A New Verification Approach? Using Coupled Natural–Human Models to Evaluate the Impact of Forecast Errors on Evacuations

Austin Harris, Paul Roebber, and Rebecca Morss

ABSTRACT: In addition to measuring forecast accuracy in terms of errors in a tropical system's forecast track and other meteorological characteristics, it is important to measure the impact of those errors on society. With this in mind, the authors designed a coupled natural–human modeling framework with high-level representations of the natural hazard (hurricane), the human system (information flow, evacuation decisions), the built environment (road infrastructure), and connections between elements (forecasts and warning information, traffic). Using the model, this article begins exploring how tropical cyclone forecast errors impact evacuations and, in doing so, builds toward the development of new verification approaches. Specifically, the authors implement track errors representative of 2007 and 2022, and create situations with unexpected rapid intensification and/or rapid onset, and evaluate their impact on evacuations across real and hypothetical forecast scenarios (e.g., Hurricane Irma, Hurricane Dorian making landfall across east Florida). The results provide first-order evidence that 1) reduced forecast track errors across the 2007–22 period translate to improvements in evacuation outcomes across these cases and 2) unexpected rapid intensification and/or rapid onset scenarios can reduce evacuation rates, and increase traffic, across the most impacted areas. In exploring these relationships, the results demonstrate how experiments with coupled natural–human models can offer a societally relevant complement to traditional metrics of forecast accuracy. In doing so, this work points toward further development of natural–human models and associated methodologies to address these types of questions and improve forecast verification across the weather enterprise.

KEYWORDS: Forecast verification/skill; Coupled models; Decision making; Emergency preparedness; Societal impacts

AFFILIATIONS: Harris—University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, and National Center for Atmospheric Research, Boulder, Colorado; **Roebber**—University of Wisconsin–Milwaukee, Milwaukee, Wisconsin; **Morss**—National Center for Atmospheric Research, Boulder, Colorado

As discussed in Murphy (1993), Roebber and Bosart (1996), and Morss (2005), forecasts that are "accurate" according to meteorological metrics are not necessarily helpful for everyone using them and can sometimes lead to undesirable outcomes. This is exemplified by Hurricane Irma, a case where meteorologically accurate but inevitably uncertain forecasts—combined with many additional physical–social factors and uncertainties—triggered mass evacuations and severe traffic across Florida in 2017. As the forecasts shifted and traffic worsened, some evacuees became more exposed to hazardous conditions than had they remained in place (Cangialosi et al. 2018; Wong et al. 2018). For these types of reasons, the National Academy of Sciences recommends the weather enterprise measure the impact of forecasts on society (NASEM 2018). Extending this idea to tropical systems like Irma, in addition to measuring forecast accuracy through errors in the system's track, intensity, and other meteorological characteristics, traditional verification metrics should be supplemented with new approaches that measure how forecasts and their errors influence societal impacts of interest, such as evacuation outcomes.

One approach for such work is using computational models to run virtual experiments studying evacuation behaviors across hurricane forecast scenarios. Here, we demonstrate this potential by using a coupled natural–human model named Forecasting Laboratory for Exploring the Evacuation system (FLEE). Introduced by Harris et al. (2021, hereafter HRM21), FLEE builds on previous work using models for studying evacuation communication (e.g., Morss et al. 2017; Watts et al. 2019), evacuation decision-making (e.g., Yin et al. 2014; Widener et al. 2013; Davidson et al. 2018), and evacuation traffic (e.g., Yang et al. 2019; Yi et al. 2017), and links these components together in a computationally feasible framework. By representing the features together with a quasi-realistic representation of forecast data, FLEE enables us to investigate relationships among forecasts, evacuation decisions, and traffic, as a step toward new approaches to evaluating forecasts based on their societal impacts.

FLEE simulates the natural hazard (hurricane), the human system (information flow, evacuation decisions), the built environment (road infrastructure), and connections between systems (forecast information, evacuation orders, traffic). Consistent with the study's goals, FLEE represents key aspects of these subsystems at a high level, but none with the full details of the real system. Decisions on what to include were informed by our understanding of hurricanes and their forecasts combined with empirical knowledge of evacuation gained through surveys and interviews of decision-makers in past hurricanes (e.g., Lindell and Perry 2012; Baker 1991; Huang et al. 2016; Lindell et al. 2019). By integrating the features, FLEE becomes a "virtual laboratory" for exploring how changes in forecasts propagate across subsystems. For example, Harris et al. (2023, hereafter HMR22) used FLEE to explore how evacuations change with various forecast scenarios impacting the Florida Peninsula and how that compares with other factors influencing evacuation, such as evacuation management strategies and population characteristics. As part of this, FLEE's evacuations were validated against empirical evacuation data collected during Hurricanes Irma and Dorian, adding confidence the modeling framework captures the important features for a first-order analysis of the system dynamics.

Building on this work, this paper's objective is to use FLEE to begin assessing how forecast errors influence evacuation outcomes. Results in this article focus on Hurricane Irma, while results for a hypothetical scenario (Hurricane Dorian making landfall across east Florida) are provided in supplemental material. Starting with the National Hurricane Center's (NHC) official (OFCL) forecasts for Irma, we first compare FLEE's simulated evacuations with observed evacuation outcomes. Then, based around the NHC OFCL forecasts, we create hypothetical scenarios with large intensity and forward speed errors, such as when the storm undergoes rapid intensification (RI) and moves faster than expected (rapid onset), and assess their impact on FLEE's evacuations. We then assess the role of reduced track errors on evacuation by using cones of uncertainty (track forecast cones) representative of forecast errors today (2022) and in the past (2007). Through the analysis, we ask the following questions:

RQ1: Do large intensity and forward speed errors, such as those in unexpected rapidly intensifying and/or poorly forecast rapid onset scenarios, negatively impact evacuations?
RQ2: Do improvements in forecast track accuracy over time—as expressed through smaller cones of uncertainty—translate to improved evacuations?

In exploring these questions for this real and hypothetical forecast scenario, we demonstrate how experiments using coupled natural–human models like FLEE can offer a societally relevant complement to traditional metrics of forecast accuracy. As part of this, we point toward the development of more detailed models to explore these types of verification questions further, and in doing so, outline a new approach to help make forecasts more useful across society.

Design and approach
Model overview. Though details regarding FLEE's implementation and design are provided in HRM21, here we highlight important aspects of the model to note when interpreting this study's experiments:

- *Virtual world*—The modeled area is a 10 × 4 cellular depiction of the Florida Peninsula (Fig. 1; grid cells are 69 km × 69 km each). FLEE includes 4.1 million household agents (groups of four individuals who collectively make evacuation decisions; Lindell et al. 2019) whose spatial distribution on the grid is approximated via Census data.
- *Forecast data*—Every 6 h, archived NHC forecast products depicting the storm's current and forecast information are synthesized to create a red–orange–yellow–green "light system" forecast of wind, storm surge, and rain risk for each of FLEE's grid cells. The information used includes the following:
  - Wind risk: Forecast category, location in the forecast wind field, location relative to the cone of uncertainty, expected time of arrival of tropical storm force winds
  - Surge risk: Forecast category, location in the forecast wind field, location relative to the cone of uncertainty, expected time of arrival of tropical storm force winds, grid cell's surge inundation potential, storm's approach angle relative to coastline
  - Rain risk: Forecast forward speed, location in the forecast wind field, location relative to the cone of uncertainty, expected time of arrival of tropical storm force winds
- *Evacuation orders*—Emergency manager agents, located within FLEE's coastal grid cells, decide whether to issue evacuation orders based on storm surge risk, clearance times, and the forecast time of arrival of tropical storm force winds.
- *Evacuation decisions*—Household agents decide to evacuate based on a combination of wind, surge, and rain risk for their location, evacuation order information, and household characteristics (mobile home ownership, age, car ownership, and socioeconomic status).
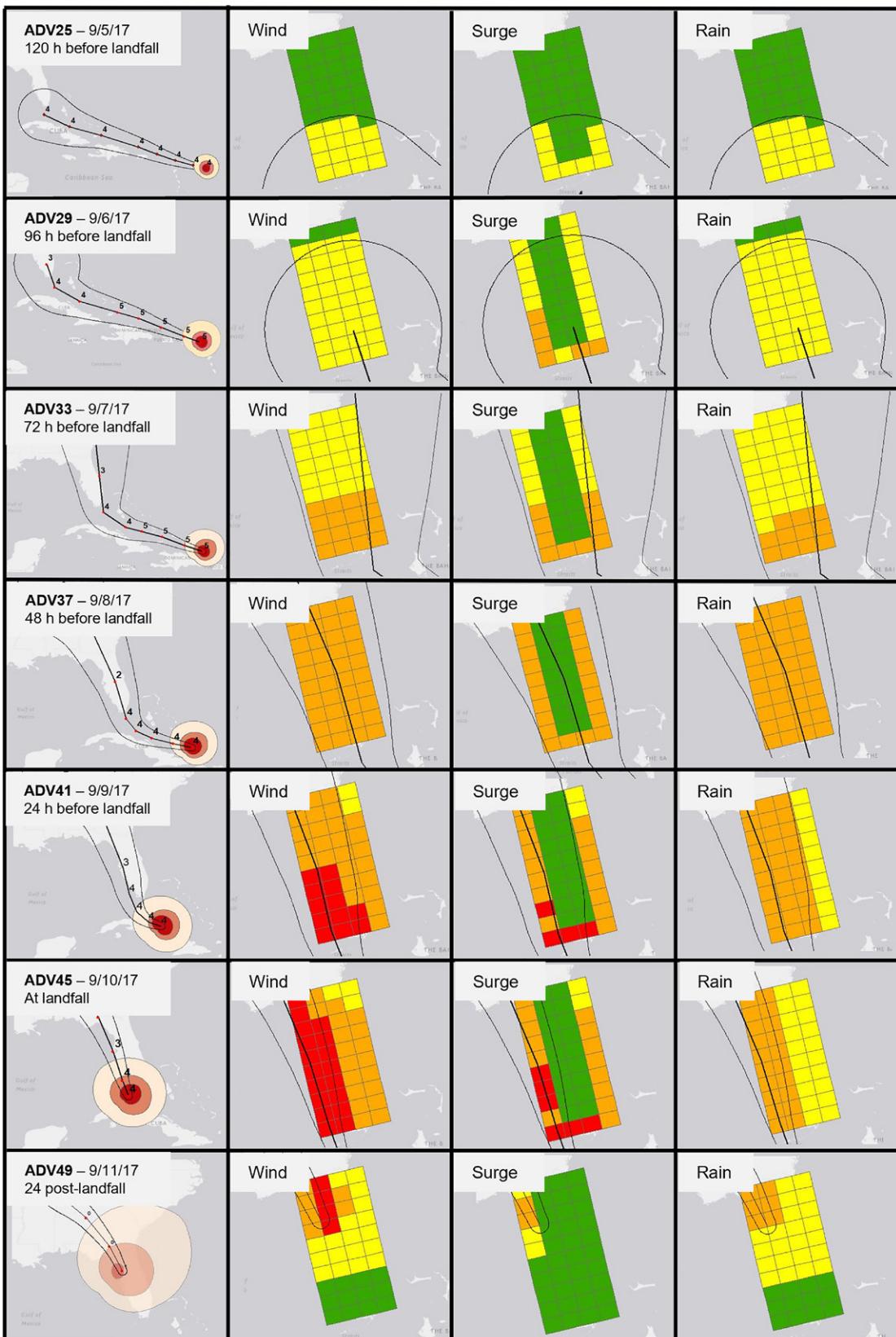
Fig. 1. NHC OFCL forecasts for Hurricane Irma (2017) and corresponding FLEE estimates of wind, surge, and rain risks as the storm approaches and travels through the Florida-like model grid. Forecasts are shown at 24 h intervals, but update every 6 h in the model simulations (not shown). (column 1) Evolving NHC OFCL forecast track (black center line), forecast storm intensity on the Saffir–Simpson scale (TS/numbers), 2017 cone of uncertainty (edges are outer black lines), and current wind radii at 34 (white), 50 (pink), and 64+ (red) kt intervals. (columns 2–4) The light system threats for wind, surge, and rain are shown on the model grid for equivalent times in the simulation, with the OFCL forecast track (center black line) and 2017 cone of uncertainty (outer black lines) included for reference. Note that threats are highest when near the center of the forecast cone and when hazards are most imminent, among other factors.

We exclude some factors known to influence evacuation decisions (e.g., fuel shortages during Irma) as we are not seeking a fully realistic algorithm, but one capturing the main processes across most cases.

- *Traffic and the built environment*—Idealized highways and interstates simulating key aspects of Florida's road network (e.g., I-75 and I-95) are overlaid on FLEE's grid. The roads allow evacuating households to move between grid cells. If spots are unavailable on roads due to traffic for an extended period, household agents who decided to evacuate will shelter in place instead.

Model validation. FLEE's evacuation response to Irma and Dorian's NHC OFCL forecasts were validated against available empirical data for the storms in HMR22, focusing on the spatial and temporal patterns of evacuation orders, evacuation rates, and traffic intensity. Key aspects of this validation are resummarized in the "Results" section. The available empirical data vary across studies; e.g., evacuation rates are calculated by county, zip code, and region. Because it is difficult to translate the empirical data by county, for example, onto FLEE's grid cells for an exact comparison, we instead aggregate the empirical data to ensure the bigger picture aspects of FLEE's evacuation behave in a sufficiently realistic manner for a first-order analysis exploring the relationships between forecast errors and evacuation outcomes.

Experimental design. To assess the impact of forecast errors on evacuation outcomes, we first identify the "typical" errors for tropical cyclone forecast elements, presently and historically. For track and intensity, average forecast errors as well as their trends over time are available for 0–120-h lead times on NHC's website (www.nhc.noaa.gov/verification/index.shtml). NHC's forward speed forecast errors (i.e., along track errors) are not readily available, though they are slightly larger than cross-track errors (as noted by Fossell et al. 2017). Storm size forecast errors are also unavailable, as it is difficult to accurately verify wind radii forecasts, and because error measurements may not paint a complete picture depending on the instrumentation used for validation (Cangialosi 2019). Based on data availability, we focus on track and intensity errors.

Forecast track errors are also reflected in the NHC's Track Forecast Cone product, also referred to as the cone of uncertainty. According to the NHC website,

> The cone represents the probable track of the center of a tropical cyclone, and is formed by enclosing the area swept out by a set of circles along the forecast track (at 12, 24, 36 hours, etc). The size of each circle is set so that two-thirds of historical official forecast errors over a 5-year sample fall within the circle.

The sizes of the circle radii defining the NHC cones—both present and historical—can be found at www.nhc.noaa.gov/aboutcone.shtml. Because track errors have been decreasing over time, the cone of uncertainty has been shrinking since its implementation in 2002.

To explore the role of forecast track errors on evacuation outcomes in FLEE, our approach is to change the cone of uncertainty to sizes representative of today (2022) and in the past (2007). This period was chosen as it represents 15 years of progress reducing forecast track errors in the weather enterprise; e.g., the 2007 cone is nearly double the size of the 2022 cone (sizes of the circles used to create the cones are provided in online supplemental Table S1; https://doi.org/10.1175/BAMS-D-22-0136.2). Another reason for choosing the period is that the cones were calculated differently before 2007 (i.e., errors over a 10-yr sample were used as opposed to the 5-yr sample used after 2007). By comparing the evacuation response in FLEE using the 2007 cone with those using the 2022 cone, we can begin to

quantify the value of reduced track errors on evacuation outcomes across this period for these scenarios (RQ2).

We note that the experiments using the 2007 and 2022 cones used the NHC best track (observed track) to create "forecasts" with varying levels of uncertainty. This was preferred over the NHC OFCL forecast track since the latter contains track errors. Best tracks were downloaded at www.nhc.noaa.gov/gis/; ArcGIS was used to overlay the cones onto the best track. The light system forecasts corresponding to these experiments are provided in Figs. S1, S2, S6, and S7.

To explore the role of forecast intensity errors on evacuations, our first approach was to introduce erroneous intensities higher and lower than OFCL forecasts by amounts representing average errors in 2007 and 2022. However, average errors are less than 20 kt (1 kt ≈ 0.51 m s⁻¹) in both cases, even at long lead times. As a result, intensity errors are too small to effectively resolve in the current implementation of FLEE, where light system forecasts are synthesized into four categories (red, orange, yellow, green) for all tropical systems.

Because of the limitation, we instead create poorly forecast rapidly intensifying/rapid onset (RI/RO) scenarios where intensity and forward speed errors are large. In these scenarios—which are a known problem in meteorology (DeMaria et al. 2021)—we shorten the NHC OFCL forecast timeline while keeping the peak magnitudes of risk the same, and simulate its effect on evacuations. More specifically, the NHC OFCL forecast timeline is shortened from 168 to 84 h and 72 h by using every other advisory (only the 0000 and 1200 UTC advisories), while only 6 h of time elapses in the simulation. This accelerates the storm's forward speed and evolution of the intensity forecasts (light system forecasts are summarized in supplemental Figs. S3, S4, S8, and S9). By comparing the evacuation from the RI/RO forecast scenarios to the NHC OFCL forecasts and to each other, we begin to tease out the potential role of poorly forecast RI/RO scenarios on evacuation outcomes (RQ1).

The full list of experiments are provided in Table 1. Simulations are run using the NHC OFCL forecast, the two RI/RO forecast scenarios, and for the two different cone sizes. Experiments were repeated for two storm scenarios: one real (Hurricane Irma) and one hypothetical (Hurricane Dorian making landfall across east Florida, shown in supplemental material). The purpose of the hypothetical Dorian scenario is to demonstrate the potential of using these types of coupled natural–human models to explore potentially impactful scenarios that may not have occurred yet. Together, these experiments allow us to answer RQ1 and RQ2 for the two scenarios, and in doing so, point toward developing more detailed models and experiments to answer related and more specific verification questions.

Table 1. Experiments in the study. We note that the NHC OFCL and RI/RO experiments use the 2017 cone of uncertainty for Irma experiments (1–3) and the 2019 cone of uncertainty for the hypothetical Dorian (landfalling) experiments (6–8). For the hypothetical Dorian experiments, tracks in all experiments (6–10) are shifted westward so the storm makes landfall along Florida's east coast.

Irma		Dorian (landfalling)	
Expt No.	Expt name	Expt No.	Expt name
1	NHC OFCL	6	NHC OFCL (shifted)
2	NHC OFCL with RI/RO	7	NHC OFCL (shifted) with RI/RO
3	NHC OFCL with RI/RO − 12 h	8	NHC OFCL (shifted) with RI/RO − 12 h
4	NHC best track with 2007 cone of uncertainty	9	NHC best track (shifted) with 2007 cone of uncertainty
5	NHC best track with 2022 cone of uncertainty	10	NHC best track (shifted) with 2022 cone of uncertainty

Data analysis. To compare evacuation behaviors across experiments, we track evacuation statistics over time for each of FLEE's grid cells. The primary model outputs analyzed are evacuation rates (the percentage of households in a grid cell that successfully evacuated) and the percentage who wanted to evacuate but were unsuccessful due to excessive traffic in their area.

In addition, we aggregate model data into multiple impact zones, designed as first-order approximations of areas likely to experience different levels of impacts based on the actual meteorological conditions produced by the storm. Here, we use four impact zones, defined by whether the grid cells are 1) coastal or inland, and 2) experience winds greater than 64 kt (hurricane force) or less than 64 kt during the storm of interest. Using the impact zones, we can determine who evacuated from locations that did not end up experiencing hazardous wind conditions (future versions could do this for rain and surge parameters as well). In interpreting results, we compare metrics that might indicate successful outcomes in different ways. For example, high evacuation rates may not be preferred if the storm ends up not having much impact in those areas, and unnecessary evacuations may not matter if those at high risk can get out safely.

Results

Comparing FLEE with observations. Figure 1 shows NHC OFCL Irma forecasts (left column) and the equivalent light system representations of wind, surge, and rain risk (right columns) at 24-h intervals. Early forecasts place FLEE's entire model grid within NHC's cone of uncertainty (cone size representative of the time of the event in 2017) with the most likely outcome (center track) being a landfalling major hurricane in southeastern Florida near Miami. However, forecasts shifted toward a west Florida landfall as the storm approached, with the storm eventually making one mainland U.S. landfall as a category 4 storm in the Florida Keys and a second landfall as a category 3 storm in southwestern Florida. Irma's hurricane-force winds impacted the western two-thirds of Florida—particularly the southwest coastlines—while tropical-storm-force winds, flooding, and power outages were observed along the eastern coastline. Based on these Irma forecasts, emergency manager agents in FLEE issued evacuation orders starting in Miami–Fort Lauderdale then expanding outward along both coasts as the storm approached (Figs. 2a,b; red cells), which was observed in Irma's actual evacuation orders (Wong et al. 2018; Darzi et al. 2021). The comparison with empirical data on evacuation orders increases our confidence that FLEE's evacuation order algorithm behaves sufficiently realistically for the purposes here.

FLEE's simulated evacuation rates based on NHC OFCL forecasts (Fig. 2a) vary from 20% to 40% along Florida's east coasts, 40%–70% along the south and west coasts, and 10%–40% inland. This closely resembles the observational data, which also suggest evacuation rates vary from 20% to 40% along Florida's east coast, to 40%–70% across the south and west coasts, and around 10%–30% inland (data aggregated from Wong et al. 2018; Long et al. 2020; Martín et al. 2020; Feng and Lin 2021).

Since this was the largest evacuation in U.S. history, severe traffic was observed across Florida before Irma (Wong et al. 2018). In FLEE, traffic is most severe around Tampa Bay–Saint Petersburg (Fig. 2b) with 5%–20% of households in the metropolitan area unsuccessful at evacuating due to traffic. This broadly matches observations of traffic rates, which shows severe traffic across the Tampa Bay, I-75, and surrounding areas (Feng and Lin 2021; Staes et al. 2021), and our general understanding of south and western Florida being difficult to evacuate based on the high population and limited evacuation routes.

Across FLEE's entire model grid, 32.0% of households evacuated, which equals 5.4 million people (Table 2, column 1). FDEM (2018) suggest actual evacuation numbers
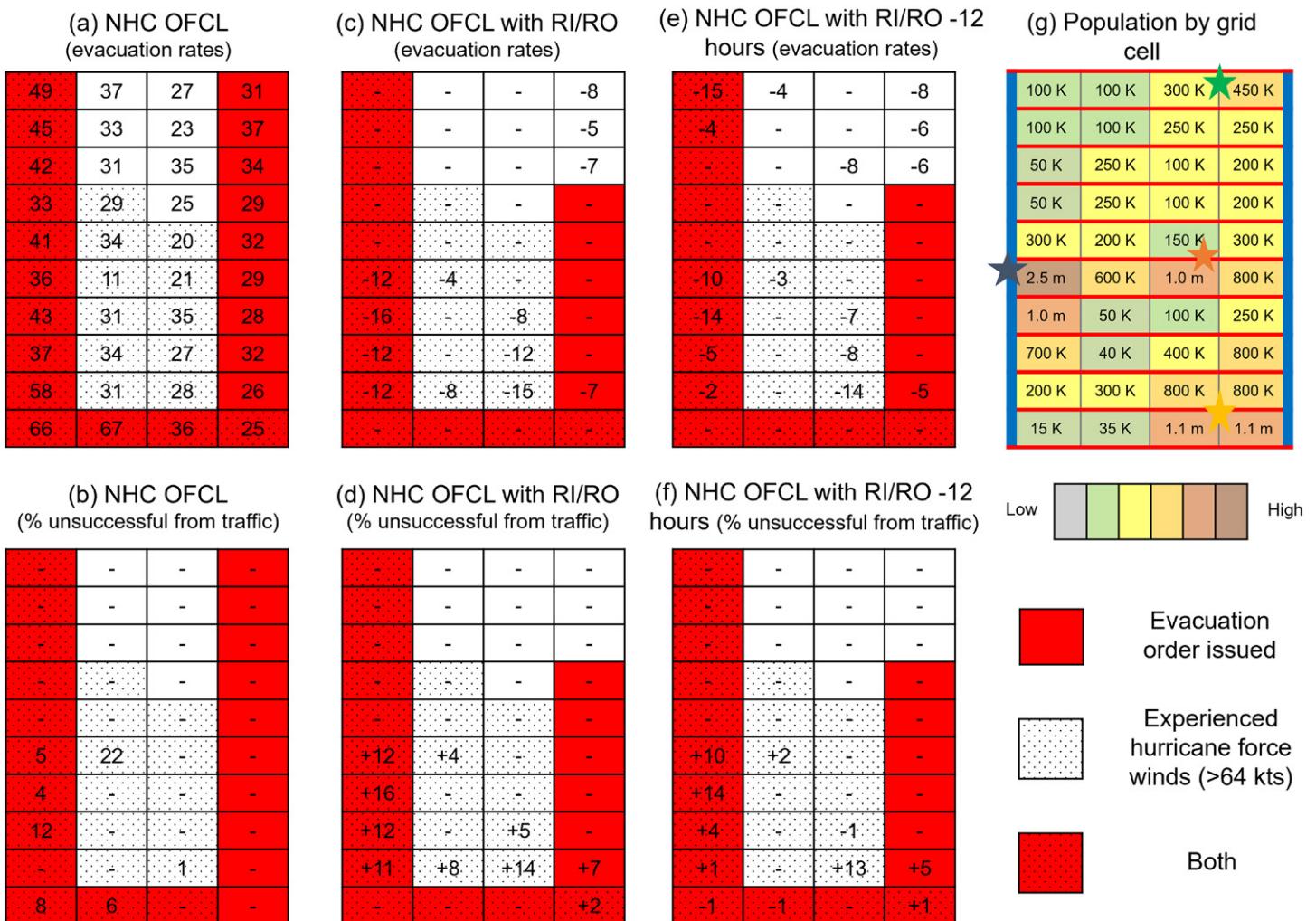
Fig. 2. Irma's simulated evacuation outcomes across grid cells for the NHC OFCL and RI/RO experiments. The first items presented are (a) evacuation rates and (b) percentage of households who decided to evacuate but were unsuccessful due to traffic, both for the experiment using Irma's NHC OFCL forecast. Evacuation rates and the percentage of households who decided to evacuate but were unsuccessful due to traffic are presented for (c),(d) the NHC OFCL with RI/RO experiments and (e),(f) the NHC OFCL with RI/RO – 12 h experiments with values expressed as the departure from those in (a),(b) NHC OFCL. Also shown in (a)–(f) are grid cells that experienced hurricane force winds (dotted cells) and evacuation orders (red cells). (g) To provide a frame of reference for interpreting the simulations, the population by grid cell along with the approximate location of several major cities on the model grid are shown: Miami–Fort Lauderdale (yellow star), Tampa Bay–Saint Petersburg (blue star), Jacksonville (green star), and Orlando (orange star).

totaled 6.9 million, and when considering households evacuating to local shelters in FLEE (not shown), the modeled evacuation rates closely resemble the observations. The temporal evolution of FLEE's evacuation rates (supplemental Fig. 5) is linear during the event, matching observations in Wong et al. (2018).

This comparison of FLEE's simulated evacuation using NHC's OFCL forecasts with empirical data on evacuation orders, rates, and traffic suggests FLEE captures the broader patterns of evacuation for Irma and thus provides a realistic baseline for interpreting results from other experiments.

Impact of poorly forecast RI/RO. Regarding the NHC OFCL with RI/RO experiments where forward speed and intensity errors are significant (RQ1), evacuation rates decreased everywhere relative to NHC OFCL while evacuation traffic increased (Table 2, Figs. 2c,d). For example, evacuation rates in the NHC OFCL with RI/RO experiment decreased by 4.7% (630,000 fewer evacuees) relative to NHC OFCL across the entire model grid, while the

Table 2. Irma's simulated evacuation behaviors averaged across all grid cells for all experiments. In addition to evacuation rates and total numbers evacuated, evacuation rates are broken down into impact zones (coastal vs inland and areas experiencing vs not experiencing hurricane force winds of 64+ kt) and the percentage and numbers of evacuees who attempted to evacuate but were unsuccessful due to excessive traffic. In columns 2 and 8, m indicates million and K indicates thousand.

Irma expt	Evacuation rates (%)	Evacuated	Coastal > 64 kt zone (%)	Inland > 64 kt zone (%)	Coastal < 64 kt zone (%)	Inland < 64 kt zone (%)	Unsuccessful evacuation due to traffic (%)	Unsuccessful evacuation due to traffic
NHC OFCL	32.0	5.24m	39.3	24.5	29.6	29.4	2.5	410K
NHC OFCL with RI/RO	27.3	4.61m	32.0	20.3	27.2	28.4	6.3	1.06m
NHC OFCL with RI/RO − 12 h	26.3	4.44m	27.7	18.2	26.5	25.9	7.6	1.28m
Best track with 2007 cone	32.5	5.33m	38.2	27.1	31.2	26.0	2.3	380K
Best track with 2022 cone	31.2	5.12m	38.6	26.0	27.7	26.0	1.7	286K

number of unsuccessful evacuations due to traffic increased by similar amounts. These impacts were most pronounced in heavily impacted areas, such as Tampa Bay and Fort Myers, where a 12%–16% reduction in evacuation rates was observed (Fig. 2c). Similarly, RI/RO experiments in the Dorian (landfalling) scenario reduced evacuation rates across impacted areas (while increasing evacuation rates in less impacted areas; Fig. S8, Table S2).

When comparing the two NHC OFCL with RI/RO experiments—in both Irma and Dorian (landfalling) forecast scenarios—cases with larger forward speed and intensity errors (RI/RO − 12 h) resulted in worse evacuation rates and more traffic (Table 2, columns 2–8, Figs. 2c–f). This suggests that an extra 12 h of forecast lead time can improve evacuations in these scenarios.

Results from the poorly forecast RI/RO scenarios with large intensity and forward speed errors make sense conceptually, as there is less time to evacuate before the storm arrives. Nevertheless, this is the first study (to our knowledge) to begin quantifying the impact of these errors on evacuation outcomes (RQ1), and to suggest that reducing these errors should translate to improved evacuations.

Impact of reduced track errors. In this section, we change from OFCL to best track as the simulated forecast, and modify NHC's cone of uncertainty to sizes representative of track errors today (2022) and in the past (2007) and examine their impact on FLEE's evacuations (RQ2). When results are averaged across the model grid, a simulation with the 2022 cone results in 210,000 fewer evacuees than a simulation with the 2007 cone (Table 2, column 2). The reduction is most pronounced along Florida's east coast (less-impacted areas), where the smaller 2022 cone results in a 1%–7% reduction in evacuation rates, while evacuation rates across west Florida (most impacted areas) remain the same (Fig. 3e). Backing up this idea, evacuation rates in the coastal <64 kt zone decrease by 3.5% overall (Table 2, column 3), while in the coastal >64 kt zone they increase by 0.4% (Table 2, column 5). The results makes sense when considering the light system forecasts for the experiments (supplemental Figs. S1 and S2), as the 2022 cone is quicker to reduce risk across Florida's east coast as the storm approached, while forecast risk in west Florida remains the same between the 2007 and 2022 cones.

The story is similar with the hypothetical Dorian scenario: while evacuation rates and traffic remain similar across impacted areas between the simulations with the 2007 and 2022 cones, the smaller 2022 cone leads to significantly reduced evacuation rates in less-impacted regions (Fig. S8, Table S2). For example, in this scenario, the 2022 cone reduced evacuation
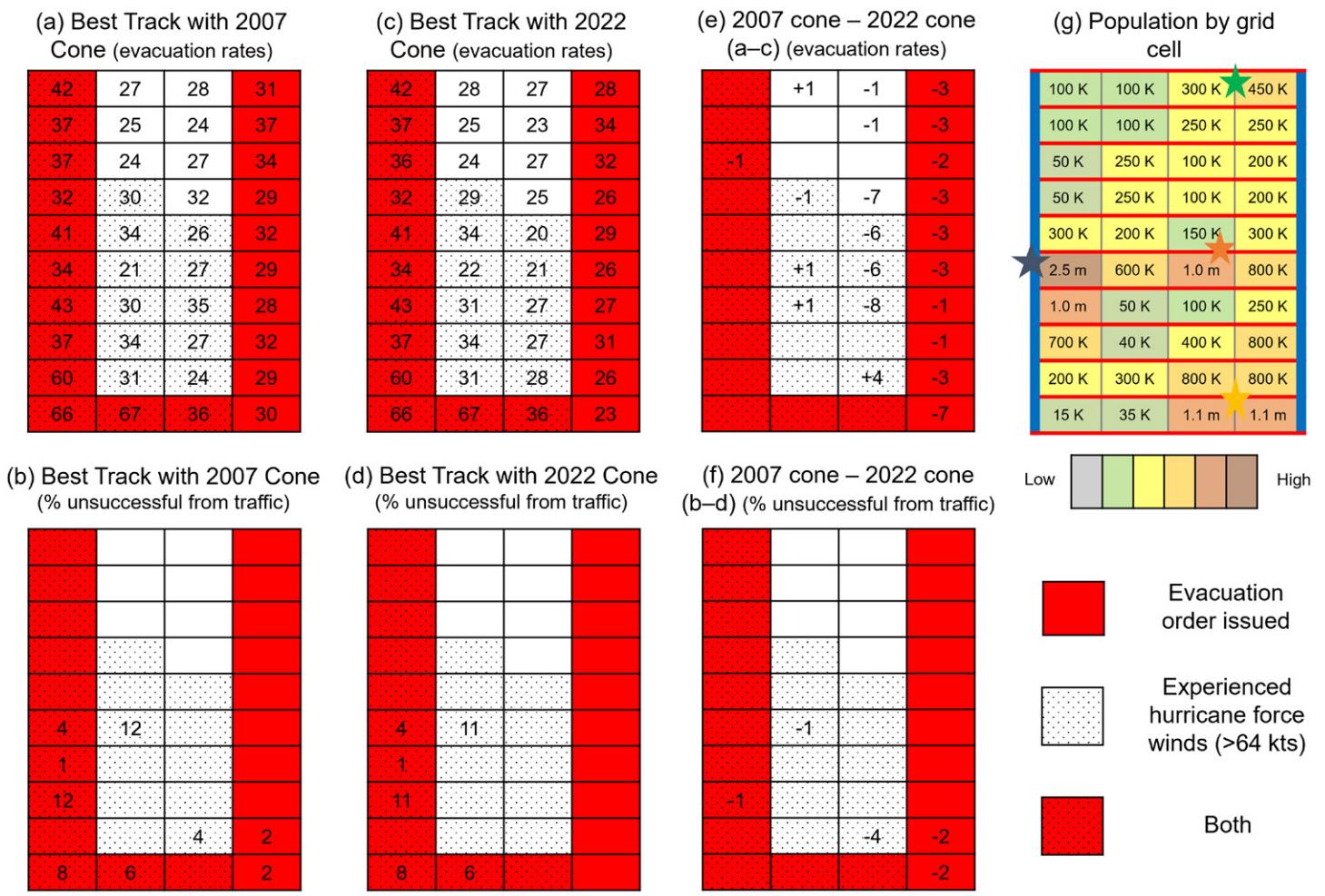
Fig. 3. Irma's simulated evacuation outcomes across grid cells for experiments using the 2007 and 2022 cones of uncertainties. (a),(b) Evacuation rates and the percentage of households who decided to evacuate but were unsuccessful due to traffic are presented for the best track with 2007 cone of uncertainty experiments. (c),(d) Similarly, for the best track with 2022 cone of uncertainty experiments, evacuation rates and the percentage of households who decided to evacuate but were unsuccessful due to traffic are presented. The difference between the 2007 and 2022 cone of uncertainty cases is then shown, both for (e) evacuation rates and (f) the percent of unsuccessful evacuations due to traffic. Also shown in (a)–(f) are grid cells that experienced hurricane force winds (dotted cells) and evacuation orders (red cells). (g) To provide a frame of reference for interpreting the simulations, the populations by grid cell along with the approximate location of several major cities on the model grid are shown: Miami–Fort Lauderdale (yellow star), Tampa Bay–Saint Petersburg (blue star), Jacksonville (green star), and Orlando (orange star).

rates by 8%–17% across west Florida. This makes sense conceptually: if the average track errors are smaller, as they are with a smaller cone, there is less uncertainty as to where (and when) a storm will make landfall, and which areas will be most heavily affected. Less uncertainty causes a smaller area to be considered at risk which influences evacuation decisions, and in this case, reduces the number of people thinking they are at high enough risk to evacuate.

Despite little differences in evacuations across the most impacted areas between the 2007 and 2022 cones—which could result from using best track forecasts in these experiments—the 2022 cone reduced evacuation rates in less impacted areas, and thus an argument can be made for improved evacuation outcomes with reduced forecast track errors over the 2007–22 year period (RQ2).

Summary and looking ahead

This article demonstrates how coupled natural–human models like FLEE can provide virtual laboratories to explore how changes in hurricane forecast errors influence evacuations

across many forecast scenarios and, in doing so, provide a new and complementary verification approach to traditional metrics of forecast accuracy. To demonstrate this potential, we conduct hypothetical experiments using cones of uncertainty representative of track errors today (2022) and in the past (2007) and examine their impacts on evacuation. We also create scenarios like when a storm undergoes rapid intensification (RI) or moves faster than expected (rapid onset), and assess their impact on evacuation outcomes in FLEE. Both sets of experiments are conducted for Irma and a hypothetical version of Dorian (supplemental material) making landfall across east Florida, and provide a first-order look at the following questions:

RQ1: Do large intensity and forward speed errors, such as those in unexpected rapidly intensifying and/or poorly forecast rapid onset scenarios, negatively impact evacuations? In these experiments for both the Irma and hypothetical Dorian landfalling cases, evacuation rates decrease considerably in the most impacted areas (e.g., by 12%–16% in areas closest to Irma's landfall) while unsuccessful evacuations due to traffic increased. Though these results make sense conceptually, we begin to measure the impact of these errors on evacuation outcomes and suggest that reducing them should translate to improved evacuations.

RQ2: Do improvements in forecast track accuracy over time—as expressed through smaller cones of uncertainty—translate to improved evacuations? In the Irma and Dorian landfalling scenarios, the 2022 (smaller) cone reduced the number of (arguably unnecessary) evacuations in less impacted areas relative to the 2007 (larger) cone. Meanwhile, evacuation rates and traffic in the most impacted areas remained similar. Since the cone of uncertainty sizes represent track errors during these periods, an argument can be made for improved evacuation outcomes with reduced forecast track errors during the period of 2007–22.

Our results are not intended to provide definitive answers to the questions above; rather, in beginning to explore these ideas, we demonstrate how coupled natural–human models offer a societally relevant complement to traditional metrics of forecast accuracy, and point toward the development of more detailed natural–human models to answer these types of questions further.

Coupled natural–human models provide several opportunities for future work to address questions of interest to the meteorological community. First, models with a more sophisticated representations of forecast intensity, an increased horizontal resolution of grid cells, and faster computational speeds—which enables running additional simulations and scenarios—could better tease out the effects of track, intensity, and forward speed errors on evacuation outcomes. Second and relatedly, coupled models with similar updates could be used to explore additional verification-related questions:

- Where and when are evacuation rates most susceptible to small changes in the forecast track?
- Are there diminishing returns in terms of how improving aspects of forecast accuracy affects evacuation?
- Does human input over models and ensembles translate to evacuation success?
- Are there fundamental differences in evacuations in well forecasted RI/RO events versus poorly forecasted ones?

Third, coupled natural–human models can be extended to additional phenomenon such as tornadoes and wildfires, potentially transforming public warning and protection

scenarios in these areas. Fourth, one could look at other ways to measure evacuation success, in addition to the ones adopted here. For example, looking at the economic impacts of evacuations could be particularly interesting. Fifth, Hurricane Ian (2022) provides another opportunity to verify FLEE against real-world evacuation data. Then, changing aspects of Ian's forecasts and seeing its evacuation effects in FLEE would be interesting given Ian's significant impacts and the complex scenario it presented for decision-makers in populated regions (e.g., rapid intensification and southward shift in track prior to landfall).

Coupled natural–human models like FLEE show promise for supporting meteorology in the long term. As computing power continues to increase, and as empirical data on hurricane evacuation behaviors and traffic become more available, that information can be codified into coupled natural–human models, thus increasing their realism, and subsequently, their ability to answer verification questions of interest. This emphasizes the value of integrating expertise in social and behavioral sciences and engineering into the weather enterprise, to address questions at the intersection of these fields. By combining such knowledge, empirical-modeling studies can provide new opportunities to advance our understanding of the hurricane forecast–evacuation system, including the development of societally relevant forecast verification techniques.

Data availability statement. The commented code, an ODD specification (a formal, detailed model description), and supporting input files are available for download at the CoMSES model library (www.comses.net/codebaserelease/4cd05855-f387-48bd-8899-9d62375518cb/).

References

Baker, E., 1991: Hurricane evacuation behavior. *Int. J. Mass Emerg. Disasters*, **9**, 287–310, https://doi.org/10.1177/028072709100900210.

Cangialosi, J. P., 2019: National Hurricane Center forecast verification report: 2019 hurricane season. NHC Tech. Rep., 75 pp., www.nhc.noaa.gov/verification/pdfs/Verification_2019.pdf.

——, A. S. Latto, and R. Berg, 2018: Tropical cyclone report. Hurricane Irma (AL112017), 30 August–12 September 2017. NHC Tech. Rep., 111 pp., www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf.

Darzi, A., V. Frias-Martinez, S. Ghader, H. Younes, and L. Zhang, 2021: Constructing evacuation evolution patterns and decisions using mobile device location data: A case study of Hurricane Irma. arXiv, 2102.12600v1, https://doi.org/10.48550/arXiv.2102.12600.

Davidson, R., and Coauthors, 2018: An integrated scenario ensemble-based framework for hurricane evacuation modeling: Part 1—Decision support system. *Risk Anal.*, **40**, 97–116, https://doi.org/10.1111/risa.12990.

DeMaria, M., J. L. Franklin, M. J. Onderlinde, and J. Kaplan, 2021: Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, **12**, 683, https://doi.org/10.3390/atmos12060683.

FDEM, 2018: Regional Emergency Management Liaison Team. FDEM, www.floridadisaster.org/dem/directors-office/regions/.

Feng, K., and N Lin, 2021: Reconstructing and analyzing the traffic flow during evacuation in Hurricane Irma (2017). *Transp. Res.*, **94D**, 102788, https://doi.org/10.1016/j.trd.2021.102788.

Fossell, K. R., D. Ahijevych, R. E. Morss, C. Snyder, and C. Davis, 2017: The practical predictability of storm tide from tropical cyclones in the Gulf of Mexico. *Mon. Wea. Rev.*, **145**, 5103–5121, https://doi.org/10.1175/MWR-D-17-0051.1.

Harris, A. R., P. J. Roebber, and R. E. Morss, 2021: An agent-based modeling framework for examining the dynamics of the hurricane-forecast-evacuation system. *Int. J. Disaster Risk Reduct.*, **67**, 102669, https://doi.org/10.1016/j.ijdrr.2021.102669.

——, R. E., Morss, and P. J. Roebber, 2023: What improves evacuations? Exploring the hurricane-forecast-evacuation system dynamics using an agent-based framework. *Nat. Hazards Rev*., https://doi.org/10.1061/NHREFO/NHENG-1671, in press.

Huang, S.-K., M. K. Lindell, and C. S. Prater, 2016: Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. *Environ. Behav.*, **48**, 991–1029, https://doi.org/10.1177/0013916515578485.

Lindell, M. K., and R. W. Perry, 2012: The Protective Action Decision Model: Theoretical modifications and additional evidence. *Risk Anal.*, **32**, 616–632, https://doi.org/10.1111/j.1539-6924.2011.01647.x.

——, P. Murray-Tuite, B. Wolshon, and E. J. Baker, 2019*: Large-Scale Evacuation: The Analysis, Modeling, and Management of Emergency Relocation from Hazardous Areas.* 1st ed. Taylor and Francis, 346 pp.

Long, E. F., M. K. Chen, and R. Rohla, 2020: Political storms: Emergent partisan skepticism of hurricane risks. *Sci. Adv.*, **6**, eabb7906, https://doi.org/10.1126/sciadv.abb7906.

Martín, Y., S. L. Cutter, and Z. Li, 2020: Bridging Twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Nat. Hazards Rev.*, **21**, 04020003, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000354.

Morss, R. E., 2005: Problem definition in atmospheric science public policy: The example of observing-system design for weather prediction. *Bull. Amer. Meteor. Soc.*, **86**, 181–192, https://doi.org/10.1175/BAMS-86-2-181.

——, and Coauthors, 2017: Hazardous weather prediction and communication in the modern information environment. *Bull. Amer. Meteor. Soc.*, **98**, 2653–2674, https://doi.org/10.1175/BAMS-D-16-0058.1.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

NASEM, 2018: *Integrating Social and Behavioral Sciences within the Weather Enterprise.* National Academies Press, 198 pp.

Roebber, P. J., and L. F. Bosart, 1996: The complex relationship between forecast skill and forecast value: A real-world analysis. *Wea. Forecasting*, **11**, 544–559, https://doi.org/10.1175/1520-0434(1996)011<0544:TCRBFS>2.0.CO;2.

Staes, B., N. Menon, and R. L. Bertini, 2021: Analyzing transportation network performance during emergency evacuations: Evidence from Hurricane Irma. *Transp. Res.*, **95D**, 102841, https://doi.org/10.1016/j.trd.2021.102841.

Watts, J., R. E. Morss, C. M. Barton, and J. L. Demuth, 2019: Conceptualizing and implementing an agent-based model of information flow and decision making during hurricane threats. *Environ. Modell. Software*, **122**, 104524, https://doi.org/10.1016/j.envsoft.2019.104524.

Widener, M. J., M. W. Horner, and S. S. Metcalf, 2013: Simulating the effects of social networks on a population's hurricane evacuation participation. *J. Geogr. Syst.*, **15**, 193–209, https://doi.org/10.1007/s10109-012-0170-3.

Wong, S., S. Shaheen, and J. Walker, 2018: Understanding evacuee behavior: A case study of Hurricane Irma. Transportation Sustainability Research Center Rep., 72 pp., https://escholarship.org/uc/item/9370z127.

Yang, K., R. A. Davidson, B. Blanton, B. Colle, K. Dresback, R. Kolar, and T. Wachtendorf, 2019: Hurricane evacuations in the face of uncertainty: Use of integrated models to support robust, adaptive, and repeated decision-making. *Int. J. Disaster Risk Reduct.*, **36**, 101093, https://doi.org/10.1016/j.ijdrr.2019.101093.

Yi, W., L. K. Nozick, R. A. Davidson, B. Blanton, and B. A. Colle, 2017: Optimization of the issuance of evacuation orders under evolving hurricane conditions. *Transp. Res.*, **95B**, 285–304, https://doi.org/10.1016/j.trb.2016.10.008.

Yin, W., P. Murray-Tuite, S. V. Ukkusuri, and H. Gladwin, 2014: An agent-based modeling system for travel demand simulation for hurricane evacuation. *Transp. Res.*, **42C**, 44–59, https://doi.org/10.1016/j.trc.2014.02.015.