).
1111 Supercomputer resources were utilized under a grant of computer time from the City University
1112 of New York High Performance Computing Center under NSF Grants CNS-0855217, CNS-
1113 0958379 and ACI-1126113. The model data, compiled NOAA water elevation observations, and
1114 Matlab codes necessary to reproduce the study are available at
1115 http://personal.stevens.edu/~porton/Orton_etal_JGR16.zip.
1116
1117

References

Aerts, J., W. W. Botzen, K. Emanuel, N. Lin, H. de Moel, and E. O. Michel-Kerjan (2014), Evaluating Flood Resilience Strategies for Coastal Megacities, *Science*, *344*(6183), 473-475.

Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven (2013), Tropical Cyclone Report: Hurricane Sandy (AL182012), National Hurricane Cent., Miami, Fla.

Blumberg, A. F., L. A. Khan, and J. St John (1999), Three-dimensional hydrodynamic model of New York Harbor region, *J. Hydraul. Engin.*, *125*(8), 799-816.

Boose, E. R., K. E. Chamberlin, and D. R. Foster (2001), Landscape and regional impacts of hurricanes in New England, *Ecological Monographs*, *71*(1), 27-48.

Boose, E. R., D. R. Foster, and K. E. Chamberlin (2003), Harvard Forest Data Archive, HF011: Landscape and Regional Impacts of Hurricanes in New England 1620-1997, http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf011, accessed March, 2013.

Botts, H., W. Du, T. Jeffery, S. Kolk, Z. Pennycook, and L. Suhr (2013), CoreLogic Storm Surge Report 2013, 44 pp, CoreLogic, Irvine, CA.

Brandon, C. M., J. D. Woodruff, J. P. Donnelly, and R. M. Sullivan (2014), How unique was Hurricane Sandy? Sedimentary reconstructions of extreme flooding from New York Harbor, *Scientific reports*, *4*.

Bretschneider, C. (1972), A non-dimensional stationary hurricane wave model, *Proceedings of the Offshore Technology Conference, Vol 1*, 51-68.

Cardone, V., R. Jensen, D. Resio, V. Swail, and A. Cox (1996), Evaluation of contemporary ocean wave models in rare extreme events: the "Halloween Storm" of October 1991 and the "Storm of the Century" of March 1993, *J. Atmos. Oceanic Technol.*, *13*(1), 198-230.

1141 Cardone, V., and A. Cox (2009), Tropical cyclone wind field forcing for surge models: critical
1142 issues and sensitivities, *Natural hazards*, *51*(1), 29-47.
1143 City of New York (2013), A stronger, More Resilient New York, 445 pp, New York, NY.
1144 Coch, N. K. (1994), Hurricane hazards along the northeastern Atlantic Coast of the United
1145 States, *J. Coast. Res.*, 115-147.
1146 Colle, B. A., F. Buonaiuto, M. J. Bowman, R. E. Wilson, R. Flood, R. Hunter, A. Mintz, and D.
1147 Hill (2008), New York City's vulnerability to coastal flooding, *Bull. Amer. Meteorol. Soc.*,
1148 *89*(6), 829-841.
1149 Colle, B. A., K. Rojowsky, and F. Buonaito (2010), New York City Storm Surges: Climatology
1150 and an Analysis of the Wind and Cyclone Evolution, *Journal of Applied Meteorology and*
1151 *Climatology*, *49*(1), 85-100.
1152 Colle, B. A., M. J. Bowman, K. J. Roberts, M. H. Bowman, C. N. Flagg, J. Kuang, Y. Weng, E.
1153 B. Munsell, and F. Zhang (2015), Exploring Water Level Sensitivity for Metropolitan New
1154 York during Sandy (2012) Using Ensemble Storm Surge Simulations, *Journal of Marine*
1155 *Science and Engineering*, *3*(2), 428-443.
1156 Condon, A. J., and Y. P. Sheng (2012), Optimal storm generation for evaluation of the storm
1157 surge inundation threat, *Ocean Engin.*, *43*, 13-22.
1158 Cox, A., J. Greenwood, V. Cardone, and V. Swail (1995), An interactive objective kinematic
1159 analysis system, paper presented at Fourth International Workshop on Wave Hindcasting and
1160 Forecasting.
1161 Demuth, J. L., M. DeMaria, and J. A. Knaff (2006), Improvement of advanced microwave
1162 sounding unit tropical cyclone intensity and size estimation algorithms, *Journal of Applied*
1163 *Meteorology and Climatology*, *45*(11), 1573-1581.

1164 Dolan, R., and R. E. Davis (1992), An intensity scale for Atlantic coast northeast storms, *J.*
1165 *Coast. Res.*, 840-853.
1166 Donelan, M. (1977), A simple numerical model for wave and wind stress prediction, *National*
1167 *Water Research Institute, Burlington, Ontario, 28pp*.
1168 Donnelly, J. P., S. Roll, M. Wengren, J. Butler, R. Lederer, and T. Webb (2001), Sedimentary
1169 evidence of intense hurricane strikes from New Jersey, *Geology*, *29*(7), 615-618.
1170 Elsner, J. B. (2006), Historical Hurricane Information Tool (HHIT), Available online, accessed
1171 January 19, 2013.
1172 Emanuel, K. (2000), A statistical analysis of tropical cyclone intensity, *Mon. Weather. Rev.*,
1173 *128*(4), 1139-1152.
1174 Espy (1839), Facts collected by Mr. Espy, taken from Newspapers of the time, *Journal of the*
1175 *Franklin Institute 23*, 156-158.
1176 FEMA (2014a), Region II Storm Surge Project - Development of Wind and Pressure Forcing in
1177 Tropical and Extra-Tropical Storms 62 pp, Federal Emergency Management Agency,
1178 Washington DC.
1179 FEMA (2014b), Region II Coastal Storm Surge Study: Overview, 15 pp, Federal Emergency
1180 Management Agency, Washington, DC.
1181 FEMA (2014c), Region II Storm Surge Project - Joint probability analysis of hurricane and
1182 extratropical flood hazards, 95 pp, Federal Emergency Management Agency, Washington,
1183 DC.
1184 Galarneau, T. J., C. A. Davis, and M. A. Shapiro (2013), Intensification of Hurricane Sandy
1185 (2012) through extratropical warm core seclusion, *Mon. Weather. Rev.*, *141*(12), 4296-4321.

1186 Garratt, J. (1977), Review of drag coefficients over oceans and continents, *Mon. Weather. Rev.*,
1187 *105*(7), 915-929.
1188 Georgas, N., A. Blumberg, and T. Herrington (2007), An operational coastal wave forecasting
1189 model for New Jersey and Long Island waters, *Shore Beach*, *75*(2), 30-35.
1190 Georgas, N., and A. F. Blumberg (2009), Establishing Confidence in Marine Forecast Systems:
1191 The Design and Skill Assessment of the New York Harbor Observation and Prediction
1192 System, Version 3 (NYHOPS v3), paper presented at Eleventh International Conference in
1193 Estuarine and Coastal Modeling (ECM11), ASCE, Seattle, Washington, USA, 4-6
1194 November.
1195 Georgas, N. (2010), Establishing confidence in marine forecast systems: The design of a high
1196 fidelity marine forecast model for the NY/NJ harbor estuary and its adjoining coastal waters,
1197 272 pp, Stevens Institute of Technology.
1198 Georgas, N., P. Orton, A. Blumberg, L. Cohen, D. Zarrilli, and L. Yin (2014), The impact of
1199 tidal phase on Hurricane Sandy's flooding around New York City and Long Island Sound,
1200 *Journal of Extreme Events*, DOI: 10.1142/S2345737614500067.
1201 Grant, W. D., and O. S. Madsen (1979), Combined wave and current interaction with a rough
1202 bottom, *J. Geophys. Res.*, *84*(C4), 1797-1808.
1203 Hall, T., and S. Jewson (2007), Statistical modelling of North Atlantic tropical cyclone tracks,
1204 *Tellus A*, *59*(4), 486-498.
1205 Hall, T., and E. Yonekura (2013), North American tropical cyclone landfall and SST: A
1206 statistical model study, *J. Clim.*, *26*(21), 8422-8439.
1207 Hall, T. M., and A. H. Sobel (2013), On the impact angle of Hurricane Sandy's New Jersey
1208 landfall, *Geophys. Res. Lett.*, *40*(10), 2312-2315.

1209 Harper, B., J. Kepert, and J. Ginger (2010), Guidelines for converting between various wind
1210 averaging periods in tropical cyclone conditions, *World Meteor. Org., TCP Sub-Project*
1211 *Report, WMO/TD*(1555).
1212 Holthuijsen, L. H., M. D. Powell, and J. D. Pietrzak (2012), Wind and waves in extreme
1213 hurricanes, *J. Geophys. Res.*, *117*(C9).
1214 Horsburgh, K., and C. Wilson (2007), Tide-surge interaction and its role in the distribution of
1215 surge residuals in the North Sea, *J. Geophys. Res.*, *112*(C8).
1216 Jelesnianski, C., J. Chen, W. Shaffer, U. S. N. Oceanic, A. Administration, and U. S. N. W.
1217 Service (1992), *SLOSH: Sea, lake, and overland surges from hurricanes*, US Dept. of
1218 Commerce, National Oceanic and Atmospheric Administration, National Weather Service,
1219 Silver Spring, MD, USA.
1220 Kemp, A. C., and B. P. Horton (2013), Contribution of relative sea-level rise to historical
1221 hurricane flooding in New York City, *Journal of Quaternary Science*, *28*(6), 537-541.
1222 Kussman, A. S. (1957), Report on the Hurricane of September 3, 1821, U.S. Weather Bureau,
1223 New York, NY, USA, 54 pp.
1224 Landsea, C. W., C. Anderson, N. Charles, G. Clark, J. Dunion, J. Fernandez-Partagas, P.
1225 Hungerford, C. Neumann, and M. Zimmer (2004), The Atlantic hurricane database re-
1226 analysis project: Documentation for the 1851-1910 alterations and additions to the HURDAT
1227 database, *Hurricanes and Typhoons: Past, Present and Future*, 177-221.
1228 Large, W. G., and S. Pond (1981), Open ocean momentum flux measurements in moderate to
1229 strong winds, *J. Phys. Oceanogr.*, *11*, 324-336.
1230 Lin, N., K. Emanuel, J. Smith, and E. Vanmarcke (2010), Risk assessment of hurricane storm
1231 surge for New York City, *J. Geophys. Res*, *115*.

1232 Lin, N., and D. Chavas (2012), On hurricane parametric wind and applications in storm surge
1233 modeling, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *117*(D9).
1234 Lin, N., K. Emanuel, M. Oppenheimer, and E. Vanmarcke (2012), Physically based assessment
1235 of hurricane surge threat under climate change, *Nature Climate Change*, *2*(6), 462-467.
1236 Lin, N., P. Lane, K. A. Emanuel, R. M. Sullivan, and J. P. Donnelly (2014), Heightened
1237 hurricane surge risk in northwest Florida revealed from climatological-hydrodynamic
1238 modeling and paleorecord reconstruction, *Journal of Geophysical Research: Atmospheres*,
1239 *119*(14), 8606-8623.
1240 Lopeman, M., G. Deodatis, and G. Franco (2015), Extreme storm surge hazard estimation in
1241 lower Manhattan, *Natural Hazards*, 1-37.
1242 Ludlum, D. M. W. (1963), *Early American Hurricanes, 1492-1870*, American Meteorological
1243 Society, Boston.
1244 Mukai, A., J. Westerink, R. Luettich Jr, and D. Mark (2002), Eastcoast 2001, a tidal constituent
1245 database for western North Atlantic, Gulf of Mexico, and Caribbean Sea, DTIC Document.
1246 Niedoroda, A., D. Resio, G. Toro, D. Divoky, H. Das, and C. Reed (2010), Analysis of the
1247 coastal Mississippi storm surge hazard, *Ocean Engin.*, *37*(1), 82-90.
1248 NOAA-NHC-HRD (2015), Atlantic hurricane best track (HURDAT), United States National
1249 Oceanic and Atmospheric Administration, National Hurricane Center, Hurricane Research
1250 Division, accessed December 2015.
1251 NOAA (2015), Tides and Currents, National Oceanographic and Atmospheric Administration,
1252 Available online: https://tidesandcurrents.noaa.gov/, accessed, January, 2015.
1253 Orff, K., G. Wirth, P. Orton, and et al. (2014), Living Breakwaters: Volume II Staten Island and
1254 Raritan Bay, 118 pp, Housing and Urban Development, Rebuild By Design.

1255 Orton, P., N. Georgas, A. Blumberg, and J. Pullen (2012), Detailed modeling of recent severe
1256 storm tides in estuaries of the New York City region, *J. Geophys. Res.*, *117*, C09030, DOI:
1257 10.1029/2012JC008220.
1258 Orton, P., S. Vinogradov, N. Georgas, A. Blumberg, N. Lin, V. Gornitz, C. Little, K. Jacob, and
1259 R. Horton (2015a), New York City Panel on Climate Change 2015 report chapter 4: Dynamic
1260 coastal flood modeling, *Ann. N. Y. Acad. Sci.*, *1336*(1), 56-66.
1261 Orton, P. M., S. A. Talke, D. A. Jay, L. Yin, A. F. Blumberg, N. Georgas, H. Zhao, H. J.
1262 Roberts, and K. MacManus (2015b), Channel Shallowing as Mitigation of Coastal Flooding,
1263 *Journal of Marine Science and Engineering*, *3*(3), 654-673, DOI: 10.3390/jmse3030654.
1264 Phadke, A. C., C. D. Martino, K. F. Cheung, and S. H. Houston (2003), Modeling of tropical
1265 cyclone winds and waves for emergency management, *Ocean Engin.*, *30*(4), 553-578.
1266 Pickands, J. (1975), Statistical inference using extreme order statistics, *the Annals of Statistics*,
1267 119-131.
1268 Powell, M. D., P. J. Vickery, and T. A. Reinhold (2003), Reduced drag coefficient for high wind
1269 speeds in tropical cyclones, *Nature*, *422*(6929), 279-283.
1270 Pugh, D. T. (1987), *Tides, surges and mean sea-level*, John Wiley & Sons Ltd.
1271 Redfield, W. (1831), Remarks on the prevailing storms of the Atlantic coast of the North
1272 American States, *Amer. J. of Science and Arts 20*, 17–51.
1273 Reichl, B. G., T. Hara, and I. Ginis (2014), Sea state dependence of the wind stress over the
1274 ocean under hurricane winds, *J. Geophys. Res.*, *119*(1), 30-51.
1275 Resio, D. T., and J. J. Westerink (2008), Modeling the physics of storm surges, *Phys. Today*, *61*,
1276 33-38.

1277 Scileppi, E., and J. P. Donnelly (2007), Sedimentary evidence of hurricane strikes in western
1278 Long Island, New York, *Geochem. Geophys., Geosyst.*, *8*(6), DOI: 10.1029/2006GC001463.
1279 Shalowitz, A. L. (1964), Shore and Sea Boundaries, with Special Reference to the Interpretation
1280 and use of Coast and Geodetic Survey Data, Vol. 2: Interpretation and Use of Coast and
1281 Geodetic Survey Data, US Government Printing Office, Washington DC.
1282 Signell, R., R. Beardsley, H. Graber, and A. Capotondi (1990), Effect of wave-current interaction
1283 on wind-driven circulation in narrow, shallow embayments, *J. Geophys. Res.*, *95*(C6), 9671-
1284 9678.
1285 Sweet, W., C. Zervas, S. Gill, and J. Park (2013), Section 6: Hurricane Sandy Inundation
1286 Probabilities Today and Tomorrow, in *Explaining Extreme Events of 2012 from a Climate*
1287 *Perspective: Special supplement to the Bulletin of the American Meteorological Society vol.*
1288 *94, no. 9, September 2013*, edited by T. C. Peterson, et al., pp. S17-S20.
1289 Swiss Re (2014), The big one: The East Coast's USD 100 billion hurricane event, 21 pp, Swiss
1290 Re America Holding Corporation, Armonk, NY, USA.
1291 Talke, S., P. Orton, and D. Jay (2014), Increasing storm tides at New York City, 1844-2013,
1292 *Geophys. Res. Lett.*, *41*, DOI: 10.1002/2014GL059574.
1293 Tang, Y. M., B. Sanderson, G. Holland, and R. Grimshaw (1996), A numerical study of storm
1294 surges and tides, with application to the North Queensland coast, *J. Phys. Oceanogr.*, *26*(12),
1295 2700-2711.
1296 Taylor, P. K., and M. J. Yelland (2001), The dependence of sea surface roughness on the height
1297 and steepness of the waves, *J. Phys. Oceanogr.*, *31*(2), 572-590.
1298 Toro, G., A. Niedoroda, C. Reed, and D. Divoky (2010), Quadrature-based approach for the
1299 efficient evaluation of surge hazard, *Ocean Engin.*, *37*(1), 114-124.

1300 USACE (2015), North Atlantic Coast Comprehensive Study: Resilient Adaptation to Increasing
1301 Risk: Main Report, 140 pp, United States Army Corps of Engineers.
1302 Vecchi, G. A., and T. R. Knutson (2011), Estimating annual numbers of Atlantic hurricanes
1303 missing from the HURDAT database (1878-1965) using ship track density, *J. Clim.*, *24*(6),
1304 1736-1746.
1305 Vickery, P., P. Skerlj, and L. Twisdale (2000), Simulation of hurricane risk in the US using
1306 empirical track model, *Journal of Structural Engineering*, *126*(10), 1222-1237.
1307 Wang, H. V., J. D. Loftis, Z. Liu, D. Forrest, and J. Zhang (2014), The storm surge and sub-grid
1308 inundation modeling in New York City during Hurricane Sandy, *Journal of Marine Science*
1309 *and Engineering*, *2*(1), 226-246.
1310 Zervas, C. (2013), Extreme Water Levels of the United States 1893-2010, NOAA Technical
1311 Report NOS CO-OPS 067, NOAA National Ocean Service Center for Operational
1312 Oceanographic Products and Services, 200 pp, Silver Spring, Maryland.

1313

1314

1315

Figure Captions

Figure 1: The top 20 gauged historical storm tides from 1844-2013 at the Battery (New York Harbor) and the two worst prior events back to 1700 (1788, 1821) were a mixture of extratropical cyclones (ETCs) and tropical cyclones (TCs). Data are water elevation above the year's mean sea level, and 1844-2013 data are based on an updated version of the Talke et al. [2014] data set. Our estimated 1788 and 1821 storm tides have 95% confidence bars, and are discussed in **Section 3**.

Figure 2: Historical and modeled TC storm tracks affecting New York Harbor (NYH), shaded by Saffir-Simpson category (Cat). (**top**) Known TC or hybrid TC/ETC 1.25 m or greater surge events at NY Harbor, 1788-2013, with year, name (if any), and observed storm surge. (**bottom left**) Select modeled TC tracks, on a map that includes the landfall gate numbers. Storms are shown that led to storm tides close to the 100-year event (2.5-2.9m) which occur at a rate higher than 0.0001 y^{-1}. (**bottom right**) Storms that led to the largest storm tides in the set, from 5.2-5.7 m.

Figure 3: Estimated 1788 hurricane meteorology and modeled storm tide (elevation) shading and timeseries at NYH (NYHOPS model domain), for two snapshots in time. Wind velocity is shown as vectors (see 33 m/s key at bottom right) and pressure isobars are shown as white lines. The red vertical lines in the timeseries show the times represented in the maps. See **Section 3.1** for a summary of the storm's history and data sources.

Figure 4: The 1950 ETC meteorological reanalysis and modeled storm surge on the SNAP grid, near the time of peak surge. Wind velocity is shown as vectors, and pressure isobars as black lines. All storms in the study were simulated with only storm surge on the SNAP grid and then on the nested NYHOPS grid with tides.

Figure 5: Characterization of spread in modeled TC storm tides within a storm flavor. Three GEV-fitted distributions are shown that cover a range of storm intensities, each including 200 cases of randomly-chosen storms and tides. These curves are later used for TC rate-distribution smoothing.

Figure 6: Annual rates (**top**) and modeled storm tide (**bottom**) for each of the 606 synthetic TCs. Colors indicate Saffir-Simpson category, and landfall gates are also shown (0 for southern New Jersey, moving clockwise to 4 for eastern Long Island; shown in **Figure 2**). The bottom panel also shows storm size, an important determinant of storm tide.

Figure 7: Historical evaluation of the statistical TC model: Shown are landfall wind speed exceedance curves for the greater New York Bight region (across all 5 landfall gates). Yellow curves are shown for each of the 15625 simulations of the 64-yr (1950–2013) period. The orange curves indicate the lower and upper bounds of the inner 95% across the simulations. The red curve is obtained by combining the simulations in series (i.e., a 1000000-yr simulation). The blue curve is obtained directly from HURDAT landfalls.

Figure 8: Historical evaluation of the ETC coupled ocean-atmosphere model results for the Battery tide gauge (NYH), including mean bias and RMSE. Numbering of data points (storm events) is given in **Table 1**.

Figure 9: Historical evaluation of the TC coupled ocean-atmosphere modeling results for the Battery tide gauge (NYH), including bias and RMSE. Storms numbers are shown in **Table 2**. Atmospheric forcing is based on simple parametric wind and pressure fields, not detailed data.

Figure 10: Annual rate distribution for modeled ETC storm tides

Figure 11: Annual rate distribution for modeled TCs (blue), also with the distribution as smoothed to account for the random tide and storm variations (red). The inset panel shows a logarithmic y-axis, to show the data in the long tail.

Figure 12: Results -- ETC flood exceedance curve versus the 1950-2013 observation-based Generalized Pareto Distribution (GPD) assessment, both with shaded 95% confidence ranges. Open circles are empirically-estimated return levels for specific events, with the highest point (the nor'easter on 1992/12/11) appearing at a return period of 64 years (the record length).

Figure 13: Results -- TC flood exceedance curve versus a 1950-2013 observation-based GPD assessment, both with shaded 95% confidence ranges. Empirical points ("emp") from 1950-2013 and 1700-2011 are also plotted. The latter shows empirical return periods for the 1788 and 1821 hurricanes, computed using the longer period of 314 years (up to 2013) because we know these were the second and third highest events since 1700.

Figure 14: Model-based and observation-based flood exceedance curves for NYH, combining TC and ETC curves from the prior two figures. The 95% confidence intervals are shown for the combined observed curve (red dotted lines) and modeled curve (grey shading). Storm tide is water level above mean sea level (MSL), and the model-based 100-year storm tide is 2.70 m, or 2.64 m NAVD88 when added on MSL1983-2001.

Table 1: ETC storm set, including the number of tide repetitions (**Section 2.4**), and in cases where a storm's annual rate was doubled, the missing storm that is accounted for (doubling storm).

#	Code yyyymmdd	Year	Month	Day of peak	Battery surge[a] (m)	Surge rank[b]	Tide reps	Doubling storm
1	19501125	1950	11	25	2.41	1	50	
2	19610413	1961	4	13	1.15	18	50	19551014
3	19620306	1962	3	6	1.26	13	50	19610204
4	19640112	1964	1	12	0.99	40	1	
5	19660123	1966	1	23	1.07	27	50	19660130
6	19681112	1968	11	12	1.49	6	50	19801025
7	19701217	1970	12	17	0.96	41	1	
8	19710208	1971	2	8	1.17	17	50	19851105
9	19711125	1971	11	25	1.13	21	50	19780120
10	19720219	1972	2	19	1.06	30	50	19721108
11	19741202	1974	12	2	1.7	3	50	
12	19790125	1979	1	25	1.3	12	50	
13	19840329	1984	3	29	1.57	4	50	19531107
14	19870123	1987	1	23	1.07	29	50	19940104
15	19911031	1991	10	31	1.41	8	50	
16	19921211	1992	12	11	1.83	2	50	
17	19930314	1993	3	14	1.39	9	50	
18	19940303	1994	3	3	1.15	19	50	19960320
19	19941224	1994	12	24	0.93	50	1	
20	19950204	1995	2	4	0.8	103	1	
21	19951115	1995	11	15	1.32	10	50	
22	19960108	1996	1	8	1.31	11	50	
23	19961020	1996	10	20	1	39	1	
24	19961206	1996	12	6	1.04	32	1	
25	19980128	1998	1	28	0.76	123	1	
26	19980205	1998	2	5	1.03	34	1	
27	20051025	2005	10	25	1.1	25	50	19600219
28	20070416	2007	4	16	0.98	42	1	
29	20080512	2008	5	12	0.84	80	1	
30	20091113	2009	11	13	0.9	66	1	

[a] Surge is the peak water elevation minus tide, minus annual mean sea level

[b] ETC surge rankings are for the period 1950-2013, the period of the reanalysis datasets

Table 2: Historical tropical cyclones used for model validation at NYH

#	year	name	month	day	R_{max}[g]	V_{max}[ag]	$P_{central}$[g]	V_{storm}[g]	bearing	max surge[h]	max storm tide[h]
					km	m/s	mbar	m/s	degrees	m	m
1	1788	unnamed	8	19	29	51.4	952[c]	17.6	16	3.0	2.9
2	1821	unnamed	9	3	50[b]	51.4	952[c]	11.1	23	3.4	3.0
3	1893	unnamed	8	24	50[b]	41.2	968[c]	10.4	9	1.2	1.6
4	1938	unnamed	9	21	55	45.3	940	22.6	359	1.22	1.68
5	1944	unnamed	9	15	51	46.3	955	13.1	30	1.18	1.76
6	1954	Carol	8	31	48	43.4	963	15.6	18	0.89	1.29
7	1960	Donna	9	12	101	43.5	961	17.5	25	1.69	2.33
8	1976	Belle	8	10	49	42.3	975	11.6	12	1.21	1.29
9	1985	Gloria	9	27	69	44.2	950	19.2	20	2.17	1.71
10	1999	Floyd	9	16	139	30.9	974	15.2	28	1.14	1.11
11	2011	Irene	8	28	185	28.3	963	12.0	19	1.26	1.97
12	2012	Sandy	10	29	204	41.2	940	8.1	320	2.76	3.38

[a] maximum sustained 1-minute average wind speed

[b] Boose et al. [2003]

[c] Hall model-based pressure estimate – see **Section 2.1.1**

[d] FEMA [2014] data

[e] HURDAT data [*Landsea et al.*, 2004]

[f] EX-HURDAT [*Demuth et al.*, 2006]

[g] Storm data are from the 6 h time step prior to final landfall

[h] Both surge and storm tide are corrected for the changing annual mean sea level

Figure 1. Figure

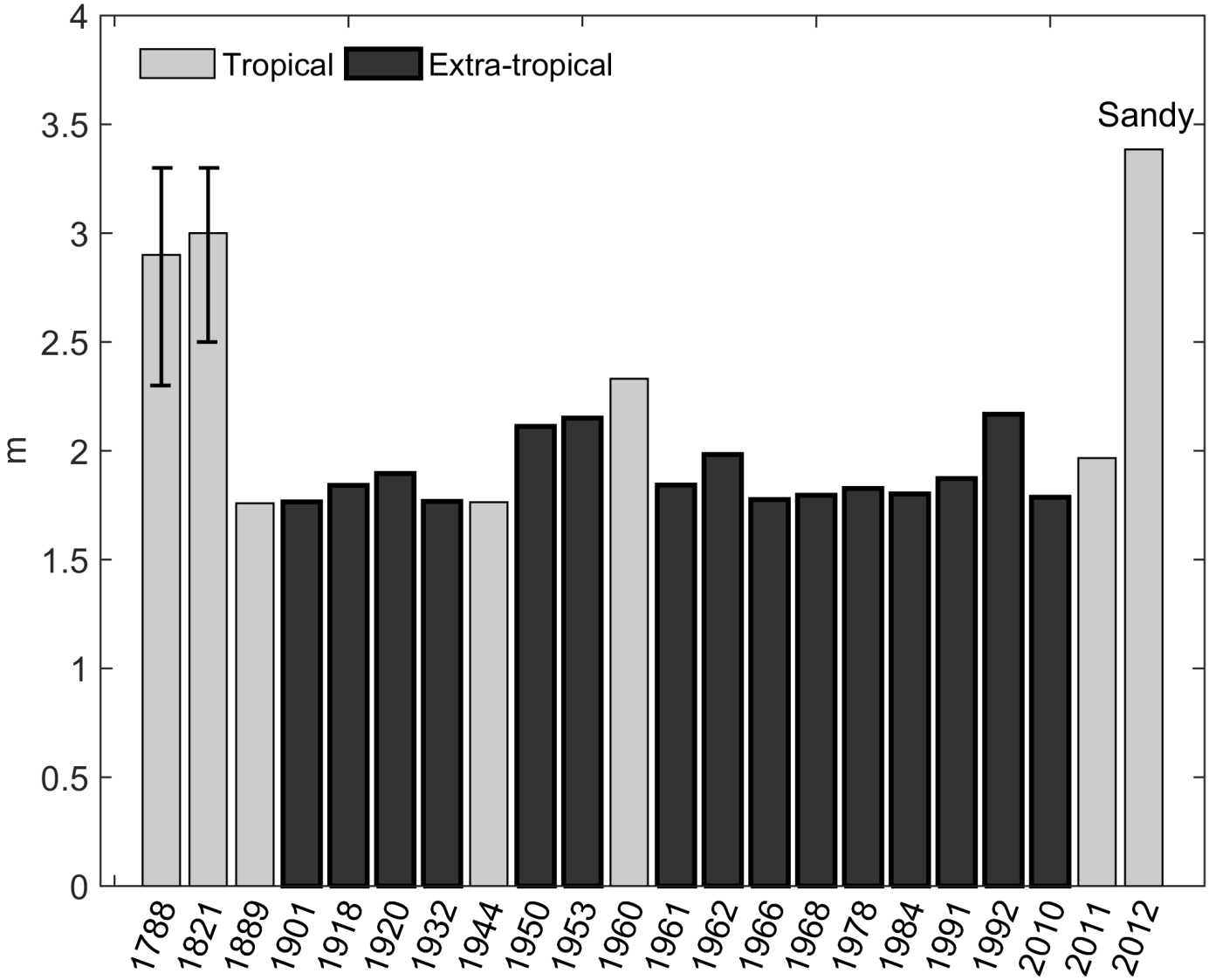

Tropical
Extra-tropical
Sandy
m
4
3.5
3
2.5
2
1.5
1
0.5
0
1788
1821
1889
1901
1918
1920
1932
1944
1950
1953
1960
1961
1962
1966
1968
1978
1984
1991
1992
2010
2011
2012

Figure 2. Figure

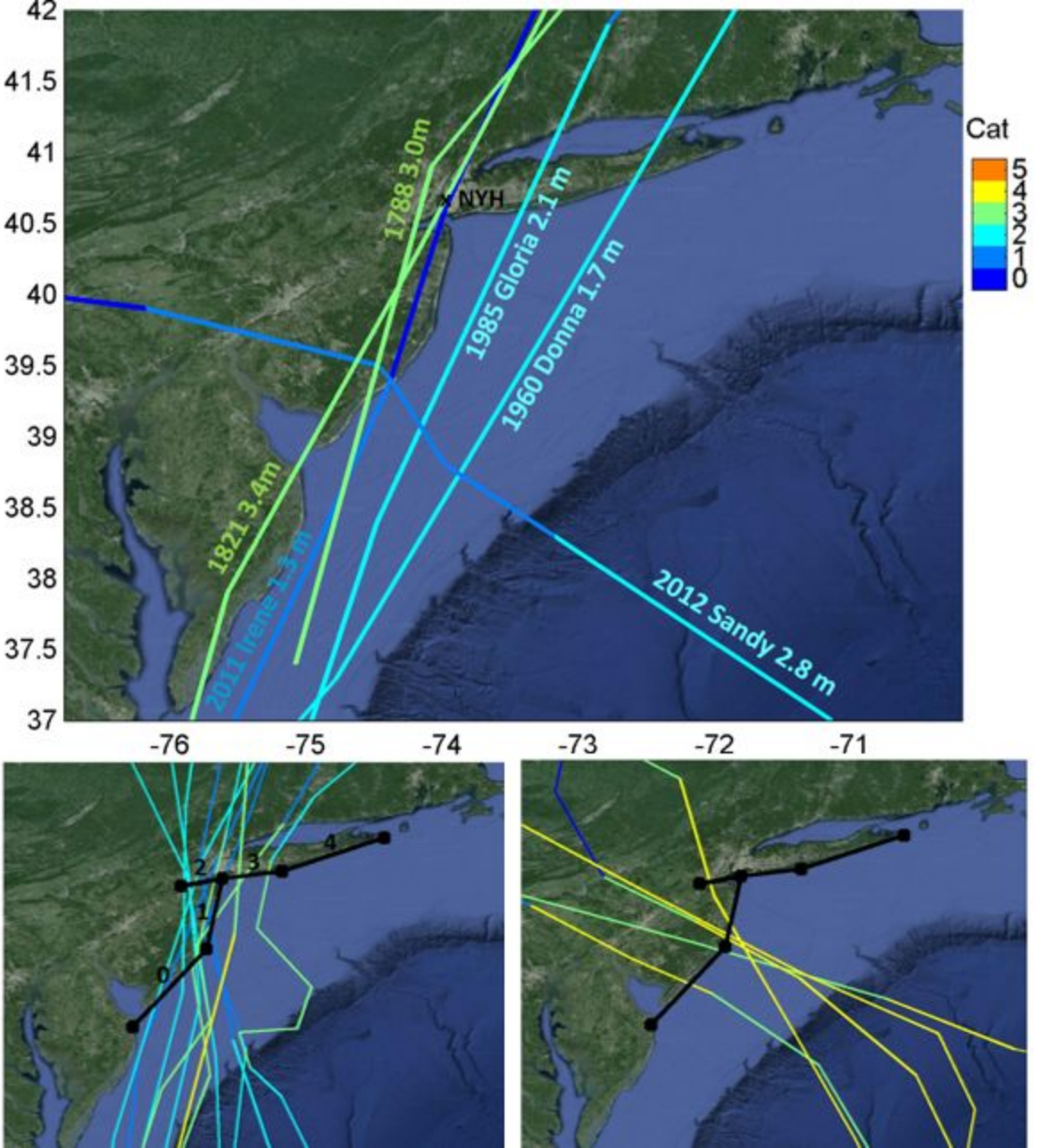

42
41.5
41
40.5
40
39.5
39
38.5
38
37.5
37
-76
-75
-74
-73
-72
-71
Cat
5
4
3
2
1
0
1788 3.0m
NYH
1985 Gloria 2.1 m
1960 Donna 1.7 m
1821 3.4m
2011 Irene 1.3 m
2012 Sandy 2.8 m
0
1
2
3
4

Figure 3. Figure

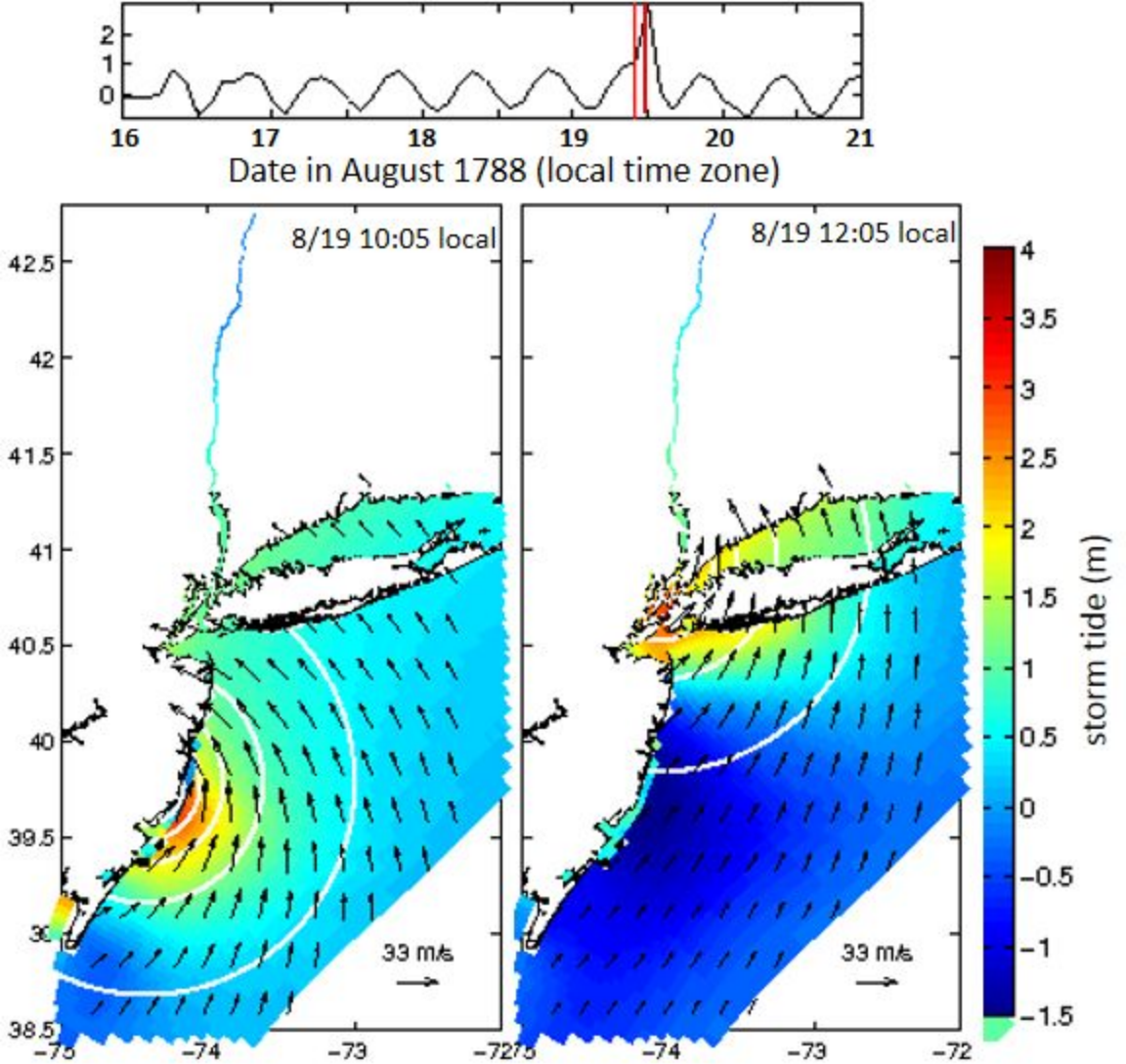

2
1
0
16
17
18
19
20
21
Date in August 1788 (local time zone)
8/19 10:05 local
8/19 12:05 local
42.5
42
41.5
41
40.5
40
39.5
38.5
−75
−74
−73
−72
33 m/s
33 m/s
4
3.5
3
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
storm tide (m)

Figure 4. Figure

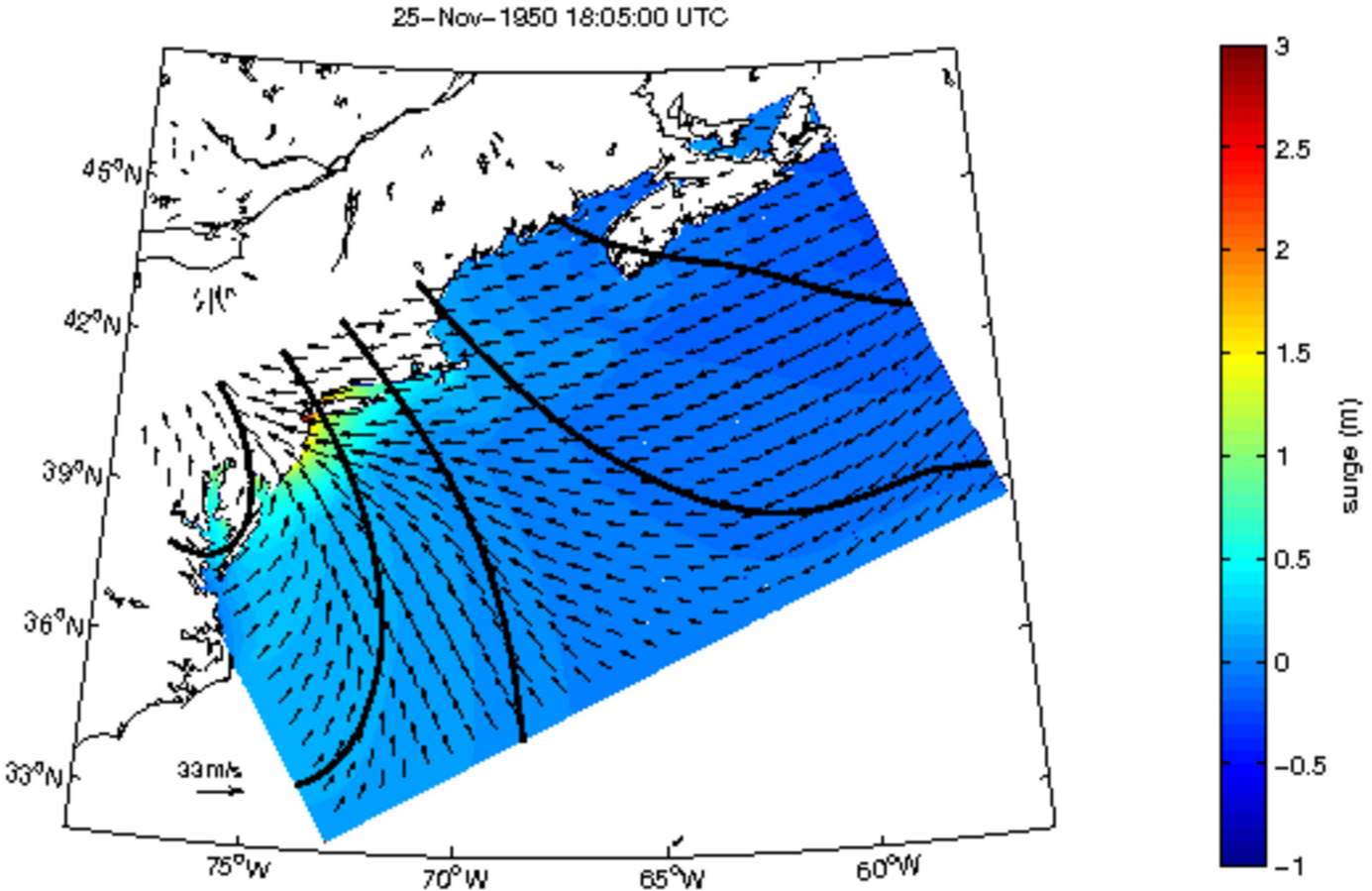

25–Nov–1950 18:05:00 UTC
45°N
42°N
39°N
36°N
33°N
75°W
70°W
65°W
60°W
33m/s
surge (m)
3
2.5
2
1.5
1
0.5
0
−0.5
−1

Figure 5. Figure

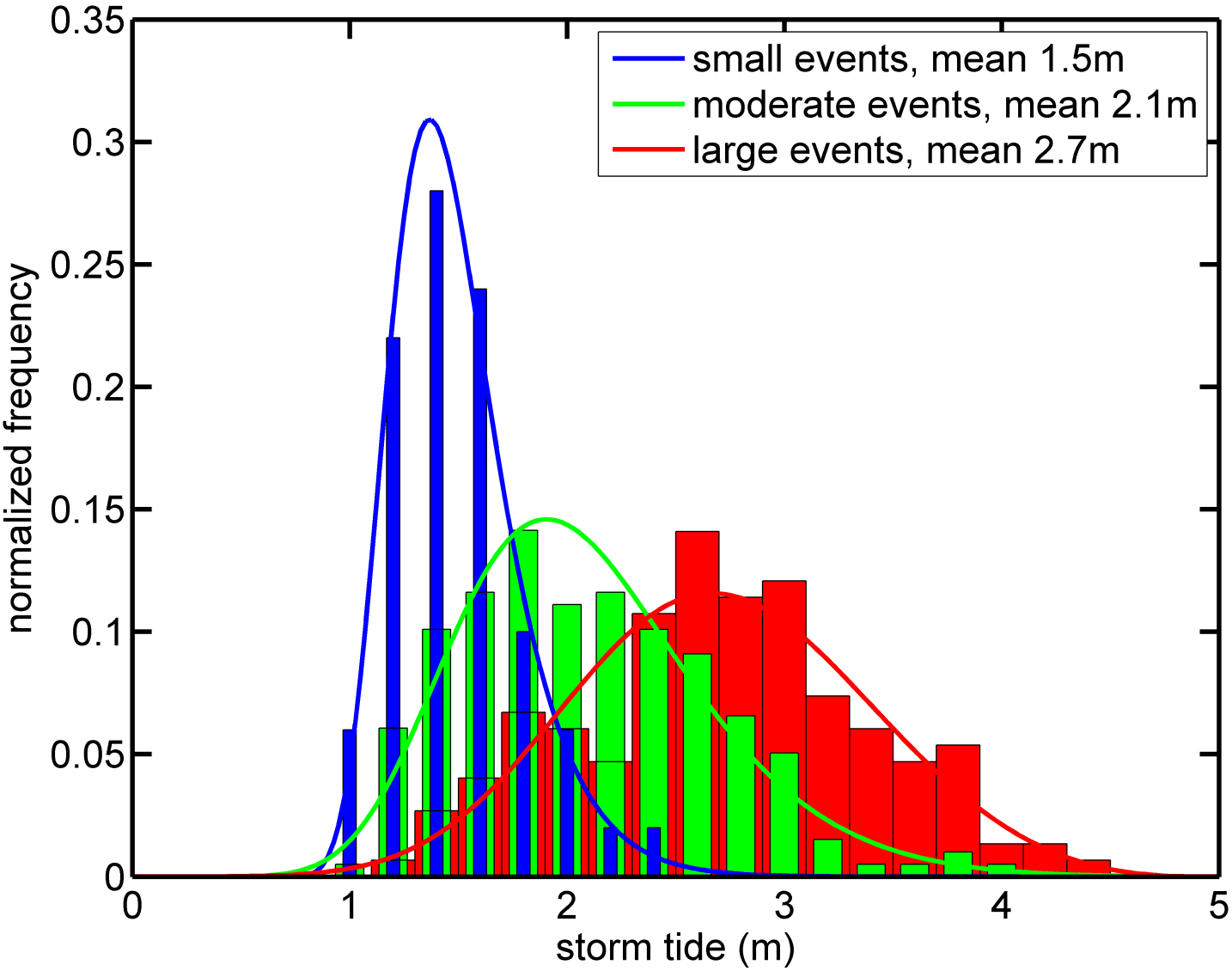

small events, mean 1.5m
moderate events, mean 2.1m
large events, mean 2.7m
normalized frequency
storm tide (m)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5

Figure 6. Figure

Cat 0
Cat 1
Cat 2
Cat 3
Cat 4
Cat 5

annual rate

Gate 0
Gate 1
Gate 2
Gate 3
Gate 4

0-50km
50-100km
>100km

storm tide (m)

storm #

Figure 7. Figure

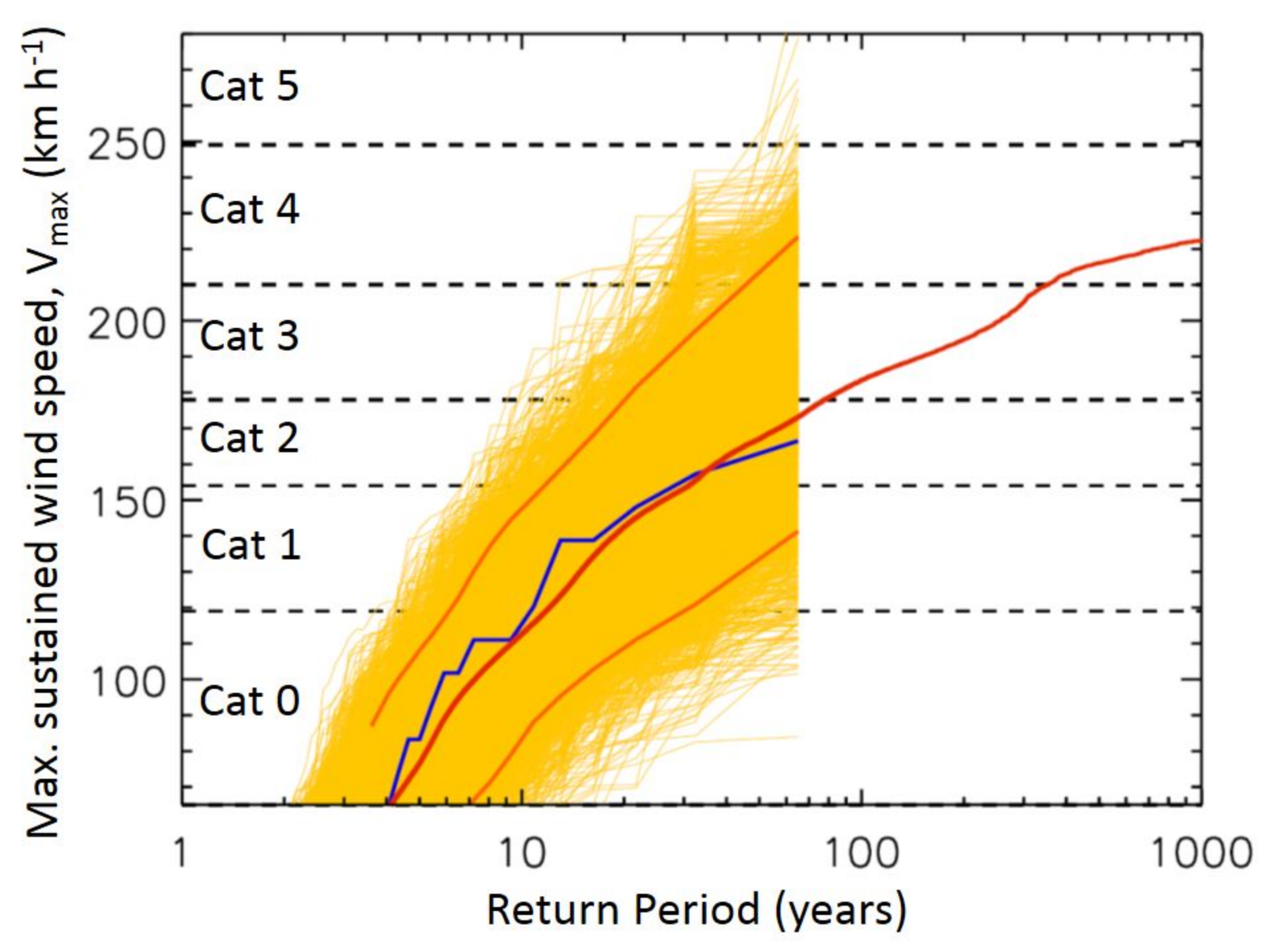

Cat 5
Cat 4
Cat 3
Cat 2
Cat 1
Cat 0
250
200
150
100
1
10
100
1000
Max. sustained wind speed, V_{max} (km h^{-1})
Return Period (years)

Figure 8. Figure

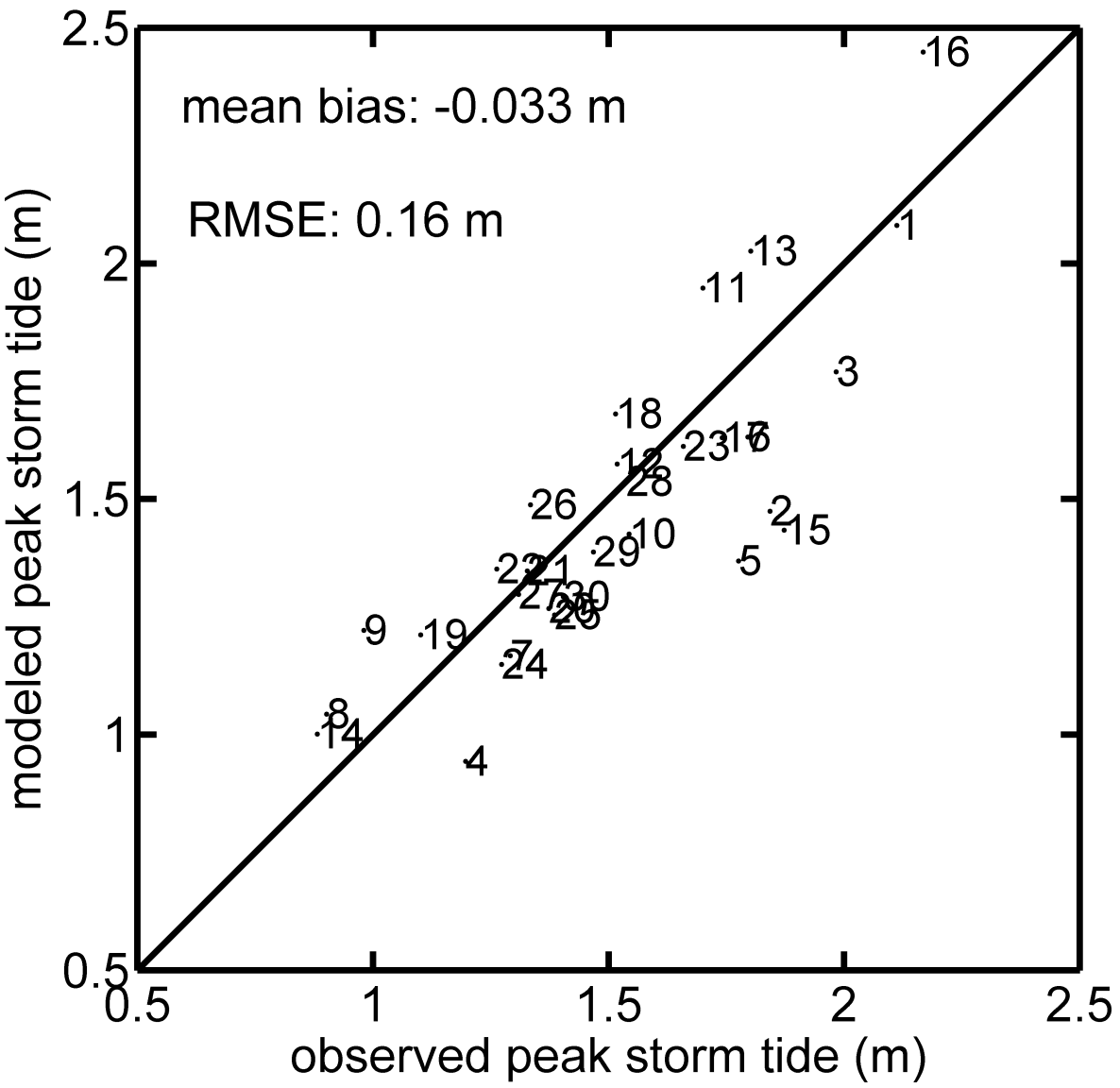

mean bias: -0.033 m
RMSE: 0.16 m
modeled peak storm tide (m)
observed peak storm tide (m)
2.5
2
1.5
1
0.5
0.5
1
1.5
2
2.5

Figure 9. Figure

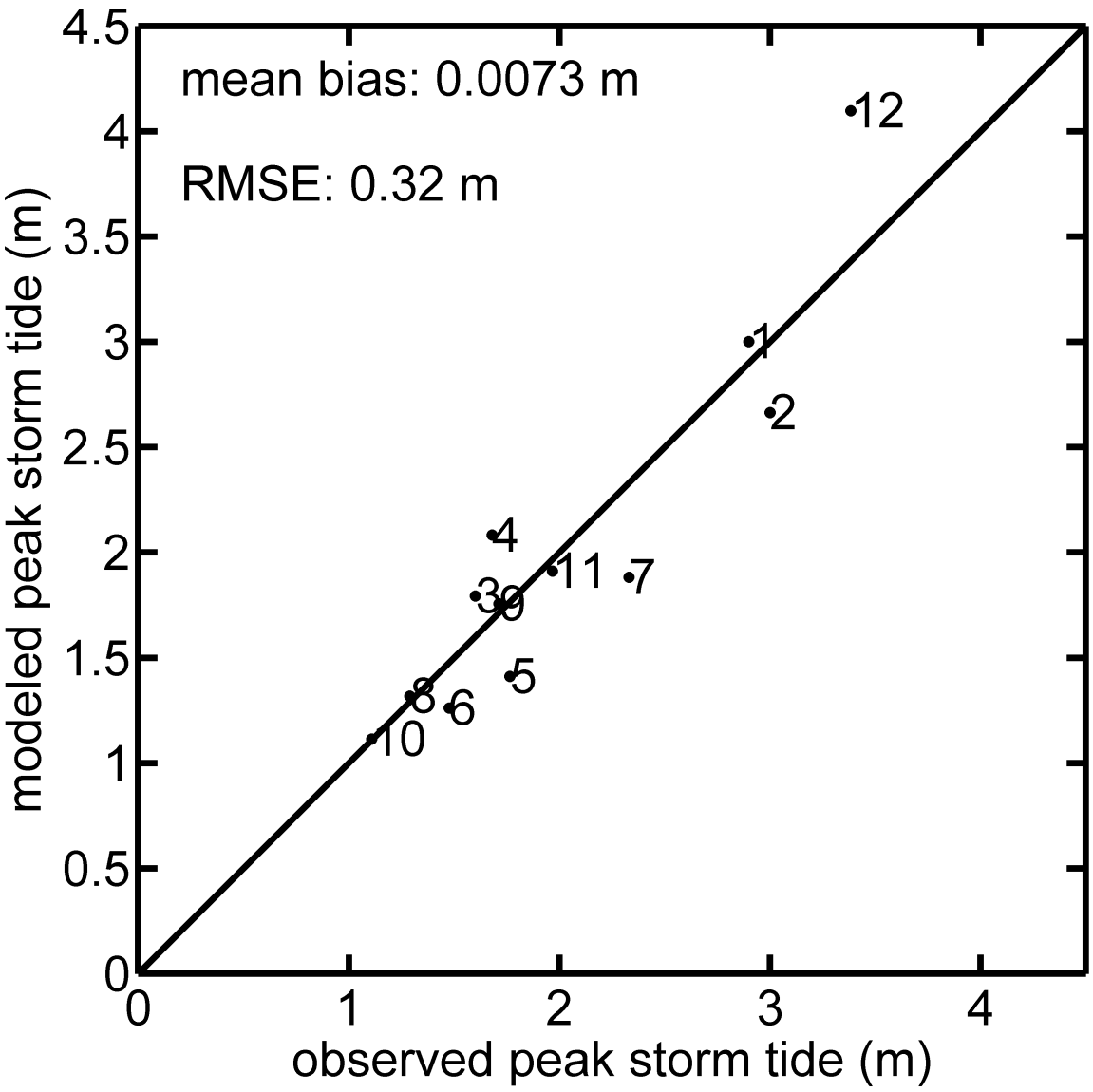

mean bias: 0.0073 m
RMSE: 0.32 m
modeled peak storm tide (m)
observed peak storm tide (m)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0
1
2
3
4
12
1
2
4
11
7
3
9
5
8
6
10

Figure 10. Figure

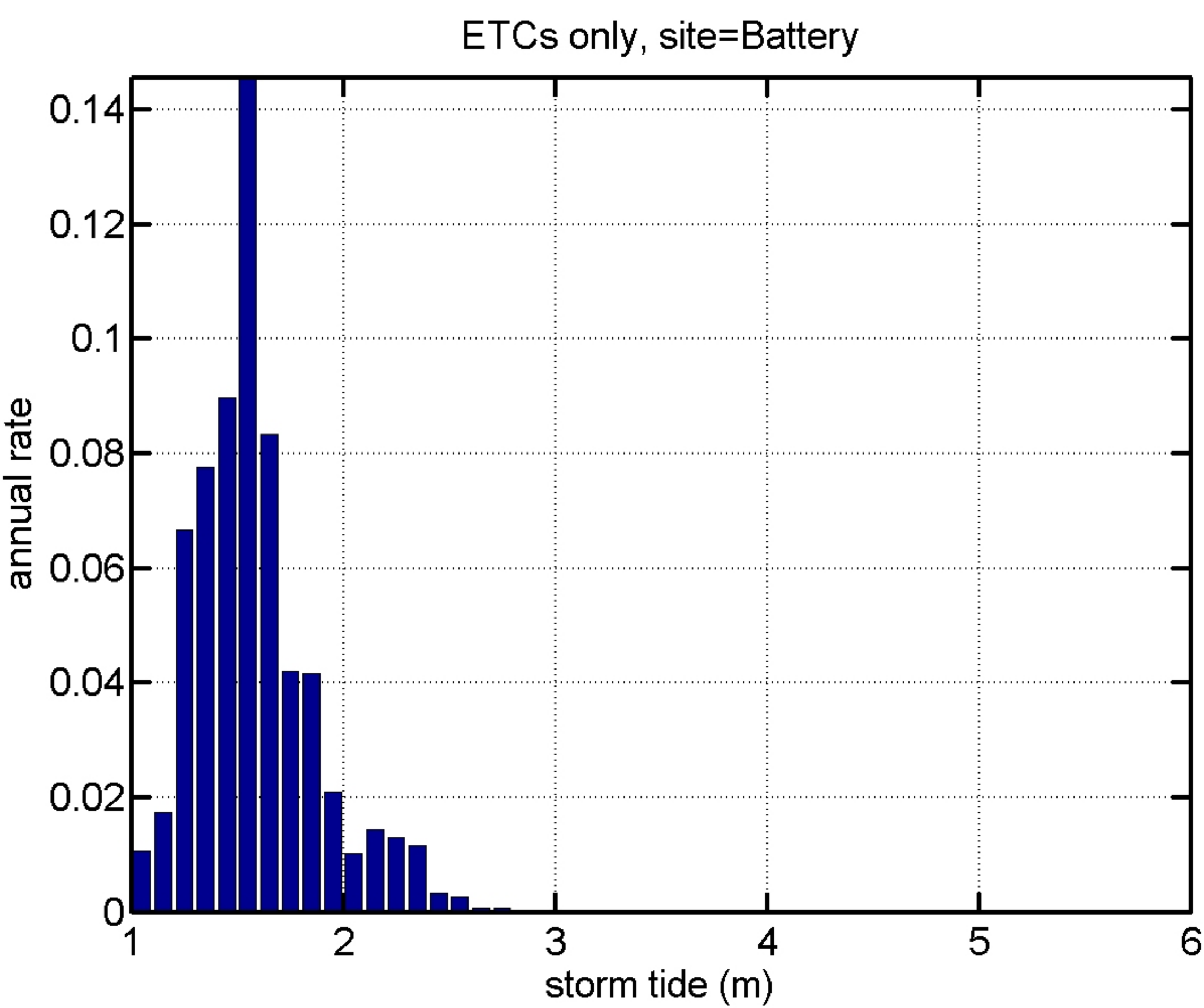

ETCs only, site=Battery
annual rate
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
storm tide (m)

Figure 11. Figure

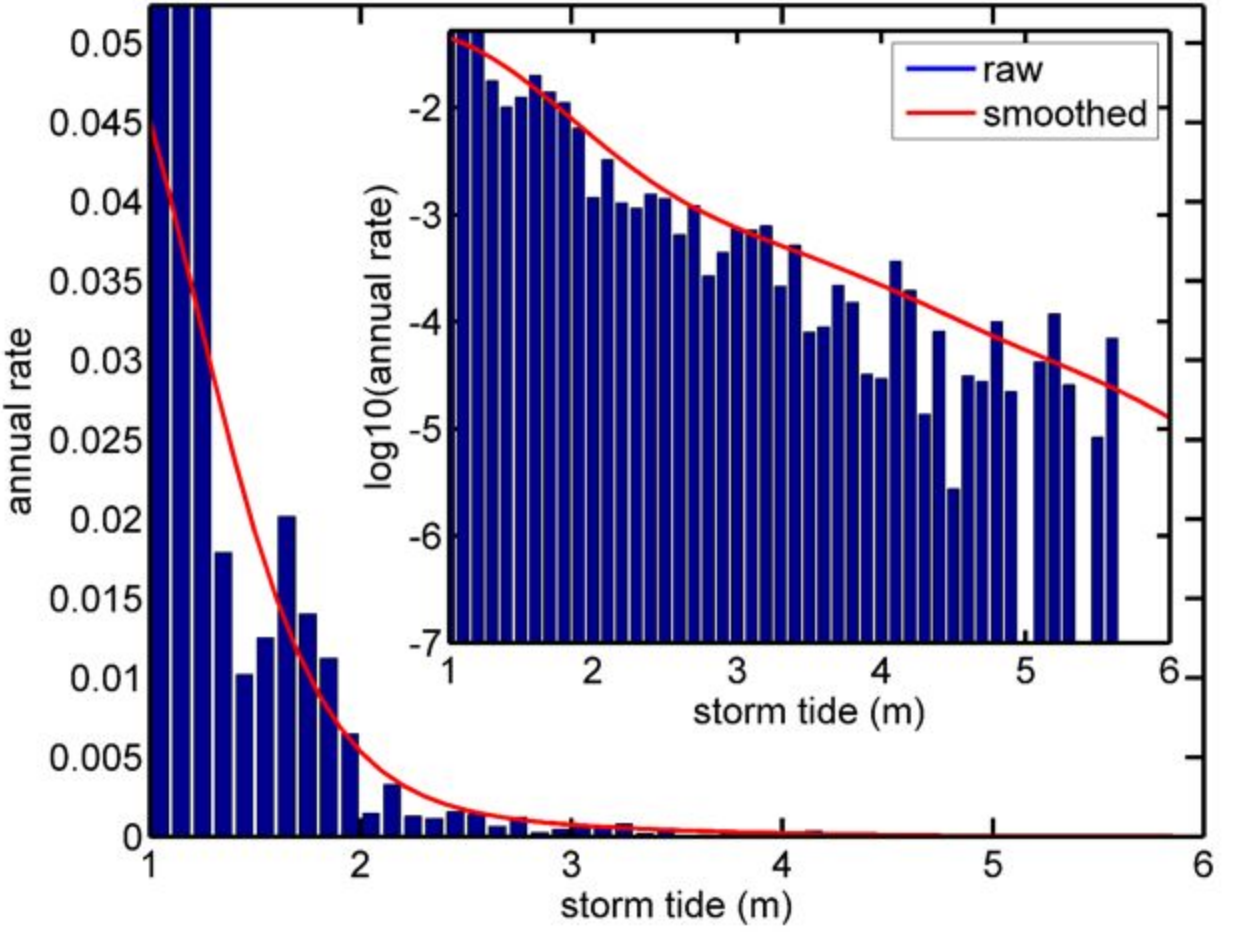

annual rate
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
1
2
3
4
5
6
storm tide (m)
log10(annual rate)
-2
-3
-4
-5
-6
-7
raw
smoothed

Figure 12. Figure

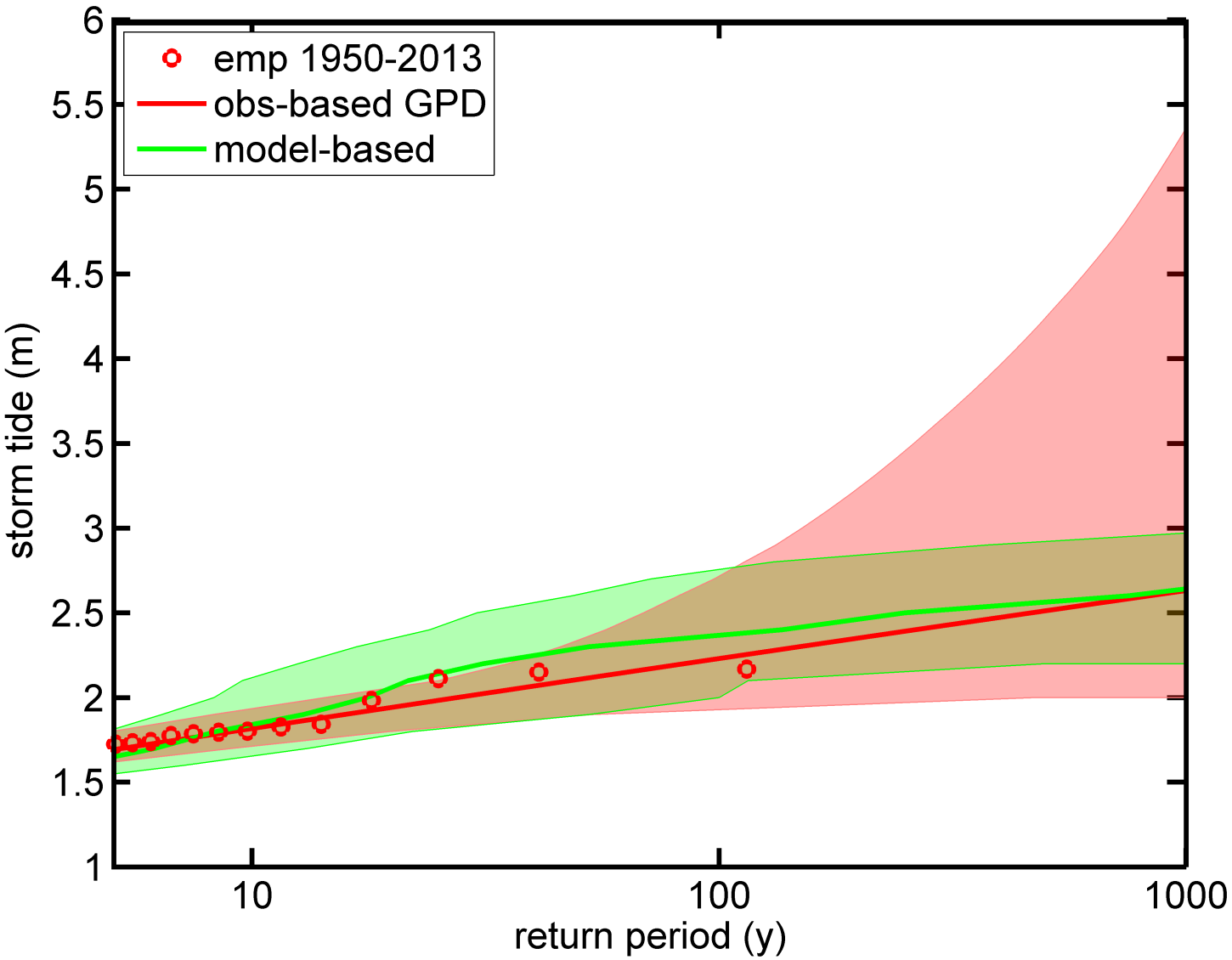

emp 1950-2013
obs-based GPD
model-based
storm tide (m)
return period (y)
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
10
100
1000

Figure 13. Figure

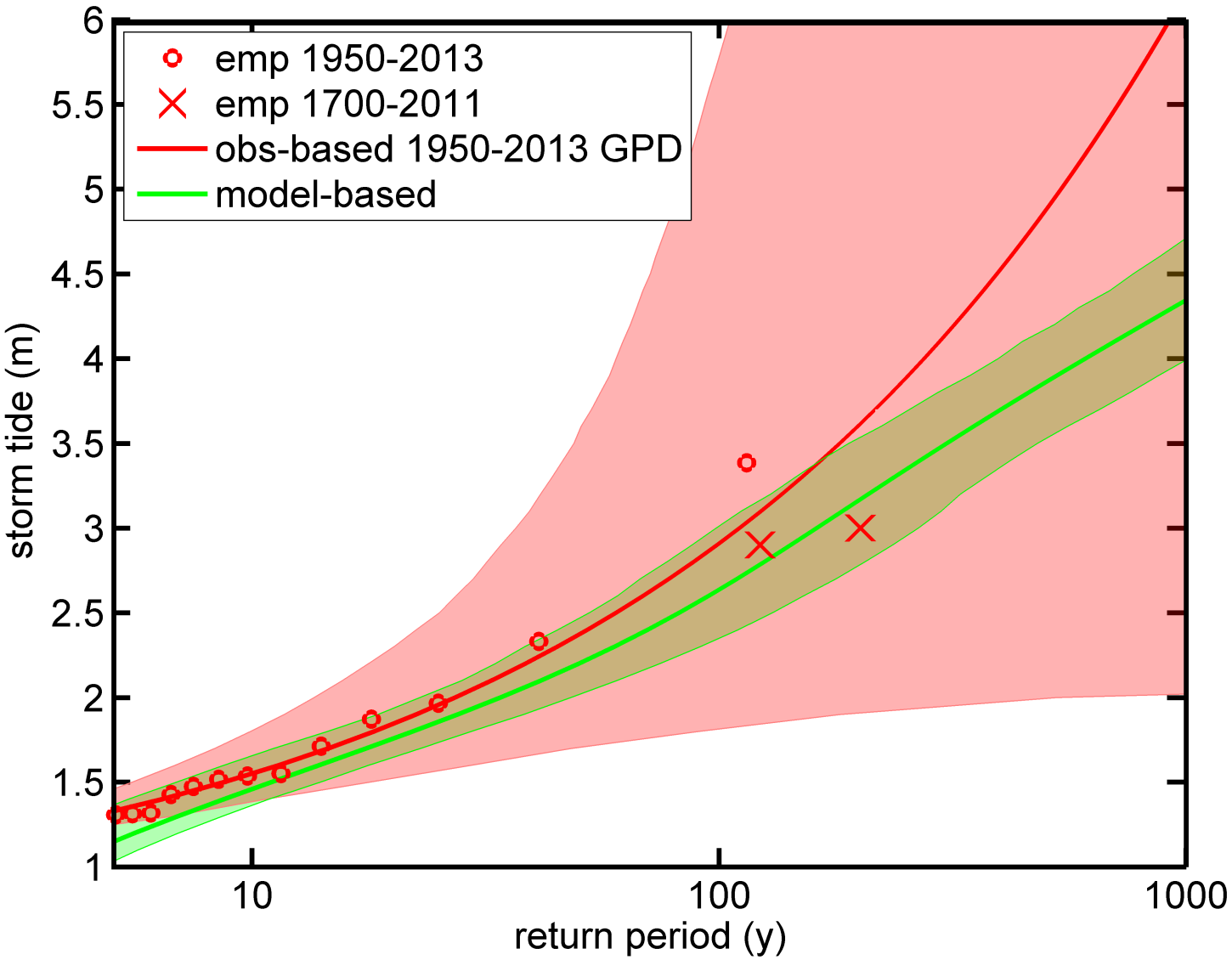
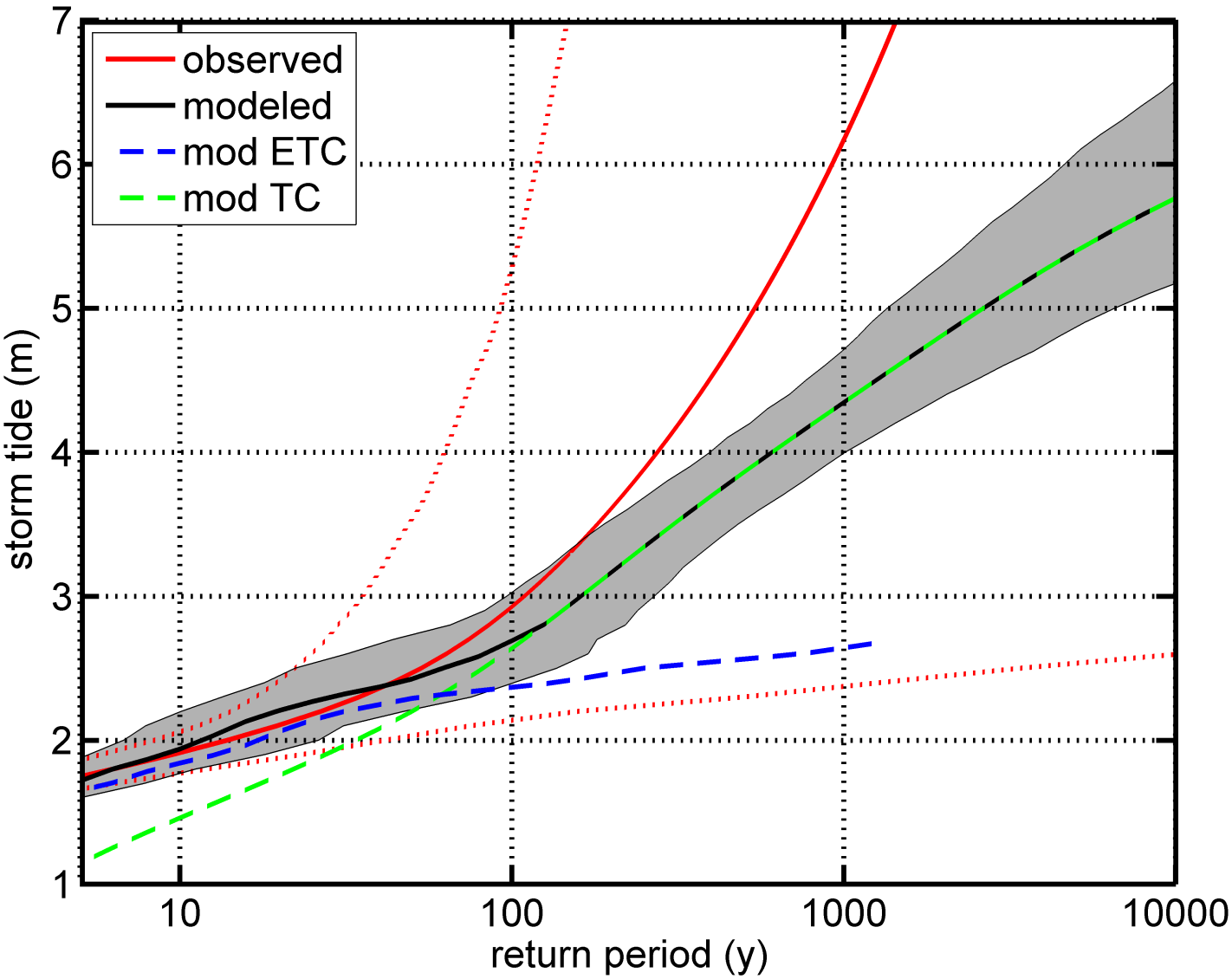

emp 1950-2013
emp 1700-2011
obs-based 1950-2013 GPD
model-based
storm tide (m)
return period (y)
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
10
100
1000

Figure 14. Figure


observed
modeled
mod ETC
mod TC
storm tide (m)
return period (y)
1
2
3
4
5
6
7
10
100
1000
10000