

Individualizing an Agent-Based Flood Evacuation Model from Survey Data

Tatsuki Fukuda

Faculty of Data Science, Shimonoseki City University, Shimonoseki, Japan

Abstract—Earlier agent-based models (ABMs) of flood evacuation by the author, first on an abstract grid and later extended to real 3D city-model geography, use a survey of Japanese residents to assign each simulated agent to one of a few discrete behavioral groups; within a group, the evacuation probability P and the neighbor-awareness radius L are, with one narrow exception, fixed at a single value shared by every agent. The same survey contains two individually answered questions – how much a respondent cares about neighbors’ evacuation status, and how many neighboring houses’ worth of information they want. The first was never used in that prior work, and the second only once, in isolation, on an abstract grid; both are natural candidates for individualizing P and L , respectively. This paper extends the model in four ways: whole respondent records are resampled, preserving the real correlation among a respondent’s evacuation threshold, concern level, and preferred radius; the preferred-neighbor count is converted into a metric radius using the actual local building density from real PLATEAU 3D city-model data for Toyohashi City; the persistent Evacuate-Now-Button regime and a direct-observation regime are unified within one model; and a real-geometry video visualization is added. Re-deriving the group distribution from the survey matches the previously published values to within 1.5 percentage points, and the observation regime reproduces the prior “no-system” rate (9.6% vs. $\approx 9\%$) as an emergent consequence rather than a calibrated target; an ablation shows that individualizing P alone, under an assumed concern-to-probability mapping, shifts the final evacuation rate by nearly 18 points (79.9% vs. 62.0%). Throughout, the parameters that remain researcher-declared assumptions are distinguished from those derived from data.

Keywords—Agent-based model; flood evacuation; evacuation rate; survey-based simulation; evacuate now button

I. INTRODUCTION

Japan is exposed every year to typhoons and heavy-rain events, and despite the routine issuance of evacuation advisories, the fraction of at-risk residents who actually evacuate remains low. A nationwide survey found that even under a Level 4 alert (nominally an evacuation order), only 16.9% of residents at home evacuated to a location outside their home, and for hydro-meteorological hazards specifically the share taking evacuation action was only 5.7% [1]. In the June 2023 heavy rain event in Toyohashi City, Aichi Prefecture, which serves as the case study of this paper, the recorded evacuation rate was approximately 0.068% [2].

Several factors have been identified as contributing to this persistently low evacuation rate in Japan: many residents do not evacuate despite receiving official information [3]; personality traits such as being “cautious and proactive” rather than “careless and proactive” predict earlier action [4]; prior

risk awareness promotes evacuation [5]; and the timing and channel of information delivery measurably affect uptake, as shown for the July 2018 Gifu Prefecture heavy rain [6] and for landslide evacuation [7]. Separately, methodological work has examined how to estimate flood evacuation rates from disaster records [8].

A body of prior work by the author has approached this problem with agent-based modeling (ABM), motivated by the observation that Japanese residents’ evacuation decisions are strongly conditioned on whether their neighbors have evacuated [9]. This is consistent with the classic sociological threshold model of collective behavior [10], in which each individual has a personal threshold – the fraction of others who must already have acted before that individual acts – and the population-level dynamics follow from the distribution of thresholds across individuals; large-scale mobile-phone location data show a similarly sharp threshold pattern in real evacuation behavior after earthquakes, with evacuation probability rising sharply once measured seismic intensity crosses a critical level [11]. Building on the “Evacuate Now Button” concept, first proposed in an earlier technical report [12] – a physical button pressed at each household to share residents’ evacuation status – prior work [9] used a first ABM prototype, on an abstract (non-geographic) grid, to explore how the number of neighbors whose status is shared affects the resulting evacuation rate. This was substantially extended [13] with real PLATEAU 3D city-model geography, an explicit warning-level schedule (Levels 3/4/5) reproducing a real disaster timeline, and a two-group behavioral model (an alert-only Group I and a neighbor-rate-triggered Group II, further split into five threshold sub-groups) calibrated from the same survey of 1,000 respondents used in [9]. The Evacuate Now Button concept was then implemented [14] as a working IoT prototype and its practical feasibility measured (end-to-end latency, delivery reliability under simultaneous presses, battery life), providing empirical grounding for the assumption – used throughout the ABM line of work, including this paper – that a button-based system can share evacuation status quickly and reliably.

This paper’s contribution is to make fuller use of the same underlying survey than these two prior papers [9], [13] did. The later extension [13] uses the survey only to assign discrete threshold groups, applying a single assumed evacuation probability ($P = 0.5$, not itself estimated from the survey) uniformly to every agent; [9]’s earlier prototype instead set separate, survey-derived probabilities for its alert-only and neighbor-threshold groups, but neither prior paper varies P within a group by an individual respondent’s stated concern, which is

the specific individualization Q7 enables here. Both papers also use, in almost every experiment, a single, population-wide neighbor-awareness parameter L (swept as an experimental condition in [13]; individualized as a per-agent neighbor *count*, but only in one isolated experiment of [9], on an abstract, non-geographic grid). The survey, however, also contains two further single-select questions – one asking how much a respondent cares about neighbors’ evacuation status, which neither prior paper uses at all, and one asking how many neighboring houses’ worth of information they would want, which the earlier prototype [9] uses only in that one isolated experiment. These two questions are used here to individualize, respectively, the evacuation probability P and the awareness radius L for each simulated agent; resampling whole respondent records (rather than each question’s marginal distribution independently) preserves, by construction, the real correlation between an agent’s threshold sensitivity, stated concern, and preferred information radius. The “information regime” an agent operates under – a persistent, always-accurate Evacuate-Now-Button-style regime on one hand, and a regime with no such system, in which agents can only directly notice a recent, nearby evacuation, on the other – is further generalized into a single switchable parameter of the same individualized model, rather than two separately implemented mechanisms. Finally, a visualization pipeline is added that renders one stochastic realization of the simulation directly onto real PLATEAU building footprints as a video, with a live numeric summary panel.

Throughout, a careful distinction is drawn between two kinds of improvement over prior work: 1) replacing a previously unfounded or population-wide assumption with a quantity estimated directly from previously unused survey data, and 2) methodological or software-engineering improvements (reproducibility, open implementation, explicit separation of data-derived and assumed parameters) that do not by themselves improve predictive accuracy. Several parameters – the base evacuation probability, the per-evaluation discount rate, and the direct-observation range and recency window – remain researcher-declared assumptions inherited from [13], because the survey contains no question from which they could be estimated, together with the CareScale mapping newly introduced here (Section IV-A); this is reported explicitly rather than presenting the whole model as empirically calibrated.

The remainder of this paper is organized as follows. Section II reviews the three prior papers this work extends and situates the contribution relative to each. Section III describes the survey and the derivation of the group-threshold distribution. Section IV presents the individualization of P and L . Section V describes the two information regimes. Section VI describes the real-geography case study. Section VII gives the simulation algorithm. Section VIII reports experimental results. Section IX discusses the findings, and Section X states limitations, followed by a conclusion in Section XI.

II. RELATED WORK

Beyond the three papers this work directly extends, ABM, a general methodology for simulating heterogeneous individuals with simple behavioral rules [15], has been applied extensively to flood evacuation and related behavior, including social-media and individual influences on evacuation decisions [16],

flood-event inhabitant behavior [17], mutual-aid mechanisms in vulnerable communities [18], flash-flood warning response [19], flood-pedestrian interaction dynamics [20], [21], flood-induced displacement forecasting [22], community flood-risk mitigation [23], coastal migration under future flood risk [24], and household adoption of private flood-mitigation measures [25]; see [26] for a broader review of agent-based flood risk models. To the author’s knowledge, none of these models individualizes both an evacuation probability and an awareness radius from the same joint survey record, or unifies a persistent-sharing and a direct-observation information regime within one model, as is done here.

A parallel strand models the physical dynamics of pedestrian and crowd evacuation rather than the decision to evacuate: cellular-automaton and continuous crowd models capture movement effects such as turning behavior on clearance time [27], controlled experiments have measured human motion speed in floodwaters [28], and recent reviews survey a decade of progress in agent-based and pedestrian evacuation simulation [29], [30]. The model developed here deliberately abstracts away physical movement – representing only the evacuate-or-stay decision and its timing – so this body of work is complementary to, rather than in competition with, the present approach.

Most closely related is work on social influence in evacuation, since the agents’ core mechanism is neighbor-conforming threshold behavior [10]. Imitation of cooperative behavior has been incorporated into pedestrian evacuation models and shown to reshape clearance dynamics [31]. At the individual level, evacuation decisions have been predicted from pre-disaster web-search behavior [32] and from large-scale mobile-phone location traces [11], indicating that heterogeneity in individual evacuation propensity is both real and, to a degree, measurable – which is precisely the heterogeneity that the survey-based individualization of P and L presented here aims to capture.

An initial ABM prototype based on the Evacuate Now Button concept, first proposed in [12], was presented in earlier work [9], which framed the evacuate-or-stay decision using game theory [33] and grounded the neighbor-conforming behavior in bandwagon-effect theory [34]. Agents were placed at random screen coordinates (no real geography). A survey of 1,000 respondents asked, among other things, which of several conditions – official alert levels near home or in a neighboring town, or five neighbor-evacuation-rate levels (10/30/50/80/almost-all%), or “I don’t care about others” – would make the respondent evacuate; this is the same question used in Section III. Ninety percent of agents were assigned to a neighbor-rate-driven behavior and 10% to an alert-only behavior; both groups’ probabilities were set directly from this same survey question rather than assumed – the alert-only group’s from the official-alert responses (18.6% at Level 3, 79.6% at Level 4), and the neighbor-rate-driven group’s from the corresponding neighbor-rate responses (e.g., 50% of respondents said they would evacuate once half of their neighbors had). A separate survey question asked how many neighboring houses’ status a respondent would want to know (10/25/50/100/200); in one experimental condition, this distribution (46%/29%/15%/7%/3%, respectively) was used to assign each agent an individual *count* of neighbors to observe,

though on the abstract, non-geographic grid this count carried no notion of physical distance, and it was tested only as one alternative to several fixed values (0, 10, 20, 50, 100, 200), not integrated with any other individual-level variation. (This split is the raw Q11 marginal reported in Section III – 46.0/29.1/15.0/6.6/3.3% – rounded to integer percentages, confirming that the earlier prototype [9] drew on the same question.) This question is built on directly in Section IV-B, where it is converted into a metric radius using real building density and combined, via joint resampling (Section IV-C), with the same agent’s threshold group and stated concern level.

A subsequent extension replaced the abstract grid with real PLATEAU 3D city-model geography (Nobeoka City and Toyohashi City) and a real warning-level timeline reproducing an actual disaster [13]. Agents were split into an alert-only Group I ($\approx 8\%$ of respondents) and a neighbor-rate-driven Group II, itself split into five threshold sub-groups (10/30/50/80/100%, in proportions 20%/31%/32%/12%/5%, Table I therein). Every agent used the same fixed evacuation probability ($P = 0.5$) and the same awareness radius L within a given experiment, with L swept from 20 to 100 m across experiments to study its effect on the final evacuation rate, and a separate, simplified “without system” baseline (observing only evacuations within 40 m and the last 5 minutes, with no threshold-group structure) was implemented once for comparison. Keeping P and L population-wide fit that design’s aim – a controlled sweep of one macro-level parameter, not a search for individual-level heterogeneity – and neither the radius conversion of Eq. (2) nor the joint-resampling of Section IV-C, both introduced here, had yet been applied there to turn Q11 and Q7 into individualized, mutually consistent parameters. In the present paper, the Group I/II(a–e) proportions are reproduced independently from the raw survey data (Section III); both P and L are individualized per agent instead of using one shared value per experiment (Section IV); and the “without system” baseline is re-implemented as one setting of a single, unified information-regime parameter that uses the same per-agent threshold groups as the main model (Section V), rather than as a separate, simplified mechanism.

A companion hardware study implemented the Evacuate Now Button as a working prototype (ESP-WROOM-02 plus a cloud pipeline) and measured end-to-end latency (median 1.9 s, 2.3 s under eight near-simultaneous presses), delivery reliability (zero losses in 200 simultaneous presses, 95% upper bound on loss rate 1.5%), and battery life (>30 months under several usage scenarios) [14]. That study does not modify the simulation model, but its results directly support the *button* regime of Section V: the regime’s core assumption – that a button press becomes known, essentially instantaneously and reliably, to a shared aggregation service – is technically achievable, not merely theoretical.

III. SURVEY AND THE GROUP-THRESHOLD DISTRIBUTION

The survey (Lancers online panel, $n = 1000$, ages spanning teens to 80s) is the same questionnaire referenced in [9], [13]. This paper uses three of its single- or multi-select items:

- **Q6** (multi-select checkbox): “If the following situation occurred while you were at home during heavy rain,

would you evacuate to a shelter? Select all that you think would make you evacuate.” Eleven options mix five official-alert conditions (not used here) with five neighbor-evacuation-rate levels (10/30/50/80%, almost all) and an “I don’t care about others” option.

- **Q7** (single-select): “When deciding whether to evacuate during heavy rain, do you care whether the houses around yours have evacuated?” – *cares* (501), *cares a little* (420), *doesn’t care at all* (79).
- **Q11** (single-select): “If you were to refer to how many neighboring houses have evacuated, how many would you want data for?” – 10 (460), 25 (291), 50 (150), 100 (66), 200-or-more (33).

A. Group Classification Rule

Because Q6 allows multiple selections, a rule is needed that maps a respondent’s checked options to a single effective threshold. A respondent is classified as Group I if they checked “I don’t care about others” and none of the five numeric levels. If they checked at least one numeric level, the *lowest* checked level is taken as their effective threshold (Group II(a–e)): checking a higher level is consistent with also being willing to evacuate at a lower one, so the minimum checked level is the most informative single estimate of the respondent’s true trigger point. Respondents who checked neither a numeric level nor “I don’t care” (i.e., who only responded to the alert-based options, or left the relevant options blank) provide no information about a neighbor-rate threshold and are excluded from the distribution.

B. Independent Reproduction of a Prior Group-Threshold Distribution

Applying this rule to the raw data yields the distribution in Table I. Of 1,000 respondents, 758 (75.8%) are classifiable; the remaining 242 (24.2%) are excluded as described above. Table I of [13] reports Group I $\approx 8\%$, and, within Group II, sub-group proportions of 20%, 31%, 32%, 12%, and 5% for the 10%, 30%, 50%, 80%, and 100% thresholds – but does not describe how these figures were derived from the raw multi-select responses. The classification rule was not fit to these published numbers; the rule above was fixed on the multi-select data itself, independent of the target values. Computed on the same basis as [13] (Group I as a share of all classified respondents, and the Group II sub-groups as shares within Group II), the resulting distribution matches [13]’s Table I to within 1.5 percentage points (the largest single deviation is Group I, 9.5% vs. $\approx 8\%$; every Group II sub-group agrees to within 0.6 points). This agreement is reported as an independent check that the (fully specified, reproducible) rule used here is likely to recover the procedure used in [13], which was not documented there, rather than as a claim of numerical novelty. The 24.2% of respondents who provide no usable answer to the neighbor-rate item are not discussed in [13]; their exclusion, rather than assignment to Group I, is treated as an explicit, open modeling choice (Section X).

TABLE I. GROUP-THRESHOLD DISTRIBUTION DERIVED INDEPENDENTLY FROM THE RAW SURVEY DATA ($n=758$; GROUP II ROWS: $n=686$), COMPARED WITH [13]'S TABLE I. RAW COUNTS IN PARENTHESES

Group	Threshold	This work	[13]
Group I ("don't care")	—	9.5% (72)	$\approx 8\%$
Group II(a)	10%	20.6% (141)	20%
Group II(b)	30%	30.5% (209)	31%
Group II(c)	50%	31.6% (217)	32%
Group II(d)	80%	12.0% (82)	12%
Group II(e)	almost all (100%)	5.4% (37)	5%

IV. INDIVIDUALIZING EVACUATION PROBABILITY AND AWARENESS RANGE

A. Evacuation Probability P via $Q7$

In [13], every agent uses the same evacuation probability $P=0.5$ at every eligible evaluation. P is instead scaled per agent, for Group II agents only, using that agent's bootstrapped answer to $Q7$:

$$P_{\text{agent}} = P_{\text{base}} \times \text{CareScale}(Q7 \text{ answer}) \quad (1)$$

where, $\text{CareScale} \in \{1.0, 0.6, 0.25\}$ for *cares*, *cares a little*, and *doesn't care at all*, respectively. Group I agents are unaffected, since $Q7$ asks specifically about neighbor status, which Group I agents do not use. The three scale values in Eq. (1) are a declared modeling assumption, not a measured quantity: no survey question asks directly how likely a respondent is to act on their stated intention, so converting an ordinal concern level into a probability multiplier necessarily involves a researcher-chosen mapping, anchored at 1.0 for the "cares" response. What is new is the use of a previously unused survey item to introduce principled, data-linked individual variation in P in place of a single population-wide constant; the specific numeric mapping is a target for future calibration (Section X).

B. Awareness Radius L via $Q11$ and Real Building Density

The later extension [13] sweeps a single radius $L \in \{20, 40, 60, 80, 100\}$ m, shared by every agent, across separate experiments. The earlier prototype [9] individualizes a neighbor count per agent from the same $Q11$ -style question, but on an abstract grid with no notion of metric distance. L is individualized here by converting each agent's bootstrapped $Q11$ answer (a preferred house count N_i) into a metric radius L_i using the real local building density ρ (households per m^2) of the loaded case-study dataset, assuming households are uniformly distributed at that density:

$$L_i = \sqrt{\frac{N_i}{\pi\rho}}. \quad (2)$$

For the Toyohashi case study of Section VI ($\rho \approx 1,752$ households/ km^2), Eq. (2) maps the five $Q11$ answers to radii of 42.6, 67.4, 95.3, 134.8, and 190.6 m, respectively – comparable in magnitude to, but individually variable around, the single fixed values (20–100 m) swept in [13]. Table II gives the resulting per-agent radius distribution over the bootstrap pool: it is strongly weighted toward the two smallest radii (72.7% of the 758-respondent bootstrap pool at ≤ 67 m; median 67.4 m,

mean ≈ 71 m), which is why the individualized-button result (62.0%) falls between the fixed $L=60$ m and $L=80$ m results (Section VIII-B).

TABLE II. PER-AGENT AWARENESS RADIUS INDUCED BY $Q11$ (EQ. (2)) OVER THE BOOTSTRAP POOL OF 758 CLASSIFIABLE RESPONDENTS

Q11 answer (houses)	Radius L_i [m]	Share of agents
10	42.6	41.7%
25	67.4	31.0%
50	95.3	16.4%
100	134.8	7.3%
200	190.6	3.7%

C. Joint Bootstrapping of $Q6$, $Q7$, and $Q11$

All three questions above are recorded for the same 758 respondents classifiable on $Q6$ ($Q7$ and $Q11$ are mandatory single-select fields answered by all 1,000 respondents). Rather than sampling each agent's group, care level, and preferred house count from three independent marginal distributions, whole respondent rows are resampled with replacement, so that whatever correlation exists in the real data among threshold sensitivity, stated concern, and preferred information depth is carried into the simulated population. Neither [9] nor [13] resamples joint respondent profiles in this way; both treat group assignment, P , and L as independent parameters (or population-wide constants).

V. INFORMATION REGIMES: PERSISTENT SHARING VS. DIRECT OBSERVATION

A Group II(a–e) agent becomes eligible to evacuate once "the evacuation rate of surrounding households" reaches its personal threshold, but what that phrase means depends on whether a button-style sharing system exists.

Let $\mathcal{N}_i = \{j \neq i : d_{ij} \leq L_i\}$ be the set of other households within agent i 's radius, where d_{ij} is the metric distance between households. In both regimes below, a rate is defined as 0 for an agent whose radius contains no other household ($|\mathcal{N}_i| = 0$), avoiding division by zero; such a Group II agent then never becomes eligible (Group I is unaffected, since it never consults neighbor information – Section IV). This is rare under the individualized radii of Eq. (2) and the observation regime's fixed $L_{\text{obs}}=40$ m (at most 0.4%, Section VI), but reaches about 4% at the smallest fixed- L condition tested ($L=20$ m, Section VIII-A).

1) *Button regime*: As in [13], a management center aggregates every button press, so at every evaluation an agent knows the true, persistent ("stock") evacuation rate among other households within its radius L_i (Section IV-B):

$$r_i^{\text{stock}}(t) = 100 \times \frac{|\{j \in \mathcal{N}_i : \text{evac}_j(t)\}|}{|\mathcal{N}_i|}, \quad (3)$$

where, $\text{evac}_j(t)$ indicates that j has evacuated by time t . This is the regime whose real-world feasibility is supported by the latency and reliability measurements of [14].

2) *Observation regime*: With no sharing system, an agent can only directly notice a neighbor's evacuation while it is happening – e.g., seeing them leave through a window – within a limited range and only if the agent happens to check during that narrow window. This is modeled as follows: at every T_1 check, the agent computes a recent (“flow”) rate over a fixed range L_{obs} (so $\mathcal{N}_i^{\text{obs}} = \{j \neq i : d_{ij} \leq L_{\text{obs}}\}$),

$$r_i^{\text{flow}}(t) = 100 \times \frac{|\{j \in \mathcal{N}_i^{\text{obs}} : t-W \leq \text{evac_time}_j \leq t\}|}{|\mathcal{N}_i^{\text{obs}}|}, \quad (4)$$

where, evac_time_j is the time at which j evacuated, and compares this recent rate (rather than the persistent stock rate of the button regime) against the same Q6-derived personal threshold. Because r^{flow} counts only evacuations in the trailing W -minute window it is almost always far smaller than r^{stock} , which is the mechanism behind the two regimes' very different outcomes (Section VIII).

Following [13]'s “without system” baseline, L_{obs} is set to 40m and W to 5min; as with P_{base} and the discount rate α (Section VII), these two values are not estimable from the survey and are adopted, as declared assumptions, from that prior baseline rather than derived from data. Because L_{obs} is specifically about *direct observation range* rather than *preferred aggregated-information depth*, the Q11-derived individualized radius is not reused here; Q11 presupposes a sharing system, which is precisely what the observation regime lacks.

Group I agents are unaffected by the choice of regime, since by construction they decide “on their own,” independent of neighbor information, in both cases.

VI. CASE STUDY: REAL BUILDING GEOMETRY AND A REAL WARNING TIMELINE

The same case-study city as [13], Toyohashi City, Aichi Prefecture, is used here for two reasons: it lets the present results be read alongside the prior paper's, and its June 2023 heavy-rain event provides real Warning Level 3/4/5 issuance times (06:40, 14:40, and 16:30 on June 2, 2023, respectively [2], [13]) that are reused as the simulation's warning schedule, rather than an arbitrary synthetic timeline.

Building positions are obtained directly from Project PLATEAU's public FY2023 CityGML package for Toyohashi [35], distributed under CC-BY 4.0. Only the needed building-model mesh tiles ($\approx 1 \text{ km} \times 1 \text{ km}$ JIS X0410 tertiary mesh cells) are fetched via HTTP range requests against the remote ZIP's central directory, avoiding the full $\approx 850 \text{ MB}$ city-wide archive. Of the city's 288 such tiles, the one with the largest residential building count is used, mesh 52371229 (centered near 34.77°N , 137.37°E , in central Toyohashi), which contains 1,856 residential buildings (bldg:usage codes 411/412/413/414/415, i.e. detached houses, apartment buildings, and mixed-use residential buildings) within a bounding box of approximately 1.06 km^2 , giving a local density $\rho \approx 1,752$ households/ km^2 . Each building's centroid and footprint (LOD0 roof-edge geometry, with a small placeholder square for the minority lacking it) are projected to local metric coordinates via an equirectangular approximation, adequate at this sub-10km scale. One agent is instantiated per residential

building, following the same convention as [13]; each building – including apartment and mixed-use buildings that may house multiple households – is therefore treated as a single household for the purposes of ρ and Eq. (2). Computed directly from this building layout, the fraction of households with zero neighbors within range is at most 0.4% for the individualized Q11-derived radii of Table II (minimum 42.6m) and for the observation regime's fixed $L_{\text{obs}}=40 \text{ m}$.

VII. SIMULATOR

The simulator is implemented in Python (NumPy/pandas for the model; lxml/Shapely for geometry parsing), in contrast to the Unity/C# implementation of [13]. Agents are represented as parallel NumPy arrays rather than individual objects, since every per-tick operation is naturally vectorized across all agents due for evaluation at that tick.

1) *Timing*: As in [13], each agent re-evaluates whether to evacuate every $T_1 \sim U(7.5, 22.5) \text{ min}$ ($15 \text{ min} \pm 50\%$), where $U(a, b)$ denotes the uniform distribution on the interval $[a, b]$, after an initial random delay $T_2 \sim U(15, 45) \text{ min}$ ($30 \text{ min} \pm 50\%$) from the start of the simulation. The simulation clock advances in discrete steps of $\Delta t = 1 \text{ min}$, beginning at Warning Level 3 ($t=0$) and ending at Level 5 ($t=T_{\text{end}}$, i.e. 590 min later). Level 4's issuance time (Table III) is shown on the temporal figures for reference only; it does not itself alter agent behavior in Algorithm 1.

2) *Decision rule*: At each eligible evaluation, a Group I agent is always eligible once the simulation has started; a Group II(a–e) agent is eligible only if the neighbor rate (Section V) meets or exceeds its personal threshold. An eligible agent evacuates with probability P_{agent} (Eq. (1)); if it does not, its probability is discounted in place for its next eligible evaluation, $P_{\text{agent}} \leftarrow \alpha P_{\text{agent}}$ (the multiplicative discount of [13], Eq. (1)), with $\alpha=0.8$. Both $P_{\text{base}}=0.5$ and $\alpha=0.8$ are carried over unchanged from [13] as declared assumptions, since the survey contains no question from which either could be estimated.

3) *Neighbor-rate computation*: At each tick, for every agent due for evaluation, pairwise distances to all other agents are computed and thresholded at that agent's own radius (individualized, Eq. (2), in the button regime; fixed at L_{obs} in the observation regime) to obtain the relevant neighbor set, from which the persistent or recent evacuation rate is computed as described in Section V. Algorithm 1 summarizes one stochastic run.

Table III consolidates every model parameter together with its value and its provenance. This is deliberate: a recurring theme of this paper is that only some of the model is estimated from data, and the table makes the boundary explicit at a glance. Two quantities – the Group distribution and the per-agent radius – are estimated directly from survey questions Q6 and Q11, respectively; Q7 supplies each agent's concern level, which is then converted into a per-agent probability multiplier via the assumed CareScale mapping (Eq. (1)); the local building density and the warning-issuance times come from real PLATEAU and disaster-record data; and the remaining five (P_{base} , α , the CareScale mapping, L_{obs} , W) are declared assumptions – four carried over unchanged from [13], and the CareScale mapping newly introduced in this paper – none of

TABLE III. MODEL PARAMETERS AND PROVENANCE. “SURVEY” = FROM THE QUESTIONNAIRE; “REAL DATA” = PLATEAU OR DISASTER RECORDS; “ASSUMED” = RESEARCHER-DECLARED; [13] = TIMING PARAMETERS UNCHANGED FROM THAT PAPER

Symbol	Meaning	Value	Provenance
Group dist.	Group I / II(a-e) shares	Table I	Survey (Q6)
CareScale	concern $\rightarrow P$ multiplier	{1.0, 0.6, 0.25}	Assumed
P_{base}	base evac. probability	0.5	Assumed
α	per-eval. discount	0.8	Assumed
T_1	re-evaluation interval	15 min $\pm 50\%$	[13]
T_2	initial delay	30 min $\pm 50\%$	[13]
L	awareness radius (button)	per agent, Eq. (2)	Survey (Q11)
ρ	local building density	1,752 /km ²	Real data
L_{obs}	observation range	40 m	Assumed
W	observation window	5 min	Assumed
$L3/L4/L5$	warning-issuance times	06:40/14:40/16:30	Real data

Algorithm 1 One stochastic run of the ABM (main individualized-button and observation regimes; the fixed- L experiments of Section VIII-A replace line 2’s radius with a constant L shared by all agents).

```

1: init: bootstrap each agent  $i$  a survey record  $(g_i, c_i, N_i)$ 
   ( $g_i$ : Q6 group,  $c_i$ : Q7 concern,  $N_i$ : Q11 house count)
2:  $\text{thr}_i \leftarrow \text{threshold}(g_i)$ ;  $L_i \leftarrow \sqrt{N_i}/(\pi\rho)$  [button] or  $L_{\text{obs}}$  [obs.]
3:  $P_i \leftarrow P_{\text{base}} \cdot \text{CareScale}(c_i)$  if  $g_i \in \text{II}$  else  $P_{\text{base}}$ 
4:  $\text{evac}_i \leftarrow \text{FALSE}$ ;  $\tau_i \leftarrow \text{draw } T_2$ 
5: for  $t = 0, \Delta t, \dots, T_{\text{end}}$  do
6:    $D \leftarrow \{i : \neg \text{evac}_i \wedge \tau_i \leq t\}$   $\triangleright$  agents due
7:   for  $i \in D$  do
8:      $r_i \leftarrow \text{neighbor rate within } L_i$  (stock [button] / last-  

        $W$  flow [obs.])
9:     if  $g_i \in \text{I}$  or  $r_i \geq \text{thr}_i$  then
10:      if  $\text{rand}() < P_i$  then
11:         $\text{evac}_i \leftarrow \text{TRUE}$ ; record time  $t$ 
12:      else  $P_i \leftarrow \alpha P_i$   $\triangleright$  decline  $\Rightarrow$  discount
13:      end if
14:    end if
15:    if  $\neg \text{evac}_i$  then  $\tau_i \leftarrow t + \text{draw } T_1$ 
16:    end if  $\triangleright$  re-check later
17:  end for
18: end for

```

which the survey can constrain directly; the timing parameters T_1 and T_2 are likewise inherited unchanged from [13].

VIII. EXPERIMENTS AND RESULTS

Unless otherwise noted, results are the mean $\pm$ one standard deviation of the final (Level 5) evacuation rate over 50 independent Monte Carlo replications, all on the Toyohashi case study of Section VI. All conditions and their exact figures are produced by a single reproducible experiment pipeline, so that every number below can be regenerated directly from the underlying data.

A. Effect of a Fixed, Population-Wide L (Button Regime)

Reproducing [13]’s experimental design of sweeping a single radius shared by every agent (with P individualized from Q7 via Eq. (1), as in all button conditions except the flat- P ablation of Section VIII-C), using the same five points

$L \in \{20, 40, 60, 80, 100\}$ m, yields final evacuation rates of $32.0\% \pm 2.8$, $48.6\% \pm 5.0$, $58.9\% \pm 6.4$, $65.7\% \pm 7.5$, and $70.3\% \pm 6.3$, respectively – confirming, in this implementation, the same qualitative relationship reported in [13]: a larger awareness radius yields a higher final evacuation rate. A larger radius gives higher-threshold Group II agents a bigger, more stable neighbor pool, and avoids leaving the 4% of Group II agents with no neighbor at $L=20$ m permanently ineligible (Section V) – though with diminishing returns once the pool is large enough for these effects to saturate. This reproduces only [13]’s radius-sweep design, not its treatment of P : [13] held P fixed at 0.5 for every agent, whereas P here remains individualized via Q7, so the absolute rates above are not directly comparable to [13]’s own figures.

B. Individualized L and P (Button Regime)

With L individualized per agent from Q11 (Eq. (2)) and P individualized per agent from Q7 (Eq. (1)), the final evacuation rate is $62.0\% \pm 7.9$, between the fixed- $L=60$ m and $L=80$ m results (58.9% and 65.7%) and slightly closer to the former, consistent with the Q11-induced radius distribution of Table II (median 67.4 m, mean ≈ 71 m).

C. Ablation: The Effect of Individualizing P

To isolate how much of this result is attributable to individualizing P specifically (as opposed to L), the individualized- L condition was re-run with the Q7 scaling of Eq. (1) switched off (every agent uses the flat $P_{\text{base}}=0.5$ regardless of its Q7 answer, while L remains individualized from Q11). This flat- P ablation yields $79.9\% \pm 4.8$ – almost 18 percentage points higher than the 62.0% obtained with Q7-scaled P . This gap is expected given Eq. (1)’s CareScale values: only 50.1% of all 1,000 respondents on Q7 (the *cares* answer) keep the full P_{base} , while 42.0% (*cares a little*) and 7.9% (*doesn’t care at all*) are scaled down to 0.6 and 0.25 of it, respectively. These shares are of the full Q7 sample; CareScale itself is applied only to the Group II agents within the 758-respondent bootstrap pool (Sections IV-A and IV-C). Individualizing P via Q7 therefore mechanically lowers the population-average evacuation probability relative to the flat- $P=0.5$ assumption of [13]. This ablation is reported, rather than only the individualized result, because it shows that Q7 individualization is not a second-order refinement on top of L individualization: in this case study, it changes the final evacuation rate by more than any single step of the L sweep in Section VIII-A.

D. Information Regime: Button vs. Observation

With the observation regime at [13]’s baseline parameters ($L_{\text{obs}}=40$ m, $W=5$ min), the final evacuation rate is $9.6\% \pm 0.7$, reaching 90% of this value by 89 min (Table VI) and rising only slightly thereafter. This closely matches the $\approx 9\%$ “without system” final rate reported in [13] for both of its case-study cities, even though the observation regime here is implemented as one setting of the same individualized, survey-derived model used for the button regime, rather than as a separately calibrated mechanism. The result is intuitive in hindsight: because the observation regime only ever reveals a small, recent, “flow” rate rather than a persistent, cumulative one, even the low-threshold Group II sub-groups rarely become eligible (Section VIII-F), so the final rate is dominated

almost entirely by Group I's alert-only evacuations (9.5% of the population per Table I), closely tracking the observed 9.6%. Because Group I agents are unaffected by CareScale (Section IV-A), this comparison with [13] is not subject to the caveat on absolute rates noted in Section VIII-A. Fig. 1 plots a representative subset of conditions together (the full L sweep is in Table V).

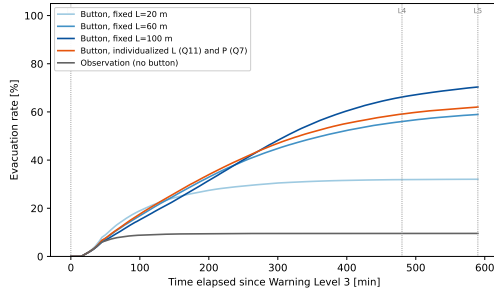


Fig. 1. Evacuation rate over time for representative conditions – fixed radii $L \in \{20, 60, 100\}$ m, individualized L and P , and the observation regime ($L_{\text{obs}}=40$ m, $W=5$ min) – on the Toyohashi case study. Each curve is the mean of 50 Monte Carlo replications; vertical dotted lines mark Warning Levels 3, 4, and 5.

E. Sensitivity of the Observation Regime to L_{obs} and W

Because L_{obs} and W are declared assumptions rather than survey-derived quantities (Section V), each was swept one-at-a-time around [13]'s baseline (Table IV). Two patterns emerge. First, the final rate is relatively insensitive to both parameters compared to the button regime's sensitivity to L (Section VIII-A): even at the extremes tested, the observation regime stays within roughly 2 percentage points of its 9.6% baseline, versus a 38-point range for button $L \in \{20, \dots, 100\}$ m. Second, and more surprising, *larger* L_{obs} is associated with a slightly *lower* final rate (10.2% at 20 m vs. 9.1% at 60 m) – the opposite direction from the button regime, where larger L raises the rate. This is a consequence of the observation regime's rate being a *flow* rate (recent evacuations divided by all neighbors in range) rather than a *stock* rate: a single nearby evacuation is a larger fraction of a smaller neighbor pool, so a small L_{obs} makes it mechanically easier for a low-threshold Group II agent to observe an apparent rate at or above 10% from just one or two witnessed departures, even though the same event would barely move the rate computed over a larger pool. Larger W , by contrast, behaves as expected (more time in which to witness an evacuation raises the rate: 9.6% at 5 min rising to 11.4% at 15 min).

TABLE IV. OBSERVATION-REGIME SENSITIVITY TO RANGE L_{obs} AND RECENCY WINDOW W (ONE-AT-A-TIME AROUND [13]'S BASELINE, IN BOLD), $n=50$

Condition	Final evacuation rate
$L_{\text{obs}}=20$ m, $W=5$ min	$10.2\% \pm 0.6$
$L_{\text{obs}}=40$ m, $W=5$ min	$9.6\% \pm 0.7$
$L_{\text{obs}}=60$ m, $W=5$ min	$9.1\% \pm 0.6$
$L_{\text{obs}}=40$ m, $W=10$ min	$10.3\% \pm 0.8$
$L_{\text{obs}}=40$ m, $W=15$ min	$11.4\% \pm 0.9$

Table V summarizes the main conditions (excluding the sensitivity sweep of Table IV).

TABLE V. FINAL (LEVEL 5) EVACUATION RATE BY CONDITION, MEAN $\pm$ SD OVER 50 REPLICATIONS. ALL BUTTON ROWS USE Q7-INDIVIDUALIZED P (EQ. (1)) EXCEPT THE FLAT- P ABLATION

Condition	Final evacuation rate
Button, fixed $L=20$ m	$32.0\% \pm 2.8$
Button, fixed $L=40$ m	$48.6\% \pm 5.0$
Button, fixed $L=60$ m	$58.9\% \pm 6.4$
Button, fixed $L=80$ m	$65.7\% \pm 7.5$
Button, fixed $L=100$ m	$70.3\% \pm 6.3$
Button, individualized L (Q11) and P (Q7)	$62.0\% \pm 7.9$
Button, individualized L , flat P (ablation)	$79.9\% \pm 4.8$
Observation ($L_{\text{obs}}=40$ m, $W=5$ min)	$9.6\% \pm 0.7$

F. Temporal Dynamics and Per-Group Contribution

The final rates above summarize only the endpoint; the dynamics that produce them are themselves informative. Table VI reports, for each main condition, the time to reach several absolute evacuation levels and the time to reach 90% of the condition's own final rate (a settling time). Two observations stand out. First, the early phase is remarkably similar across all button conditions: every one crosses 10% at 55–70 min, essentially independent of L , because the first evacuations are dominated by the low-threshold agents (Group I and Group II(a)) that a small radius already suffices to trigger. Every condition but the flat- P ablation also crosses 25% at 138–162 min; the flat- P ablation reaches 25% markedly earlier (115 min) because removing the CareScale discount raises the per-evaluation evacuation probability of the Group II agents who answered *cares a little* or *doesn't care at all* (Section VIII-C), accelerating the cascade once it is under way. The remaining conditions diverge only later, once sustaining the cascade to the higher-threshold agents requires the larger neighbor pools that only a larger L provides. Second, the observation regime settles fastest of all (90% of its final rate by 89 min) precisely because it never engages the neighbor-triggered cascade at all: it plateaus early at the Group I level and then rises only slowly.

TABLE VI. TIME (MINUTES SINCE LEVEL 3) TO REACH ABSOLUTE EVACUATION LEVELS, AND TO REACH 90% OF EACH CONDITION'S OWN FINAL RATE. “–” MEANS THE LEVEL IS NEVER REACHED

Condition	10%	25%	50%	90% final
Button, fixed $L=20$ m	55	159	–	237
Button, fixed $L=40$ m	58	138	–	358
Button, fixed $L=60$ m	63	149	365	413
Button, fixed $L=80$ m	66	155	328	436
Button, fixed $L=100$ m	70	162	312	435
Button, individualized L , P	62	144	331	411
Button, individualized L , flat P	58	115	214	367
Observation (40 m, 5 min)	–	–	–	89

Fig. 2 decomposes the cumulative evacuation rate of the main individualized-button condition by the agents' Group. It makes the cascade explicit: Group I (alert-only) rises first, reaching 90% of its own final $\approx 9\%$ share by about 90 min, consistent with the observation regime's own 89-min milestone in Table VI since Group I dominates that regime almost entirely (Section VIII-D); the lowest-threshold neighbor group, II(a), is the single largest contributor and rises early; the mid-threshold groups II(b) and II(c) accumulate steadily throughout; and the high-threshold groups II(d) and II(e) contribute little and late, since the neighbor rate rarely

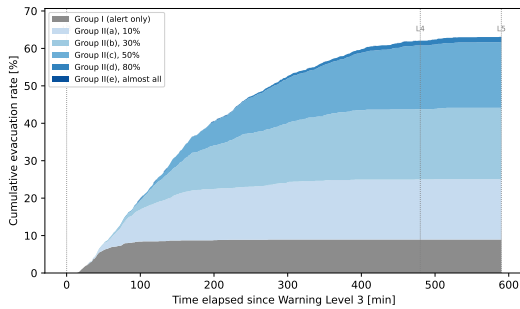


Fig. 2. Cumulative evacuation rate over time, decomposed by agent Group, for the individualized-button condition (one representative run; bands stack to the total evacuation rate).

climbs high enough to trigger them within the disaster window. This is the mechanistic picture behind both the button regime's success (low-threshold agents seed a cascade that carries the mid-threshold majority) and the observation regime's failure (without a persistent shared rate, the cascade never leaves the Group I level).

G. Visualization

Because the simulator only ever transitions an agent from not-evacuated to evacuated, a full spatial-temporal reconstruction of any single stochastic realization requires only that run's per-agent evacuation times: the state at any past time t is exactly $\{\text{evac_time} \leq t\}$. This makes it possible to render an animation without modifying the core simulation loop to record full per-tick snapshots. Fig. 3 shows one frame of such a rendering, in which each real PLATEAU building footprint is colored by its household's status; the full animation is exported as an MP4 video, together with a live numeric panel reporting the evacuation rate, evacuated count (split by alert-only vs. neighbor-triggered), current warning level, and elapsed time.

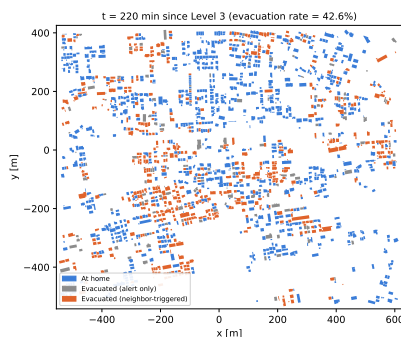


Fig. 3. One frame ($t = 220$ min since Level 3) of the video visualization, showing real PLATEAU building footprints colored by evacuation status. Spatial clustering of neighbor-triggered evacuations (orange) is visible as the simulated evacuation propagates locally.

IX. DISCUSSION

Two results are, in a precise sense, *consistency checks* rather than accuracy improvements per se: the independent reproduction of [13]'s Group distribution to within 1.5 points,

and the observation regime's 9.6% closely matching that paper's $\approx 9\%$ "without system" baseline. Both show that a more granular, more data-driven model composed from previously unused parts of the same survey behaves consistently with the coarser model's published, aggregate results, which is a necessary condition for trusting the finer-grained model but not sufficient evidence that it more accurately predicts real evacuation behavior (no such validation against real evacuation data is attempted here; see Section X). The genuinely new empirical content is the individualization itself (Section IV), which depends on survey items Q7 and Q11 that neither prior paper [9], [13] used for this purpose, and the unification of the button and observation information regimes under one model (Section V), which carries the Q7-individualized P into both regimes even though the observation regime's range is fixed rather than Q11-derived. In particular, the magnitude of the P -individualization ablation (Section VIII-C) is itself informative: an 18-point swing under the assumed CareScale mapping (Limitations item 2) shows that how Q7 is mapped to P is not a minor detail, and suggests that further covariates already present in the survey but not used here (Limitations item 5) could also matter.

The reversal found in the observation-regime sensitivity analysis (Section VIII-E) – smaller L_{obs} associated with a slightly *higher* rate, opposite to the button regime's response to L – is a reminder that "more information range" is not uniformly "more evacuation" once information is a recent flow rather than a persistent stock; the sign of the effect depends on which quantity a neighbor-rate threshold is actually being compared against.

The roughly $6.5\times$ gap between the button regime's individualized-parameter result (62.0%) and the observation regime's result (9.6%) in Table V illustrates, strictly within this model, the scale of difference the two information regimes produce under a common, survey-derived population. It should not be read as a validated forecast of real-world gain from deploying such a system: the model has not itself been validated against real evacuation records (Section X), so the $6.5\times$ figure is best treated as an illustrative, model-internal comparison of what the mechanism could be worth under the model's assumptions, not a deployment-ready estimate.

The temporal analysis (Section VIII-F) adds a practical nuance to this comparison. Because the early phase of the cascade is dominated by low-threshold agents and is nearly independent of L (Table VI), the marginal value of a wider sharing radius is realized late – after the first two to three hours – when it sustains the cascade into the mid- and high-threshold majority. Within the model, this suggests that the button system's benefit is greatest precisely in long-duration events, and that its effect on the crucial *early* evacuation window may be modest unless paired with measures that raise the low-threshold agents' baseline probability (which, per the ablation, has an effect larger than any single step of the L sweep).

It is worth recalling how far the model's outputs lie from observed behavior: the real recorded evacuation rate in the Toyohashi event was about 0.068% [2], far below even the observation regime's 9.6%. This gap is itself informative for how the model's outputs should be used in deployment decisions. Survey responses record stated intentions under a hypothetical

scenario, not revealed behavior under an actual, uncertain, and often ambiguous warning; real evacuation additionally depends on factors the model does not represent, such as physical and logistical barriers to leaving, risk misperception, and habituation to repeated false alarms, which are widely reported to push real uptake far below stated intention. The model is therefore best read not as a point prediction of real uptake – for which it is not validated (Section X) – but as a controlled comparison of information regimes and behavioral parameters holding the survey-derived population fixed: the evacuation rates reported in Section VIII are informative mainly as *relative* comparisons between information-sharing strategies (button vs. observation, individualized vs. flat P) evaluated under a common, internally consistent set of behavioral assumptions, and a deployment decision informed by this model should weight the *ranking* of conditions rather than their absolute levels.

Finally, a note on internal validity. Because the button and observation regimes share the same agent population, group thresholds, P , and warning schedule, and differ only in how the neighbor rate is computed and over what radius (stock rate over L_i vs. flow rate over L_{obs}), the $\sim 6.5\times$ gap between them is attributable to the information mechanism alone rather than to any difference in the modeled population – an isolation that is possible only because the two regimes are unified within one model (Section V). The main residual threats to that internal validity are the five declared-assumption parameters (Table III); the sensitivity analysis (Section VIII-E) bounds the observation regime's dependence on two of them (L_{obs} , W), and the ablation (Section VIII-C) shows that switching the CareScale mapping off entirely shifts the final rate by nearly 18 percentage points; alternative CareScale values, α , and P_{base} , however, have not been swept and remain unexplored degrees of freedom for future work.

X. LIMITATIONS

Limitations are listed explicitly here, both to bound the claims of this paper and to set an agenda for future work.

1) *No validation against real evacuation data:* All comparisons in this paper are against another simulation's published summary statistics [13], not against ground-truth evacuation records.

2) *Several parameters remain declared assumptions:* $P_{\text{base}}=0.5$, $\alpha=0.8$, the CareScale mapping $\{1.0, 0.6, 0.25\}$ in Eq. (1), and the observation regime's $L_{\text{obs}}=40$ m and $W=5$ min are all carried over from [13] or newly introduced without a survey item that could estimate them directly. Future work should pursue field trials or a purpose-built follow-up survey (e.g., asking directly about repeated-exposure effects, or about realistic direct observation range/recency) to calibrate these. Because the ablation in Section VIII-C shows that individualizing P alone, via the CareScale mapping, changes the final evacuation rate by close to 18 percentage points, that mapping in particular is a priority target for a systematic sensitivity sweep over alternative configurations in future work, rather than remaining a single researcher-chosen mapping as it does throughout this paper.

3) *Unclassified respondents:* The 24.2% of respondents excluded from Table I (Section III-A) could instead be assigned to Group I. Exclusion was chosen on the classification logic

of Section III-A itself – these respondents checked neither a numeric neighbor-rate level nor “I don't care,” so they provide no information about a neighbor-rate threshold either way – and this choice happens also to reproduce [13]'s published proportions closely (Table I), which is reported as a consistency check rather than as the reason for the choice. The correct treatment is nonetheless not fully determinable from [13] alone and should be revisited in future work, together with a sensitivity comparison of how the two treatments affect the population-level cascade dynamics of Section VIII.

4) *Single case-study area:* All experiments use one 1.06 km^2 mesh tile of Toyohashi City (1,856 residential buildings, each treated as one household). Although Eq. (2) is designed to hold each agent's *expected* neighbor count invariant to local density – converting a preferred house count N_i into a radius L_i scaled to ρ – the fixed-radius conditions of Section VIII-A and the observation regime's fixed L_{obs} are not, and non-uniform building layouts make each agent's *realized* neighbor count deviate from its target N_i regardless of the conversion. The cascade dynamics reported in Section VIII therefore cannot be cleanly attributed to the behavioral model alone as opposed to this particular tile's density and building layout. Disentangling the two will require repeating the same pipeline on additional tiles spanning a wider range of densities and geometries, which is left for future work; generalization to other densities, building layouts, or cities is therefore untested here.

5) *Unused survey covariates:* Age, gender, region, and prior earthquake/flood evacuation experience are present in the survey but not used to further individualize behavior in this paper.

6) *No physical evacuation constraints:* As in [13], only the evacuate/stay decision is modeled, not flood depth, route availability, or evacuation travel time.

7) *No modeling of imprecise information sharing:* [13]'s Patterns 1–2 (in which the shared evacuation rate is deliberately coarsened, e.g. to increase perceived urgency) are not reproduced here; the button regime always shares the exact true rate.

XI. CONCLUSION AND FUTURE WORK

This paper extended a survey-calibrated agent-based model of flood evacuation [9], [13] by individualizing, per agent, both the evacuation probability and the spatial range over which neighbor status is considered, using two survey questions that prior work had used, at most, only in isolation on an abstract grid, and a joint-resampling procedure that preserves the real correlation between an agent's threshold sensitivity, stated concern, and preferred information depth. The persistent-sharing and direct-observation information regimes were further unified into a single model, and a real-geometry, video-based visualization of individual simulation runs was added. The prior model's group distribution, re-derived independently from the same underlying survey data, and its “without system” baseline, reproduced as an emergent output of the individualized simulation rather than a calibrated target, both match the published values to within a few percentage points, supporting the consistency of the extension, though not, on its own, its

predictive accuracy against real evacuation behavior. An ablation showed that individualizing P alone, under the assumed CareScale mapping, shifts the final evacuation rate by almost 18 percentage points, and a sensitivity sweep of the observation regime's assumed parameters showed final rates within about 2 points of baseline but with a sign reversal, relative to the button regime, in how the awareness range affects the rate – both results that a single-information-regime, non-ablated model could not have surfaced. Future work should prioritize: calibrating the remaining declared-assumption parameters (Limitations item 2) against field data, including a systematic sensitivity sweep over alternative CareScale configurations given the size of the P -individualization effect; resolving the treatment of unclassified respondents (Limitations item 3) with the sensitivity comparison outlined there; extending the model to additional case-study areas to separate the behavioral model's effects from this tile's density and geometry (Limitations item 4); and incorporating the demographic and prior-experience covariates already present in the survey but not yet used (Limitations item 5).

DECLARATION ON GENERATIVE AI-ASSISTED TOOLS

Generative AI tools were used to translate the manuscript from Japanese into English and to improve the clarity and fluency of the English writing. The author critically reviewed and edited all resulting text and takes full responsibility for the scientific content, analysis, interpretation, references, and the final manuscript.

REFERENCES

- [1] MS&AD InterRisk Research & Consulting, Inc., "Survey results on evacuation awareness during natural disasters: Findings from a questionnaire survey," Research Letter 2022 No. 1, Tokyo, Japan, Feb. 2023.
- [2] Disaster Management Division of Toyohashi City, "Record of the Heavy Rain in June 2023 (Main Volume)," Toyohashi City, 2023.
- [3] A. Kajiya, K. Akaishi, T. Yokota, and H. Tsurusaki, "Comparative Analysis of Evacuation Behavior in Response to Typhoons No. 15 and 19 in 2019 and Proposal for Providing Disaster Information Reflecting Regional Characteristics," *Journal of Disaster Information*, vol. 19, no. 2, pp. 145–155, 2021.
- [4] Y. Takenouchi, K. Yamaguchi, and Y. Iwasaki, "Analysis of the Influence of Residents' Personality on Evacuation Awareness," *Proceedings of the Kansai Branch of the Japan Society of Urban Planning*, vol. 14, pp. 117–120, 2016.
- [5] S. Kubo, H. Yoshida, T. Ichimura, M. L. L. Wijerathne, and M. Hori, "Study on Influence of Prior Recognition of Flooding State on Evacuation Behavior," *International Journal of Disaster Risk Reduction*, vol. 63, p. 102437, 2021.
- [6] A. Takagi, S. Sugiura, H. Mori, and H. Iwata, "Analysis of Evacuation Behavior During the July 2018 Heavy Rain Disaster – A Case Study of Gifu Prefecture," *Journal of Natural Disaster Science*, vol. 38, no. S06, pp. 133–151, 2019.
- [7] M. Mizuno, Y. Tomita, S. Katsura, N. Koyanagi, R. Hanada, and T. Yasuda, "Analysis of Evacuation Rate During Landslides Using Disaster Information," *Journal of the Japan Sabo Society*, vol. 65, no. 3, pp. 29–34, 2012.
- [8] S. Tomura, B. Maaskant, M. Chiba, S. Masuya, and T. J. Yamada, "Basic Study on Estimating Evacuation Rate at Flood Events," *Advances in River Engineering*, vol. 28, pp. 349–354, 2022.
- [9] T. Fukuda, "A Development of Simulator Considering Behavioral Psychology of Japanese to Improve Evacuation Ratio in Flood," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 735–740, 2020.
- [10] M. Granovetter, "Threshold Models of Collective Behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.
- [11] T. Yabe, Y. Sekimoto, K. Tsubouchi, and S. Ikemoto, "Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data," *PLOS ONE*, vol. 14, no. 2, pp. 1–12, 2019.
- [12] T. Fukuda, "Development of a System for Evacuation Promotion with Buttons Insist on 'Begin to Evacuate'," *Research Report of National Institute of Technology, Kitakyushu College*, vol. 53, pp. 41–44, 2020, (in Japanese). [Online]. Available: <https://kitakyushu.repo.nii.ac.jp/records/2000205>
- [13] T. Fukuda, "Simulation-Based Analysis of Evacuation Information Sharing Systems Using Geographical Data," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 11, pp. 90–97, 2024.
- [14] T. Fukuda, "Evacuate Now Button: A Resident-Operated IoT Prototype for Real-Time Evacuation Sharing," *WSEAS Transactions on Circuits and Systems*, vol. 24, pp. 259–267, 2025.
- [15] E. Bonabeau, "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.
- [16] E. Du, X. Cai, Z. Sun, and B. Minsker, "Exploring the role of social media and individual behaviors in flood evacuation processes: An agent-based modeling approach," *Water Resources Research*, vol. 53, no. 11, pp. 9164–9180, 2017.
- [17] F. Taillandier, P. Di Maiolo, P. Taillandier, C. Jacquenod, L. Rauscher-Lauranceau, and R. Mehdizadeh, "An Agent-Based Model to Simulate Inhabitants' Behavior During a Flood Event," *International Journal of Disaster Risk Reduction*, vol. 64, p. 102503, 2021.
- [18] Z. Wang, J. Huang, H. Wang, J. Kang, and W. Cao, "Analysis of Flood Evacuation Process in Vulnerable Community with Mutual Aid Mechanism: An Agent-Based Simulation Framework," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, p. 560, 2020.
- [19] R. Zhang, D. Liu, E. Du, L. Xiong, J. Chen, and H. Chen, "An Agent-Based Model to Simulate Human Responses to Flash Flood Warnings for Improving Evacuation Performance," *Journal of Hydrology*, vol. 628, p. 130452, 2024.
- [20] M. Shirvani, G. Kesserwani, and P. Richmond, "Agent-Based Simulator of Dynamic Flood-People Interactions," *Journal of Flood Risk Management*, vol. 14, no. 2, p. e12695, 2021.
- [21] M. Shirvani, G. Kesserwani, and P. Richmond, "Agent-Based Modelling of Pedestrian Responses During Flood Emergency: Mobility Behavioural Rules and Implications for Flood Risk Analysis," *Journal of Hydroinformatics*, vol. 22, no. 5, pp. 1078–1092, 2020.
- [22] A. Jahani, S. Jess, D. Groen, D. Suleimenova, and Y. Xue, "Developing an Agent-Based Simulation Model to Forecast Flood-Induced Evacuation and Internally Displaced Persons," in *Proceedings of the International Conference on Computational Science*, pp. 550–563, 2023.
- [23] Y. Han, L. Mao, X. Chen, W. Zhai, Z. Peng, and P. Mozumder, "Agent-Based Modeling to Evaluate Human-Environment Interactions in Community Flood Risk Mitigation," *Risk Analysis*, vol. 42, no. 9, pp. 2041–2061, 2022.
- [24] L. Tierolf, T. Haer, W. J. W. Botzen, J. A. de Bruijn, M. J. Ton, L. Reimann, and J. C. Aerts, "A Coupled Agent-Based Model for France for Simulating Adaptation and Migration Decisions Under Future Coastal Flood Risk," *Scientific Reports*, vol. 13, no. 1, p. 4176, 2023.
- [25] S. Nabinejad and H. Schüttrumpf, "Agent-Based Modeling for Household Decision-Making in Adoption of Private Flood Mitigation Measures: The Upper Kan Catchment Case Study," *Water*, vol. 16, no. 14, p. 2027, 2024.
- [26] A. Taberna, T. Filatova, D. Roy, and B. Noll, "Tracing Resilience, Social Dynamics, and Behavioral Change: A Review of Agent-Based Flood Risk Models," *Socio-Environmental Systems Modelling*, vol. 2, 2020.
- [27] D. Miyagawa and G. Ichinose, "Cellular Automaton Model with Turning Behavior in Crowd Evacuation," *Physica A: Statistical Mechanics and its Applications*, vol. 549, p. 124376, 2020.
- [28] G. Bernardini, E. Quagliarini, M. D'Orazio, and M. Brocchini, "Towards the Simulation of Flood Evacuation in Urban Scenarios: Experiments to Estimate Human Motion Speed in Floodwaters," *Safety Science*, vol. 123, p. 104563, 2020.

- [29] G. P. D. P. Senanayake, M. Kieu, Y. Zou, and K. Dirks, "Agent-Based Simulation for Pedestrian Evacuation: A Systematic Literature Review," *International Journal of Disaster Risk Reduction*, vol. 111, p. 104705, 2024.
- [30] Q. Duan, "Pedestrian Evacuation Dynamics: A Literature Review," *Journal of Human Movement Science*, vol. 5, pp. 54–61, 2024.
- [31] A. Zablotzky, M. Kuperman, and S. Bouzat, "Pedestrian Evacuations with Imitation of Cooperative Behavior," *Physical Review E*, vol. 109, no. 5, p. 054304, 2024.
- [32] T. Yabe, K. Tsubouchi, T. Shimizu, Y. Sekimoto, and S. V. Ukkusuri, "Predicting Evacuation Decisions Using Representations of Individuals' Pre-Disaster Web Search Behavior," in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD)*, pp. 2707–2717, 2019.
- [33] R. B. Myerson, *Game Theory*. Cambridge, MA: Harvard University Press, 2013.
- [34] H. Leibenstein, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," *Quarterly Journal of Economics*, vol. 64, no. 2, pp. 183–207, 1950.
- [35] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), "PLATEAU," <https://www.mlit.go.jp/plateau/>, accessed 2026.