

# Hazard Function Guided Agent-Based Models: A Case Study of Return Migration from Poland to Ukraine

Technical report BII-2025-7

Zakaria Mehrab<sup>1,2</sup>, S.S. Ravi<sup>2</sup>, Logan Stundal<sup>3</sup>, Samarth Swarup<sup>2</sup>, Srinivasa Venkatramanan<sup>2</sup>, Bryan Lewis<sup>2</sup>, Henning Mortveit<sup>2,4</sup>, David Leblang<sup>3</sup>, Madhav Marathe<sup>1,2</sup>

<sup>1</sup>Department of Computer Science, University of Virginia

<sup>2</sup>Biocomplexity Institute & Initiative, University of Virginia,

<sup>3</sup>Department of Politics, University of Virginia

<sup>4</sup>Department of Systems and Information Engineering, University of Virginia

## Abstract

The Russian invasion of Ukraine in February 2022 has led to the largest forced migration crisis in Europe since World War II, with millions displaced both internally and internationally. Among the displaced, approximately 4.2 million individuals have returned, highlighting the significance of return migration as a critical phase in the migration continuum. Existing studies on return migration are limited in scope, relying on survey-based approaches that suffer from demographic bias, lack of validation against ground truth, and inability to account for uncertainty. We propose a novel computational framework for modeling the return of conflict-induced migrants, using agent-based models (ABMs) and their surrogates. These models are grounded in hazard functions and account for sociopolitical contexts. Our proposed ABMs outperform baseline methods in estimating return migration from Poland to Ukraine by at least 42% and by as much as 57% in terms of normalized root mean squared error (NRMSE). Further, to illustrate the utility of such models for policymakers, we conduct two case studies that estimate the duration of displacement and characterize the demographic breakdown among the returnees.

## 1 Introduction

The Russian invasion of Ukraine that began on February 24, 2022, has caused, among other things, the largest forced migration in Europe since the end of World War II [UNHCR, 2023]. As of August 2024, around 3.7M people have been reported to be internally displaced in various parts of Ukraine [IOM, 2024]. Around 6.8M people have been displaced as refugees, among which around 6.2M have taken shelter in various European countries [UNHCR, 2024]. According to the 2025 Humanitarian Needs and Response Plan, 12.7M people are estimated to be in need of multidimensional humanitarian assistance. One such dimension considers potential migrants who have already returned or

who are seeking to return to Ukraine once the situation escalates. The return migration to Ukraine began as early as April 2022. With the out-migration from Ukraine seemingly halted, return migration is the most prominent dimension that needs addressing at present.

Reports suggest that around 4.2M people have returned to Ukraine as of October 2024 [IOM, 2024]. Even though return migration is identified to be among the highest of priorities among the multi-dimensional top-down management issues surrounding international migration [Cassarino, 2008], there is a paucity of systematic literature surrounding return migration [Adhikari and Hansen, 2014; Şahin-Mencütek, 2024; Zetter, 2021; Toth-Bos *et al.*, 2019]. In fact, available studies attempting to tackle return migration in the context of the conflict in Ukraine have relied on survey-based approaches [van Tubergen *et al.*, 2024; Maidanik, 2024; Studien, 2024]. While potentially useful in identifying key factors driving return migration, three limitations raise concerns regarding the applicability of these survey-based approaches in the long run. *First*, these surveys are conducted with a subsample of the population, which can make these studies demographically **biased**. *Second*, models fitted from these survey responses are not validated against any ground truth observations, making their **validity** questionable. *Third*, models developed from these studies are not probabilistic, making them unable to account for the **uncertainty** associated with return migration, which is inherent in spatiotemporal human mobility [Zhou *et al.*, 2021]. Thus, a computational tool that can overcome these limitations is imperative.

**Contributions:** To address the above limitations, we assembled a **multidisciplinary team** consisting of *computer scientists* and *political scientists*. By integrating AI-driven modeling and political science insights, our work makes the following contributions to address one dimension of the pressing societal challenge of forced displacement.

- *First*, Using the concept of **hazard functions** (see e.g., [Cox and Oakes, 1984; Miller, 1981]), we propose three agent-based models (ABMs) to study return migration from a conflict-induced country by considering relevant contexts (e.g., social and political), with each ABM adding more contextual layers than the previous. We also propose sur-

80 rogates to these ABMs that generate aggregated summary  
 81 reports with lower computational expense. *To the best of*  
 82 *our knowledge*, this is the first work that computationally  
 83 models conflict-induced return migration without relying  
 84 on survey-based approaches.  
 85 • *Second*, we experimentally evaluate our models against  
 86 several baseline regression methods and show how our  
 87 models outperform the baseline methods in studying the  
 88 first wave of return migration from Poland to Ukraine,  
 89 which lasted roughly until the end of July 2022. We also  
 90 perform additional experiments to quantify the uncertainties  
 91 in the model.  
 92 • *Third*, we conduct two case studies that underscore the utility  
 93 of these models in generating policy-relevant information.  
 94 The first case study involves finding the mean length  
 95 of displacement. Validated against ground truth estimates  
 96 from subsidiary reports, this study shows the accuracy of  
 97 the emergent behavior of the ABM. The second case study,  
 98 which analyzes the demographic patterns of returnees using  
 99 the fast surrogate models, shows the efficacy of our model  
 100 in ensuring fair resource allocation based on requirements.

## 2 Related Work

101 **Return Migration:** An early work [Biondo *et al.*, 2012] that  
 102 employed a computational approach to model return migration  
 103 in the context of voluntary migration assumed that each  
 104 migrant has two associated social networks, one at home and  
 105 another abroad. Using a single-agent perspective, they cal-  
 106 culated utility at home and abroad as functions of the net-  
 107 works and decaying expectation, and when the utility abroad  
 108 becomes less than the utility at home, they assumed that the  
 109 migrant would return. Since the work was done in the con-  
 110 text of voluntary migration and used a single-agent perspec-  
 111 tive, it is not directly applicable to our case. A more re-  
 112 cent work [Alrababah *et al.*, 2023] attempted to model re-  
 113 turn intentions of Syrian refugees based on survey responses  
 114 of around 3000 refugees living in Lebanon. They identified  
 115 safety at home and network effects to be the key factors be-  
 116 hind return migration and commented that the effect of host  
 117 countries do not play a significant role during short-term mi-  
 118 gration. A recent work [van Tubergen *et al.*, 2024] studied  
 119 return migration in the context of the Ukraine war by collect-  
 120 ing survey responses from 18,000 Ukrainian refugees. They  
 121 attempted to model return migration considering three per-  
 122 spectives: (a) contextual (e.g., economic, social attachment),  
 123 (b) source country (e.g., security) and (c) cross country (e.g.,  
 124 language). Although done in the context of conflict-induced  
 125 forced migration, these two studies are based on surveys,  
 126 which are often time consuming. Furthermore, both works  
 127 have acknowledged the respondent group to be demographi-  
 128 cally biased, which may fail to paint a general picture.

129 **Hazard Function:** Although hazard functions have been  
 130 used extensively in areas such as biometrics and industrial  
 131 engineering to examine life-expectancy and product fail-  
 132 ures [Lee and Horvitz, 2017; Rizzuto *et al.*, 2017; Saikia  
 133 and Barman, 2017], researchers have also applied them in  
 134 other contexts. For example, Liu *et al.* (2017) applied hazard  
 135 function to study influence of one user over another in

136 social networks. Auld *et al.* (2011) developed a compet-  
 137 ing hazard model based on the Weibull hazard function and  
 138 used it to simulate the daily activities of individuals. Azzarri  
 139 et al. (2009) developed a duration model based on the log-  
 140 logistic hazard to understand the factors behind migration and  
 141 return of Albanian migrants, based on surveys. However, the  
 142 method was developed in the context of voluntary migration  
 143 and the temporal resolution was yearly, making it infeasible  
 144 to study short-term conflict-induced migration. Finally, haz-  
 145 ard functions have also been used in conjunction with ABMs  
 146 in various domains. For example, Wu *et al.* (2020) used ABM  
 147 simulation to infer the hazard of transmission and that of re-  
 148 recovery. Another paper [Billari *et al.*, 2007] uses an ABM and  
 149 a hazard function to study the influence of an agent’s age and  
 150 social pressure on the likelihood of the agent getting married.  
 151 These examples illustrate applicability of ABMs in conjunc-  
 152 tion with hazard functions to model various social issues.

## 3 Methodology

### 3.1 Overview

153 The concepts of hazard and survival functions lie at the core  
 154 of our modeling. While existing literature attempts to quan-  
 155 tify return intention by considering different combinations of  
 156 factors based on survey responses, we start with the basic as-  
 157 sumption that every migrant wants to return eventually. As  
 158 we point out in this section, this latent intention can be ac-  
 159 commodated in the simplest of our modeling approaches to  
 160 the very complex ones. First, we discuss these models under  
 161 the hood of the ABM framework, with the assumption that  
 162 the time of migration of each agent is known, along with other  
 163 properties. Subsequently, we discuss learnable surrogates of  
 164 the ABMs which can be leveraged to generate important re-  
 165 sults with less computational expense.

### 3.2 Hazard and Survival Functions

166 Hazard and survival functions were conceptualized and fur-  
 167 ther developed by Cox [Cox, 1959, 1972]. We briefly review  
 168 these concepts to make our paper self-contained. Additional  
 169 information can be found in several textbooks (e.g., Cox and  
 170 Oakes [1984]; Kleinbaum and Klein [2012]; Miller [1981]).  
 171

172 Hazard and survival functions are defined for a random  
 173 variable  $T$  with an underlying probability distribution func-  
 174 tion (PDF)  $p(T = t)$  and a cumulative distribution function  
 175 (CDF)  $F(T \leq t)$  that gives the probability that an event has  
 176 occurred by time  $t$ . The **survival function**  $S(t)$  is defined  
 177 as the complement of the CDF. Thus,  $S(t) = 1 - F(t) =$   
 178  $\int_t^{\infty} p(x)dx$ , which gives the probability that the event of in-  
 179 terest has not occurred by time  $t$ .

180 The **hazard function**  $h(t)$  is often defined in conjunction  
 181 with the survival function. Considering the event of inter-  
 182 est as a failure,  $h(t)$  is interpreted as the *instantaneous rate of*  
 183 *failure* given that the failure has not occurred by time  $t$ . Math-  
 184 ematically,  $h(t)$  is defined by  $h(t) = \frac{p(t)}{S(t)} = -\frac{\partial}{\partial t} \log S(t)$ . It  
 185 follows that  $S(t) = \exp(-\int_0^t h(u)du)$ .

### 3.3 Problem Formulation

186 Let  $A$  be the set of agents and  $X$  be the set of conflict-induced  
 187 regions. Each agent  $a \in A$  has the following attributes. First,

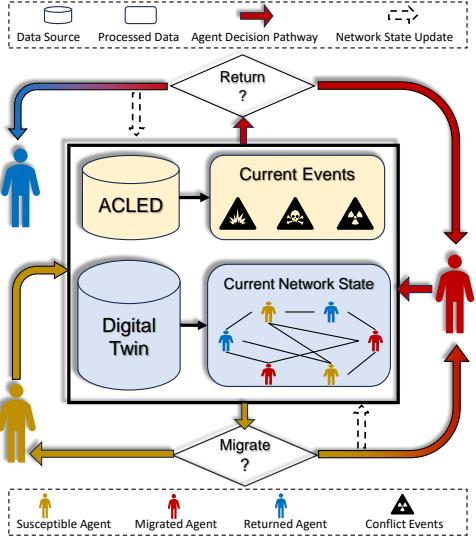


Figure 1: Architecture of a conceptual ABM to capture one cycle of migration of an agent. (From bottom left) a susceptible agent from a conflict-induced region decides to migrate based on conflict context and network influence and becomes a migrated agent. Afterward, when the conflict situation improves and their peer start to return, they may also return. Our proposed model primarily focuses on the decision pathway shown in red.

their migration time  $T_a^M$ , the time at which agent  $a$  migrates to a different country. For agents  $a$  who never migrate, we set  $T_a^M = \infty$ . Second, their origin  $x_a \in X$ . Additionally, let  $G_A(A, E_A)$  be the network formed by the agent set  $A$  where  $E_A$  corresponds to the agent pair who may exert peer influence over one another. Let  $\mathcal{N}_a$  be the set of agents who are the neighbors of agent  $a$  induced by the edges  $E_A$ . We assume that the edge set  $E_A$  does not change over time.

Let  $h_a(t > T_a^M)$  denote the hazard rate of an agent at time  $t$  that represents instantaneous rate of the agent failing to stay migrated (i.e., returning to  $x_a$ ). While recurring movements are possible, we assume that agents only leave their countries of migration when their goal is return migration. Applying the formula for  $S(t)$  in the discrete-time scenario, the probability  $S_a(t+1)$  that the agent will stay migrated at time  $t+1$  given that it has remained a migrant up to time  $t$ , can be written as  $S_a(t+1) = \exp(-\sum_{t'=0}^t h_a(t'))$ . By doing a Bernoulli sampling based on  $S_a(t)$  (where  $\text{Bern}(p)$  indicates that the Bernoulli trial leads to success with probability  $p$ ), we can estimate return state of an agent at time  $t$ . Formally, let  $r_a(t)$  denote the state whether an agent has returned by time  $t$  or not. Then,  $r_a(t) \sim \text{Bern}(S_a(t))$  tells us the return state of agent  $a$  at time  $t$  (if the agent has not already returned). Thus, modeling the hazard function  $h_a(t)$  will allow us to control the return dynamics in various ways. Subsequent sections describe how we employ various strategies to progressively refine the hazard function by employing various societal factors that play a key role in conflict-induced migration scenarios. Across all the models  $h_a(t \leq T_a^M) = 0$ , which constrains agents from returning before migration. Thus, from here on,  $h_a(t)$  corresponds to  $h_a(t > T_a^M)$ .

### 3.4 ABM of Return Migration

223

#### Model 1: Basic Hazard (BASE)

224

**Motivation:** In the context of forced migration, it is well established that the general intention of displaced individuals is to eventually return to their place of origin [Zetter, 2021]. Therefore, the decision to return can be effectively seen as a time-to-event problem, where the “event” is the act of returning. The hazard function is often used for survival analysis in discrete-time simulation [Suresh *et al.*, 2022]. Thus, it is a natural choice for modeling return migration dynamics.

**Formulation:** Various forms of hazard functions are used in the literature. A comprehensive list of these various forms, along with their mathematical properties and corresponding survival functions has been described in [Van Wijk and Simonsson, 2022]. Our initial model uses the simplest of these forms where we consider the hazard rate to be constant for all agents at all times. Thus, the ABM is characterized by:

$$h_a(t) = h \quad (1)$$

It indicates that at any time  $t > T_a^M$ , agent  $a$  has a constant value  $h$  of baseline hazard rate that controls their likelihood of returning if they have not already returned by that time. It can be shown that, if the hazard rate is constant, the underlying probability distribution for the time to return can be expressed as a geometric distribution [Chakraborty and Gupta, 2015]. Thus,  $P(r_a(t) = 1) = h \times (1 - h)^{t-T_a^M}, \forall a, \forall t$ .

**Agent Dynamics:** Thus, the *Basic Hazard Guided ABM* works as follows. Initializing the baseline hazard rate  $h$ , at every time step  $t$ , all the agents  $a$  with  $T_a^M < t$  and  $r_a(t-1) = 0$  calculate  $S_a(t)$  (probability of not returning) based on the baseline hazard  $h$ . Then,  $r_a(t)$  is sampled from  $S_a(t)$ .

A more nuanced analysis can be conducted by classifying the agents into groups and assigning each group a separate hazard rate. While we explore this in a case study, our goal is to provide the foundation for adding more layers of complexity starting from the simplest of assumptions. Therefore, while acknowledging the possibility of agent class-based hazard rates, we refrain from doing so in our initial evaluations.

#### Model 2: Conflict Context Influenced Hazard ( $C^L$ )

259

**Motivation:** Having established the basic ABM for modeling return migration, we next consider what other contexts to incorporate in calculating the hazard rate and how to incorporate them. Several survey reports conducted on Ukrainian migrants to identify key factors behind their intention to return can help us in this regard. One report published by IOM [Sohst *et al.*, 2024] indicates that *improved security situation in origin community/Ukraine* is the primary motive behind Ukrainian refugees’ return in the short term. In another, safety was marked as the primary factor for return by more than 45% of the participants [Sologoub, 2024].

**Formulation:** We incorporate conflict context for modeling return migration dynamics into the ABM as follows. Let,  $Z_a \subseteq X$  denote the regions observed by agent  $a$ . Let  $c(z, t)$  denote the conflict context<sup>1</sup> of region  $z$  at time  $t$ . We use this conflict context to modify the baseline hazard rate as follows.

<sup>1</sup>Conflict context is a scalar quantity to measure the intensity of conflict in a region at a particular time. More details are in Section 4.

$$h_a(t) = h \times \left( 1 - \frac{1}{K} \sum_{z \in Z_a} \sum_{t'=0}^W c(z, t-t' - \mathcal{L}) \right) \quad (2)$$

276 Here,  $\mathcal{L}$  is a lag parameter that controls the time of the latest  
 277 conflict context. Since migration decisions involve planning  
 278 and travel time [Wycoff *et al.*, 2024], choosing such param-  
 279 eter is important for temporal mapping of a migrant from in-  
 280 tending to return to being recorded as a returnee at the border.  
 281 The window parameter  $W$  is used to reduce the effect of noise  
 282 present in the data recording conflict context [Mehrab *et al.*,  
 283 2022]. Finally,  $K$  is a normalizing constant so that the right-  
 284 side term of the multiplication is bounded between 0 and 1. In  
 285 the evaluation section, we describe how these are configured.  
 286 From Equation (2), we can see that when the conflict context  
 287 is high, the baseline hazard will be scaled down, indicating  
 288 that the likelihood of return will decrease and vice-versa.

289 **Agent Dynamics:** Thus, the *Conflict Context Influenced Haz-  
 290 ard Guided ABM* works as follows. At every discrete-time  
 291 step  $t$ , all the agents  $a$  with  $T_a^m < t$  and  $r_a(t-1) = 0$   
 292 look at the conflict contexts of each region  $z \in Z_a$  from time  
 293  $t - \mathcal{L} - W$  to  $t - \mathcal{L}$  and incorporate it to scale the baseline  
 294 hazard  $h$  following Equation (2). This scaled hazard is used  
 295 to calculate the return probability of each agent.

296 Researchers have used other forms of hazard functions  
 297 (e.g., Weibull) to include additional covariates [Auld *et al.*,  
 298 2011]. However, these covariates are added in the exponent  
 299 term while the baseline hazard rate is kept as separate. Since  
 300 conflict context is significant to return dynamics, we choose  
 301 this functional form to let it affect the baseline hazard. More-  
 302 over, the other forms of hazard functions have additional pa-  
 303 rameters. Since this study aims at developing the simplest  
 304 possible model with a minimal set of parameters, we refrain  
 305 from using functional forms that introduce additional pa-  
 306 rameters and complexities.

### 307 Model 3: Conflict and Peer influenced Hazard ( $C^{\mathcal{L}}P$ )

308 **Motivation:** According to many social and behavioral theo-  
 309 ries (e.g., theory of planned behavior [Ajzen, 1991], herd be-  
 310 havior [Banerjee, 1992]), peer influence is a key factor behind  
 311 driving one’s decisions. It has also been established that both  
 312 conflict events and peer influence are important in driving mi-  
 313 gration decision-making [Mehrab *et al.*, 2024a]. Following  
 314 this, in our final model, we incorporate peer influence along  
 315 with conflict context. We do so using the threshold model, a  
 316 well-known model used for capturing peer influence [Valente,  
 317 1996; Granovetter, 1978]. Traditionally threshold model is  
 318 used in the context of peer influence by having agents go  
 319 through state transition when the number of neighbors with a  
 320 particular state exceeds some threshold [Hancock *et al.*, 2022;  
 321 Qiu and others, 2022]. Since we are modeling the return state  
 322 with hazard function, we use the threshold model instead to  
 323 have the agents transition using different hazard functions.

324 **Formulation:** Let,  $q_a(t) = \sum_{a' \in \mathcal{N}_a} r_{a'}(t)$  be the number of  
 325 neighbors of  $a$  who have returned by time  $t$ . We express  $h_a(t)$   
 326 in the form of a piecewise hazard function as follows.

$$h_a(t) = \begin{cases} h_\ell \cdot \left( 1 - \sum_{z \in Z_a} \sum_{t'=0}^W \frac{c(z, t-t'-\mathcal{L})}{K} \right) & \text{if } \frac{q_a(t)}{|\mathcal{N}_a|} \geq \tau_a \\ h_s \cdot \left( 1 - \sum_{z \in Z_a} \sum_{t'=0}^W \frac{c(z, t-t'-\mathcal{L})}{K} \right) & \text{otherwise.} \end{cases} \quad (3)$$

327 Here,  $\tau_a$  ( $0 \leq \tau_a \leq 1$ ) is the fractional active threshold  
 328 parameter required to influence agent  $a$ . Since considering  
 329 different thresholds across agents is difficult, we consider a  
 330 simpler model where each agent has the same fractional ac-  
 331 tive threshold parameter (i.e.,  $\forall a \in A, \tau_a = \tau$ ). While we  
 332 do not impose any constraint here, ideally one can expect  $h_\ell$   
 333 to be larger than  $h_s$ , since when more people are returning it  
 334 should drive a neighboring agent more likely to return.

335 **Agent Dynamics:** Thus, the *Conflict Context and Peer Influ-  
 336 enced Hazard Guided ABM* works as follows. At every time  
 337 step  $t$ , all the agents  $a$  with  $T_a^m < t$  and  $r_a(t-1) = 0$  look  
 338 at the fraction of their neighbors who have migrated and re-  
 339 turned. If the fraction is larger (smaller) than  $\tau$ , they choose  
 340 a baseline hazard of  $h_\ell$  ( $h_s$ ). Then, they look at the conflict  
 341 contexts of each region  $z \in Z_a$  from time  $t - \mathcal{L} - W$  to  
 342  $t - \mathcal{L}$  and incorporate it to scale the baseline hazard and sub-  
 343 sequently calculate the return probability using it.

344 It can be seen that we have progressively added layers to  
 345 our model to incorporate contexts of different dimensions,  
 346 making the models progressively more powerful in capturing  
 347 return migration dynamics. However, as this has two addi-  
 348 tional parameters (a different baseline hazard and the thresh-  
 349 old), it will be computationally more expensive to calibrate.

## 350 3.5 Surrogate Models of Return Migration

351 In this section, we propose some surrogates to the ABMs op-  
 352 erating in the aggregated population space rather than the in-  
 353 dividual agent space. While we lose some granularity by do-  
 354 ing so, we significantly reduce the computational cost.

355 The problem formulation for return migration in the  
 356 aggregated population space is as follows. Let  $M =$   
 357  $\langle m(1), m(2), \dots, m(T) \rangle$  be the temporal estimates of mi-  
 358 grants, where  $m(t)$  is the number of migrants at time  $t$ . We  
 359 define  $R$  to be the  $T \times T$  return matrix, where  $r_{i,j}$  is defined as  
 360 the number of returnees at time  $i$  among those who migrated  
 361 at time  $j$ . Thus, by definition, when  $i \leq j$ ,  $r_{i,j} = 0$ . Our goal  
 362 is to find the values  $r_{i,j}$ ,  $\forall i, j : i > j$ . We will refer to these  
 363 as the **non-trivial** entries of the  $R$  matrix. In the remainder of  
 364 this subsection, we describe three surrogates corresponding  
 365 to the three proposed ABMs; each description shows how the  
 366 non-trivial values of  $R$  are computed.

### 367 Surrogate to the BASE model

368 Let  $h$  be the constant hazard rate defined as the parameter  
 369 of BASE. Let  $Q_{i,j}$  denote the number of remaining migrants  
 370 at the time  $i$  from time  $j$  who survive (i.e., do not return),  
 371 calculated by  $Q_{i,j} = m(j) - \sum_{k < i} r_{k,j}$ . Thus, the surrogate  
 372 to BASE computes the non-trivial entries  $r_{i,j}$  as follows.

- 373 •  $r_{i,j} = Q_{i,j} (1 - e^{-h(i-j)})$  if  $i > j$

374 **Surrogate to the  $C^L$  Model**

375 Similar to how the problem space in the aggregated population space, the surrogate also assumes conflict context at  
 376 a global level. Let  $h_c^t$  be the conflict-influenced hazard ex-  
 377 pressed as  $h_c^t = h \times (1 - \frac{1}{K} \sum_{x \in X} \sum_{t'=0}^W c(x, t - t' - \mathcal{L}))$ ,  
 378 where  $W, K, \mathcal{L}$  have similar meanings as described for  $C^L$ .  
 379 Thus, the non-trivial entries  $r_{i,j}$  can be calculated as follows.  
 380

381 •  $r_{i,j} = Q_{i,j} (1 - e^{-h_c^t(i-j)})$  if  $i > j$

382 **Surrogate to the  $C^L$ P model**

383 Since  $C^L$ P considers peer influence at the local neighborhood  
 384 level of agents, an exact formulation of this peer influence is  
 385 difficult for the surrogate working at aggregated population  
 386 space. We approximate the peer influence as follows. Let,  
 387  $\tilde{r}(t) = \sum_{t'=1}^t \sum_{j=1}^{t'} r_{i,j}$  be the total number of people who  
 388 have returned by time  $t$ . Let  $h_{cp}^t$  be the conflict and peer  
 389 influenced hazard expressed as follows.

$$h_{cp}^t = \begin{cases} h_\ell \times \left(1 - \sum_{x \in X} \sum_{t'=0}^W \frac{c(x, t - t' - \mathcal{L})}{K}\right) & \text{if } \frac{\tilde{r}(t)}{N} \geq \tilde{\tau} \\ h_s \times \left(1 - \sum_{x \in X} \sum_{t'=0}^W \frac{c(x, t - t' - \mathcal{L})}{K}\right) & \text{otherwise} \end{cases}$$

390 where  $N$  is the total number of agents. This essentially is a  
 391 peer influence considered at the overall population level in-  
 392 stead of the local neighborhood level. Using this, this surro-  
 393 gate calculates non-trivial  $r_{i,j}$  as follows.

394 •  $r_{i,j} = Q_{i,j} (1 - e^{-h_{cp}^t(i-j)})$  if  $i > j$

395 Here, the exponent term contains the conflict and peer-  
 396 influenced hazard term  $h_{cp}^t$  which is defined above. It can be  
 397 seen that while these surrogate models can generate aggregated  
 398 level outputs, they cannot generate detailed individual  
 399 agent level behavior. However, since the outputs of surrogate  
 400 models can be generated with great computational efficiency,  
 401 if they are similar to the results of ABM, they can be used as  
 402 components of beneficial software tools for policymakers.

403 **3.6 Model Calibration**

404 In order to find a parameter configuration (e.g., hazard rate,  
 405 threshold parameter) that best fits the model, we define a *loss*  
 406 function that combines the capability of the model to cap-  
 407 ture the scale and trend of the observed return estimate and  
 408 apply a black-box optimization technique, namely *Bayesian*  
 409 *optimization*, to optimize this loss function. Bayesian opti-  
 410 mization is applied when the parameter space has a small  
 411 dimension (typically  $< 20$ ), the objective function is com-  
 412 putationally extensive and its gradient is non-trivial to evaluate  
 413 (Frazier [2018]).

414 Let  $R_\theta = \langle r_\theta(1), r_\theta(2), \dots, r_\theta(T) \rangle$  be daily return esti-  
 415 mates provided by our models (ABM or surrogate) parame-  
 416 terized by  $\theta$ . Let,  $\tilde{R} = \langle \tilde{r}(1), \tilde{r}(2), \dots, \tilde{r}(T) \rangle$  be the observed  
 417 number of returnees which we want to calibrate our model.  
 418 We define the following loss function.

$$L(\theta) = \lambda_e \text{RMSE}(R_\theta, \tilde{R}) + \lambda_c (1 - \text{PCC}(R_\theta, \tilde{R})) \quad (4)$$

The first term calculates the root mean squared error (RMSE) 419  
 between the model estimate and the ground truth estimate 420  
 whereas the second term calculates the Pearson correlation 421  
 coefficient (PCC) between the model estimate and the ob- 422  
 served estimate. Here,  $\lambda_e$  and  $\lambda_c$  are weight coefficient hy- 423  
 perparameters. Our proposed loss can help the model avoid 424  
 getting stuck in local minima in terms of just the traditional 425  
 RMSE loss. In the Appendix, we present a justification be- 426  
 hind this claim and an ablation study on the loss function to 427  
 experimentally validate our claim. 428

Baseline	NACRPS		NRMSE		PCC	
$L^{21}$	N/A		0.223		0.569	
$L^{30}$	N/A		0.213		0.617	
$L^{40}$	N/A		0.211		0.628	
$L^{45}$	N/A		0.207		0.647	

Ours	NACRPS		NRMSE		PCC	
	ABM	S-ABM	ABM	S-ABM	ABM	S-ABM
BASE	0.067	0.062	0.12	0.131	0.95	0.92
$C^{21}$	0.059	0.060	0.112	0.135	0.93	0.91
$C^{21}P$	0.057	0.062	0.106	0.126	0.93	0.91
$C^{30}P$	0.051	0.083	0.089	0.102	0.95	0.93
$C^{35}P$	0.052	0.051	0.094	0.128	0.96	0.94
$C^{40}P$	0.056	0.054	0.107	0.169	0.95	0.89

Table 1: Evaluation of return estimation of proposed ABMs and their corresponding surrogates (denoted as S-ABM). The metrics are reported by comparing the median estimate of the ABM (or surrogate) with the ground truth. Each instance of ABM takes around 10 minutes and the surrogate takes around 3 seconds to run.

4 **Evaluation**

The following subsection provides an overview of the config- 430  
 431urations of our experiments and the subsequent subsections  
 432 describe the results and case studies. All evaluation and stud-  
 433ies involve the case of the return migrants from Poland to  
 434 Ukraine during the early period of the Russian invasion of  
 435 Ukraine. While we cannot comprehensively evaluate the gen-  
 436 eralizability of the model due to inadequacy of ground truth  
 437 pertaining to diverse scenario, we emphasize that our model  
 438 can be applied in other conflict scenarios as well.

4.1 **Experimental Setup**

**Datasets:** We parse the conflict events of Ukraine during the 440  
 441 Feb-Aug 2022 period from ACLED [ACLED, 2010]. For the  
 442 agent data, we use the synthetic household data described  
 443 in [Mortveit *et al.*, 2020]. We collect the synthetic house-  
 444 holds of Ukraine and the associated synthetic individuals.  
 445 These households represent the agents in our model. Fi-  
 446 nally, we obtain the border-crossing data from Polish border  
 447 guards [Portal, 2022] and use the number of Ukrainians cross-  
 448 ing a Poland-Ukraine border between February 24, 2022 and  
 449 August 01, 2022 as the ground truth.

**Building Block Models:** Since our problem formulation 450  
 requires knowing the migration time of the agents to esti- 451  
 452

mate their return time (Figure 1), we first apply the ABM-TPB model [Mehrab *et al.*, 2024b] to assign each migrant a refugee or internally displaced (IDP) status. Then, we apply the method proposed in [Pandey *et al.*, 2023] to place each refugee in one of the six neighboring countries of Ukraine and apply our methodology on the refugees placed in Poland. To account for the uncertainties associated with these models, we run 100 realizations of ABM-TPB model to create a Digital Library (DL) of simulations (see appendix for details) by using Latin Hypercube Sampling (LHS) over the parameter spaces for ABM-TPB. From the DL, we select the 30 best performing simulations. Then, we run our models for return migration over each of these simulations 30 times.

**Hyperparameters:** We create the agent network  $G_A$  using Kleinberg’s Small World Model (KSW) [Kleinberg, 2000] using the methods and parameters described in [Mehrab *et al.*, 2024a]. The same network is also used for the ABM-TPB model. Across all the  $C^L$  and  $C^L P$  models, we use a constant window parameter  $W = 7$  days and the lag parameter  $L \in \{21, 30, 35, 40\}$  days. To reduce the number of possible covariate combinations, we also set  $Z_a = X, \forall a \in A$  for our ABM models. Future research can explore how limiting the scope of observation to different radii of neighboring regions can change the outcome of the model and how this can be done across agents of various demographic groups. Finally, we use daily total fatalities from ACLED as the conflict context to our ABMs and their surrogates and  $K$  is chosen to be the maximum value among all conflict contexts.

**Metrics:** Our first evaluated metric is Normalized Root Mean Squared Error (NRMSE), used to calculate the average error in the model estimation. The second metric is Normalized Average Continuous Ranked Probability Score (NACRPS), used to quantify how well the model can account for uncertainty in its estimation. Finally, we also evaluate the Pearson Correlation Coefficient (PCC) to quantify how well the model can capture the trend in the observed data. These metrics are described in detail in the Appendix.

**Baseline:** As baselines for our models, we create some vanilla linear regression models  $L^D$ ,  $D \in \{20, 30, 40, 45\}$ . Here,  $D$  represents the number of days of historical conflict contexts each model uses as features. For these models, we consider both the number of events and the fatalities of events as conflict context. Note that these models are not probabilistic in nature; so, they cannot account for uncertainties.

## 4.2 Experimental Validation

Table 1 summarizes the performance of our proposed models against the baseline methods. We describe our main findings in the next two subsections. Results from a more comprehensive set of experiments can be found in the Appendix.

**Error and Correlation:** The best ABM ( $C^{30}P$ ) has a median NRMSE of 0.089, an improvement of around 57% compared to the best-performing regressor ( $L^{45}$ , with an NRMSE of 0.207). Even the simplest ABM (BASE) has a 42% improvement in NRMSE compared to  $L^{45}$ . The surrogates also outperform the baseline methods. The best-performing surrogate (surrogate of  $C^{30}P$ ) has a median NRMSE of 0.102, a 51% improvement over  $L^{45}$ . Further, our proposed models capture the trend of the return migration better than baseline methods,

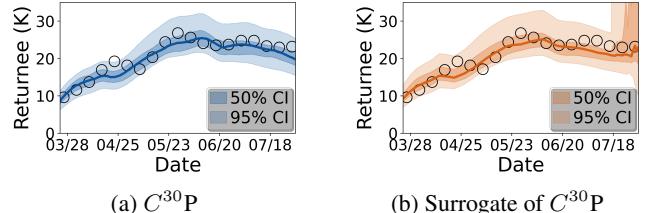


Figure 2: Best Performing ABM and Surrogates for estimating return migrants from Poland to Ukraine. Circles represent ground truth data. Visualizations corresponding to other models of Table 1 can be found in the Appendix. 50% (95%) CI corresponds to the 50% (95%) confidence interval of the model estimate.

with PCC constantly being higher than 0.9. These results underscore the capability of the proposed hazard-based models to capture the scale and trend of the return migration accurately. Notably, the  $C^L P$  models always seem to perform better than BASE or  $C^L$ , thus bringing out the importance of considering different layers of sociopolitical contexts in capturing return dynamics. We elaborate on this through an ablation study of the model components in the Appendix.

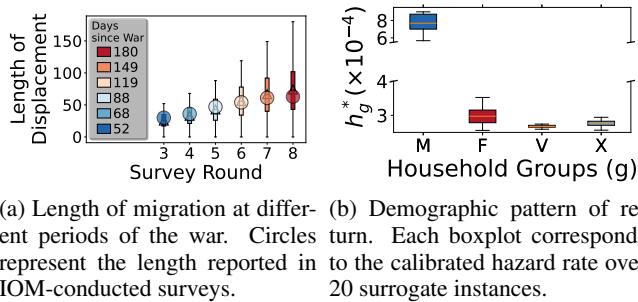
**Surrogate vs ABM:** The proposed surrogate models are able to produce estimates comparable to their ABM counterparts. We take the  $C^{30}P$  as an example, which has the best performance in terms of NRMSE. Its surrogate has an NRMSE of 0.102, which is worse only by 14.6% compared to the NRMSE of its ABM counterpart. However, the surrogate model takes around three seconds to generate the return estimate whereas the ABM takes close to 10 minutes for one instance to execute. Such a significant improvement in computation time may compensate for the penalty in accuracy. However, it must be noted that the more sub-optimal the ABM gets, the performance of its surrogate gets progressively worse. For example, the performance drop in NRSME of the surrogate for  $C^{35}P$  is almost 36.1% (0.128 vs 0.094) and for  $C^{40}P$  is around 57%. This behavior also underscores the importance of choosing a good lag parameter. We study the effect of the lag parameter in the Appendix.

## 4.3 Uncertainty Quantification

It is important to quantify the uncertainty associated with a probabilistic model designed for generating return migration estimates. To do so, we analyze the NACRPS of the models tabulated in Table 1. *First*, we notice that adding layers of complexity to BASE decreases NACRPS, suggesting improvement of both point forecasts and probabilistic forecasts. *Second*, the  $C^L P$  ABMs with lower NRMSE have low NACRPS as well; thus, these models not only have reasonable point prediction accuracy, but the uncertainty associated with probabilistic return estimation is also accounted for more effectively. *Third*, adding layers of complexity to the surrogate models do not necessarily improve NACRPS. As example, when conflict context is introduced with the BASE, the NACRPS slightly improves with a decline in NRMSE. But, when peer influence is added, the NACRPS declines as NRMSE improves. Thus, there is a tradeoff between accuracy and uncertainty across these surrogate models. If we compare

the  $C^{30}P$  surrogate with its ABM counterpart in Figure 2, we realize that the poor NACRPS of the surrogate is due to high uncertainty interval (shaded region) at the later period (Figure 2b), which does not happen for the ABM (Figure 2a). In the ABM, since peer influence comes from the local neighborhood of each agent, the threshold is unlikely to be met by all agents at once. For the surrogates, peer influence is calculated by looking at the entire population. Thus, the transition occurs for everybody simultaneously. Across multiple simulations, uncertainty propagated by this sharp transition is observed intensely, leading to higher NACRPS values.

## 5 Case Studies



(a) Length of migration at different periods of the war. Circles represent the length reported in IOM-conducted surveys. (b) Demographic pattern of return. Each boxplot corresponds to the calibrated hazard rate over 20 surrogate instances.

Figure 3: Two case Studies conducted using our proposed models

### 5.1 Return time estimation

For our first case study, we aim to understand an aspect of utmost importance to policymakers. The question we want to address is the following. *What is the expected length of displacement before someone conducts return migration?* An answer to this question helps policymakers to decide whether the migrants require long-term integration programs or short-term emergency responses [World Bank, 2025]. It can also help policymakers plan for advocating a ceasefire on various frontlines, given migrants displaced from vicinities return after a certain period [Associated Press, 2024].

To answer this question, we calculate the length of stay across agents at different periods of the Ukraine war using the output from the  $C^{30}P$  ABM. Figure 3a shows distribution of the length of displacement, along with the reported mean length of displacement provided in the IOM general survey reports [IOM, 2022]. We can see that the reported mean length agrees well with the model estimates. Moreover, the increase in the length of displacement before return with the progression of the war is correctly estimated by our model. This signifies that our proposed hazard-guided ABM is appropriate for time-to-return modeling as this emergent behavior matches qualitative observation and captures overall return migration dynamics.

### 5.2 Demographic Pattern of Returnees

Our second case study aims to estimate the demographic composition of returning migrants. Since ground truth containing demographic details is not available regarding re-

Group	Household Description	Migrant ratio (%)
M	Adult males only	1.47±0.41
F	Adult females only	6.07±0.23
V	Elderly (and/or) children	71.64±0.34
X	Adult males and females	20.81±0.16

Table 2: Group categorization based on demographic attributes. These groups were chosen upon careful discussion with the Political scientists in our team. Right column shows the distribution of each group as a migrant from ABM-TPB simulation.

turn migrants, such estimates can help policymakers in tailoring targeted reintegration initiatives [Battistella, 2018] or allocating social/medical services based on the psychological needs of specific demographic groups [Migration Observatory, 2011]. Here, the question we want to address is: *What is the likelihood of return for migrants from different demographic groups?* Fitting hazard rate parameters across various demographic groups can answer this question.

Therefore, we modify the surrogate of BASE as follows. We first categorize each household into one of the four groups outlined in Table 2. Then, we initialize  $h_g, \forall g \in \{M, F, V, X\}$  and the surrogate model produces return estimates for the groups individually. The aggregation of return estimates of all groups is calibrated to match the total return estimates. Once calibrated, the hazard rates of individual groups in Figure 3b reveal interesting emergent properties.

First, we notice that the adult male-only households have a high hazard rate compared to other groups. This indicates their high propensity to return to Ukraine, which can be attributed to the mandate requiring adult males to participate in the war. Second, we observe that households with vulnerable groups (i.e., elderly, children) are the least likely to return. This implies that the threat of security is greatest among families with children and elderly. Third, we observe that female-only households have hazard rates similar to the vulnerable groups compared to the male-only households, again indicating their greater concern for security compared to adult males. Consequently, females without partners are slightly more likely to return than females with male partners, which corroborates with prior findings [Sologoub, 2024], indicating the possibility of having a partner alongside to be more motivating to settle in a new place. Finally, there is less variance in hazard rate across the V group, compared to the other groups. Since the largest portion of migrants is comprised of this group, a slight variation from the calibrated hazard rate may cause a large deviation from the ground truth. Overall, the ability to produce demographically detailed estimates underscores the usefulness of the model in building decision-support tools for policymakers.

## 6 Concluding Remarks

We developed novel agent-based models (ABMs) for return migration, experimentally evaluated these models and conducted case studies using these models to demonstrate their usefulness in generating policy-relevant information. Our models can be seamlessly incorporated to work with prior computational models that generate outflow of Ukrainian mi-

639 grants and subsequently place them in one of the neighboring  
640 countries. Therefore, our work provides the path to develop  
641 the first end-to-end ABM of forced migration that is jointly  
642 able to capture these heterogeneous migration dynamics. Future  
643 work can explore the return dynamics of internally displaced  
644 people and consequently, address the dynamics of recurring  
645 migration.

## 646 Ethical Impact

647 The work reported in this paper proposes agent-based models  
648 for studying several aspects of returning migrants during conflicts.  
649 Such models are useful in developing software tools that will be  
650 beneficial to policymakers. The paper presents simulation-based  
651 evaluations of the models using a synthetic dataset discussed in  
652 one of the cited references and other public domain datasets.

## 654 Code Availability

655 All scripts are available in [github.com/dmehrab06/ukr\\_migration\\_return](https://github.com/dmehrab06/ukr_migration_return)

## 657 References

- 658 ACLED. Introducing ACLED-Armed Conflict Location and  
659 Event Data. *Journal of Peace Research*, 47(5):651–660,  
660 2010.
- 661 Prakash Adhikari and Wendy L Hansen. Reversing the flood  
662 of forced displacement: Shedding light on important de-  
663 terminants of return migration. Report by the Nepal Study  
664 Center, 2014.
- 665 Icek Ajzen. The theory of planned behavior. *Organizational  
666 behavior and human decision processes*, 50(2):179–211,  
667 1991.
- 668 Nawras Al Husein and Natascha Wagner. Determinants of  
669 intended return migration among refugees: A comparison  
670 of syrian refugees in germany and turkey. *International  
671 Migration Review*, 57(4):1771–1805, 2023.
- 672 Ala Alrababah, Daniel Masterson, Marine Casalis, Dominik  
673 Hangartner, and Jeremy Weinstein. The dynamics of  
674 refugee return: Syrian refugees and their migration intentions.  
675 *British Journal of Political Science*, 53(4):1108–1131, 2023.
- 677 Associated Press. Israel-hezbollah ceasefire: Refugees return  
678 home to lebanon, 2024. Accessed: 2025-02-08.
- 679 Joshua Auld, Taha Hossein Rashidi, Mahmoud Javanmardi,  
680 and Abolfazl Mohammadian. Dynamic activity generation  
681 model using competing hazard formulation. *Transportation  
682 research record*, 2254(1):28–35, 2011.
- 683 Carlo Azzarri and Calogero Carletto. Modelling migration  
684 dynamics in albania: a hazard function approach. *South-  
685 east European and Black Sea Studies*, 9(4):407–433, 2009.
- 686 Abhijit V Banerjee. A simple model of herd behavior. *The  
687 quarterly journal of economics*, 107(3):797–817, 1992.
- 688 Graziano Battistella. Return migration: A conceptual and  
689 policy framework, 2018. Accessed: 2025-02-08.
- Francesco C Billari, Alexia Prskawetz, Belinda Aparicio Diaz, and Thomas Fent. The “wedding-ring” an agent-based marriage model based on social interaction. *Demographic research*, 17:59–82, 2007.
- Alessio Emanuele Biondo, Alessandro Pluchino, and Andrea Rapisarda. Return migration after brain drain: A simulation approach. *arXiv preprint arXiv:1206.4280*, 2012.
- Jean-Pierre Cassarino. Conditions of modern return migrants—Editorial introduction. *International Journal on Multicultural Societies*, 10(2):95–105, 2008.
- Subrata Chakraborty and Rameshwar D Gupta. Exponentiated geometric distribution: another generalization of geometric distribution. *Communications in Statistics-Theory and Methods*, 44(6):1143–1157, 2015.
- D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall, New York, NY, 1984.
- D. R. Cox. The analysis of exponentially distributed lifetime with two types of failures. *Journal of Royal Statistical Society*, 21B:411–421, 1959.
- D. R. Cox. Regression models and life-tables. *Journal of Royal Statistical Society*, 26B:186–220, 1972.
- Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.
- M. Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420–1443, 1978.
- Matthew Hancock, Nafisa Halim, Chris J Kuhlman, Achla Marathe, Pallab Mozumder, SS Ravi, and Anil Vullikanti. Effect of peer influence and looting concerns on evacuation behavior during natural disasters. In *Proc. of COMPLEX NETWORKS*, 2022.
- IOM. Ukraine internal displacement report: General population survey round 3, April 2022.
- IOM. IOM:Displacement Tracking Matrix. <https://dtm.iom.int/ukraine>, 2024.
- David G. Kleinbaum and Mitchel Klein. *Survival Analysis – A Self-Learning Text*. Springer, Heidelberg, Germany, 2012.
- Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170, 2000.
- Dae Lee and Eric Horvitz. Predicting mortality of intensive care patients via learning about hazard. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 4953–4954, 2017.
- Shenghua Liu, Houdong Zheng, Huawei Shen, Xueqi Cheng, and Xiangwen Liao. Learning concise representations of users’ influences through online behaviors. In *IJCAI*, pages 2351–2357, 2017.
- Iryna Maidanik. The forced migration from Ukraine after the full scale Russian invasion: Dynamics and decision making drivers. *European Societies*, 26(2):469–480, 2024.

- 742 Zakaria Mehrab, Aniruddha Adiga, Madhav V Marathe,  
743 Srinivasan Venkatramanan, and Samarth Swarup. Evaluating  
744 the utility of high-resolution proximity metrics in  
745 predicting the spread of covid-19. *ACM Transactions on*  
746 *Spatial Algorithms and Systems*, 8(4):1–51, 2022.
- 747 Zakaria Mehrab, Logan Stundal, Samarth Swarup, Srinivasan  
748 Venakramanan, Bryan Lewis, Henning Mortveit, Christo-  
749 pher Barrett, Abhishek Pandey, Chad Wells, Alison Gal-  
750 vani, et al. Network agency: An agent-based model of  
751 forced migration from Ukraine. In *Proceedings of the 23rd*  
752 *International Conference on Autonomous Agents and Mul-*  
753 *tiagent Systems*, pages 1372–1380, 2024.
- 754 Zakaria Mehrab, Logan Stundal, Srinivasan Venkatramanan,  
755 Samarth Swarup, Bryan Lewis, Henning S Mortveit,  
756 Christopher L Barrett, Abhishek Pandey, Chad R Wells,  
757 Alison P Galvani, et al. An agent-based framework to study  
758 forced migration: A case study of ukraine. *PNAS nexus*,  
759 3(3):pgae080, 2024.
- 760 Migration Observatory. Demographic objectives in migration  
761 policy-making, 2011. Accessed: 2025-02-08.
- 762 Rupert G. Miller. *Survival Analysis*. John Wiley and Sons,  
763 New York, NY, 1981.
- 764 Henning. S Mortveit, Abhijin Adiga, et al. Synthetic Pop-  
765ulations and Interaction Networks for the U.S. Technical  
766 report, University of Virginia, 2020. NSSAC Technical Re-  
767 port: #2019-025.
- 768 Abhishek Pandey, Chad R Wells, Valentyn Stadnytskyi,  
769 Seyed M Moghadas, Madhav V Marathe, Pratha Sah,  
770 William Crystal, Lauren Ancel Meyers, Burton H  
771 Singer, Olena Nesterova, et al. Disease burden among  
772 Ukrainians forcibly displaced by the 2022 Russian inva-  
773 sion. *Proceedings of the National Academy of Sciences*,  
774 120(8):e2215424120, 2023.
- 775 Poland's Data Portal. Statistical data on the situation on the  
776 border with Ukraine. [https://dane.gov.pl/en/dataset/2705\\_dane-statystyczne-dotyczace-sytuacji-na-granicy-z-Ukrainą](https://dane.gov.pl/en/dataset/2705_dane-statystyczne-dotyczace-sytuacji-na-granicy-z-Ukrainą),  
777 2022.
- 779 Zirou Qiu et al. Understanding the coevolution of mask wear-  
780 ing and epidemics: A network perspective. *PNAS*, 119(26),  
781 2022.
- 782 Debora Rizzato, René JF Melis, Sara Angleman, Chengxuan  
783 Qiu, and Alessandra Marengoni. Effect of chronic dis-  
784 eases and multimorbidity on survival and functioning in  
785 elderly adults. *Journal of the American Geriatrics Society*,  
786 65(5):1056–1060, 2017.
- 787 Zeynep Şahin-Mencütek. Conceptual complexity about re-  
788 turn migration of refugees/asylum seekers. *Journal of*  
789 *Asian and African Studies*, 59(7):2125–2138, 2024.
- 790 Rinku Saikia and Manash Pratim Barman. A review on accel-  
791 erated failure time models. *Int J Stat Syst*, 12(2):311–322,  
792 2017.
- 793 Ravenna Sohst, Tino Tirado, Lucía Salgado, and Jasmijn  
794 Slootjes. *Exploring Refugees' Intentions to Return to*  
795 *Ukraine: Data Insights and Policy Responses*. Migration  
Policy Institute Europe and International Organization for  
Migration, Brussels and Geneva, 2024.
- I. Sologoub. Return or stay? what factors impact the deci-  
sions of ukrainian refugees. VoxUkraine, 2024.
- Forum Transregionale Studien. Why Ukrainians return: Mo-  
tivations of return migration and its implications. <https://trafo.hypotheses.org/53767>, 2024.
- Krithika Suresh, Cameron Severn, and Debashis Ghosh. Sur-  
vival prediction models: an introduction to discrete-time  
modeling. *BMC medical research methodology*, 22(1):207,  
2022.
- Agnes Toth-Bos, Barbara Wisse, and Klara Farago. Goal pur-  
suit during the three stages of the migration process. *In-  
ternational Journal of Intercultural Relations*, 73:25–42,  
2019.
- UNHCR. Lives on Hold: Intentions and Perspectives  
of Refugees from Ukraine #3. [https://data.unhcr.org/en/  
documents/details/99072](https://data.unhcr.org/en/documents/details/99072), 2023.
- UNHCR. Ukraine Refugee Situation. [https://data.unhcr.org/en/  
situations/ukraine](https://data.unhcr.org/en/situations/ukraine), 2024. [Accessed December 30,  
2024].
- Thomas W Valente. Social network thresholds in the diffu-  
sion of innovations. *Social networks*, 18(1):69–89, 1996.
- Frank van Tubergen, Gusta G Wachter, Yuliya Kosyakova,  
and Irena Kogan. Return intentions among Ukrainian  
refugees in Europe: A cross-national study. *International  
Migration*, 62(5):181–198, 2024.
- Rob C Van Wijk and Ulrika SH Simonsson. Finding the  
right hazard function for time-to-event modeling: A tuto-  
rial and shiny application. *CPT: Pharmacometrics & Sys-  
tems Pharmacology*, 11(8):991–1001, 2022.
- Robert L Winkler, Javier Munoz, José L Cervera, José M  
Bernardo, Gail Blattenberger, Joseph B Kadane, Dennis V  
Lindley, Allan H Murphy, Robert M Oliver, and David  
Ríos-Insua. Scoring rules and the evaluation of probabili-  
ties. *Test*, 5:1–60, 1996.
- World Bank. Forced displacement, 2025. Accessed: 2025-  
02-08.
- Sean L Wu, Andrew J Dolgert, Joseph A Lewnard, John M  
Marshall, and David L Smith. Principled simulation  
of agent-based models in epidemiology. *bioRxiv*, pages  
2020–12, 2020.
- Nathan Wycoff, Lisa O Singh, Ali Arab, Katharine M Do-  
nato, and Helge Marahrens. The digital trail of ukraine's  
2022 refugee exodus. *Journal of Computational Social Sci-  
ence*, 7(2):2147–2193, 2024.
- Roger Zetter. Refugees and their return home: Unsettling  
matters. *Journal of Refugee Studies*, 34(1):7–22, 2021.
- Zhengyang Zhou, Yang Wang, Xike Xie, Lei Qiao, and Yuan-  
tao Li. Stuanet: Understanding uncertainty in spatiotempo-  
ral collective human mobility. In *Proceedings of the Web  
Conference 2021*, pages 1868–1879, 2021.

848 **A Additional Insights about Surrogate to the  
849 BASE**

850 **Theorem 1.** Let  $m(j)$  be the total number of migrants at  
851 time  $j$ , denoted as  $j^{\text{th}}$ -day migrants. Then, the number of  
852 returnees among the  $j^{\text{th}}$ -day migrants grow linearly at each  
853 time  $i > j$ , given  $h$  is small.

854 *Proof.* It can be shown that when  $h$  is small,  $r_{i+1,j} =$   
855  $r_{i,j} \frac{i+1-j}{i-j}$ , when  $i > j + 1$  (Proof below). Let  $r_{j+1,j}$  be  
856 the number of returnee of  $j^{\text{th}}$ -day migrant at time  $j + 1$ . It  
857 follows that,  $r_{j+k,j} = kr_{j+1,j}$ , which proves that the number  
858 of returnees from the  $j^{\text{th}}$  is a linear function with respect to  
859 their length of displacement before return.  $\square$

$$\begin{aligned} r_{i,j} &= \left( m(j) - \sum_{k < i} r_{k,j} \right) \left( 1 - e^{-h(i-j)} \right) \\ r_{i+1,j} &= \left( m(j) - \sum_{k < i} r_{k,j} - r_{i,j} \right) \left( 1 - e^{-h(i+1-j)} \right) \\ r_{i+1,j} &= \left( \frac{r_{i,j}}{1 - e^{-h(i-j)}} - r_{i,j} \right) \left( 1 - e^{-h(i+1-j)} \right) \\ r_{i+1,j} &= r_{i,j} \frac{e^{-h\Delta} (1 - e^{-h(\Delta+1)})}{1 - e^{-h\Delta}} \\ r_{i+1,j} &= r_{i,j} \frac{e^{-h\Delta} - e^{-h(2\Delta+1)}}{1 - e^{-h\Delta}} \\ r_{i+1,j} &\approx r_{i,j} \frac{(1 - h\Delta) - (1 - h(2\Delta+1))}{h\Delta} \\ r_{i+1,j} &\approx r_{i,j} \frac{h\Delta + h}{h\Delta} \\ r_{i+1,j} &\approx r_{i,j} \frac{\Delta + 1}{\Delta} \\ r_{i+1,j} &\approx r_{i,j} \frac{(i-j) + 1}{(i-j)} \end{aligned} \quad (5)$$

860 Here between Line 5 and 6, we have applied Taylor's ap-  
861 proximation, assuming  $h$  is small enough.

862 **B Bayesian Optimization**

863 The Bayesian optimization process begins by placing a prior  
864 on the parameter space. After the function is evaluated at  
865 some query parameter, a statistical model is used to map  
866 the candidate input (parameter space) to the output (objective  
867 function) and simultaneously obtain the posterior distribution  
868 of the parameter. Then, an acquisition function, is used to  
869 sample the next query point(s). This process is continued it-  
870 eratively until a stopping criterion is met.

871 Since Bayesian optimization is classically a maximization  
872 task, we use  $-L(\theta)$  as the objective function during the  
873 optimization procedure. We use Gaussian process estimator as  
874 statistical model for constructing the posterior and expected  
875 improvement as the acquisition function.

876 **C Insight about our Proposed Loss Function**

877 Our proposed loss function is as follows.

$$L(\theta) = \frac{\lambda_e}{T^2} \sqrt{\sum_{i=1}^T (r_\theta(i) - \hat{r}(i))^2} + \lambda_c \left( 1 - \frac{\sum_{i=1}^T (r_\theta(i) - \mu(R_\theta))(\hat{r}(i) - \mu(\hat{R}))}{\sqrt{\sum_{i=1}^T (r_\theta(i) - \mu(R_\theta))^2} \sqrt{\sum_{i=1}^T (\hat{r}(i) - \mu(\hat{R}))^2}} \right) \quad (6)$$

which we write simplifying as:

$$L(\theta) = \lambda_e \text{NRMSE}(f_\theta, \hat{R}) + \lambda_c (1 - \text{PCC}(f_\theta, \hat{R})) \quad (7)$$

Where  $f_\theta$  corresponds to all time-series predictions  
 $\{r_\theta(1), r_\theta(2), \dots, r_\theta(T)\}$  produced by some function  $f$  pa-  
rameterized by  $\theta$  that tries to generate results similar to  $\hat{R} =$   
 $\{\hat{r}(1), \hat{r}(2), \dots, \hat{r}(T)\}$

We want to counter the argument that if  $\theta$  is a local minimum with respect  $\text{NRMSE}(f_\theta, \hat{R})$ , it is also a local minimum with respect to  $L(\theta)$ . If this argument holds true, then essentially our loss function do not provide any additional benefit over traditional RMSE loss.

Table 3: Empirical Verification of unsatisfaction of Equation (9) within a feasible range of  $\theta$

$\theta$	$\frac{\nabla \sigma_{f,\hat{R}}}{\sigma_{f,\hat{R}}}$	$\frac{\nabla \sigma_f}{\sigma_f}$	$\nabla \text{RMSE}(\theta)$
0.000100	4364.290874	2908.489304	-5088.988560
0.000150	793.349642	-384.276087	-3685.474780
0.000200	-1362.776294	-2090.909702	-2566.156525
0.000250	-3180.140218	-2453.694803	-1086.638762
0.000300	-4980.408037	-1341.170406	707.586293
0.000350	-7602.458344	486.004014	1302.247636
0.000400	-13579.184000	2092.404581	1375.002277
0.000450	-41855.273090	2644.295103	1271.565101
0.000500	38370.934346	2489.975694	1135.966719

If  $\theta$  is a local minimum with respect to the NRMSE error, it should hold that  $\nabla \text{NRMSE}(f_\theta, \hat{R}) = 0$ . In order for  $L(\theta)$  to also be a local minimum with respect to  $\theta$ , it should thus hold that  $\nabla \text{PCC}(f_\theta, \hat{R})$  should be 0. Our goal is to find whether this condition satisfies. Countering this argument can therefore justify the proposed loss function helping the model getting out of local minima with just RMSE loss.

**Definition of the Pearson Correlation Coefficient (PCC)**

The Pearson correlation coefficient between the model pre-  
dictions  $f(\theta)$  and the ground truth  $\hat{R}$  is given by

$$\text{PCC}(f(\theta), \hat{R}) = \frac{\sigma_{f\hat{R}}}{\sigma_f \sigma_{\hat{R}}}$$

where:

- $\sigma_{f\hat{R}} = \mathbb{E}[(f(\theta) - \mu_{\hat{R}})(\hat{R} - \mu_{\hat{R}})]$  is the covariance between  $f(\theta)$  and  $\hat{R}$ ,
- $\sigma_f = \sqrt{\mathbb{E}[(f(\theta) - \mu_f)^2]}$  is the standard deviation of  $f(\theta)$ ,
- $\sigma_{\hat{R}} = \sqrt{\mathbb{E}[(\hat{R} - \mu_{\hat{R}})^2]}$  is the standard deviation of  $\hat{R}$  (constant with respect to  $\theta$ ),
- $\mu_f = \mathbb{E}[f(\theta)]$  and  $\mu_{\hat{R}} = \mathbb{E}[\hat{R}]$  are the means of  $f(\theta)$  and  $\hat{R}$ , respectively.

Next, we try to compute the gradient of the PCC with respect to  $\theta$ :

$$\nabla \text{PCC}(f(\theta), \hat{R}) = \frac{1}{\sigma_{\hat{R}}} \frac{\sigma_f \nabla \sigma_{f\hat{R}} - \sigma_{f\hat{R}} \nabla \sigma_f}{\sigma_f^2}$$

Now,

$$\begin{aligned} \nabla L(\theta) = 0 &\implies \nabla \text{PCC}(f(\theta), \hat{R}) = 0 \\ &\implies \frac{1}{\sigma_{\hat{R}}} \frac{\sigma_f \nabla \sigma_{f\hat{R}} - \sigma_{f\hat{R}} \nabla \sigma_f}{\sigma_f^2} = 0 \quad (8) \\ &\implies \sigma_f \nabla \sigma_{f\hat{R}} - \sigma_{f\hat{R}} \nabla \sigma_f = 0 \end{aligned}$$

Note that,  $\sigma_f$  and  $\sigma_{\hat{R}}$  can only be zero if all the data are the same which is not possible in a dynamic scenario like migration. Rearranging the last line, we obtain the following condition for  $\theta$  to be local minima w.r.t to the proposed loss function.

$$\frac{\nabla \sigma_{f\hat{R}}}{\sigma_{f\hat{R}}} = \frac{\nabla \sigma_f}{\sigma_f} \quad (9)$$

Table 4: Empirical Verification of unsatisfaction of Equation (9) around one local minima of  $\text{RMSE}(\theta)$

$\theta$	$\frac{\nabla \sigma_{f\hat{R}}}{\sigma_{f\hat{R}}}$	$\frac{\nabla \sigma_f}{\sigma_f}$	$\nabla \text{RMSE}(\theta)$
0.0002765	-4095.029739	-1979.413278	-17.831328
0.0002766	-4098.525636	-1976.544746	-13.966556
0.0002767	-4102.023035	-1973.670578	-10.105442
0.0002768	-4105.521994	-1970.790820	-6.248005
0.0002769	-4109.022415	-1967.905394	-2.394286
0.0002777	-4112.524437	-1965.014419	1.455699
0.0002771	-4107.421302	-1941.869671	6.584414
0.0002772	-4151.437528	-1986.601121	18.054937
0.0002773	-4154.970180	-1983.640920	21.890979
0.0002774	-4158.504450	-1980.675017	25.723123
0.0002775	-4162.040293	-1977.703386	29.551335

Thus, our loss function will not be better than traditional RMSE loss if this is satisfied. Now, computing the gradients of  $\sigma_{f\hat{R}}$  and  $\sigma_f$  is difficult. However, we can compute the approximate gradient using the slope method by comparing the quantities at  $\theta$  and  $\theta + \Delta$  where  $\Delta$  is a very small quantity. We empirically validated this using the  $C^{35}$  surrogate version

over various  $\theta$ . This model has only a single hazard rate parameter, thus  $\theta$  is a single scalar value. Table 3 shows the LHS and RHS of Equation (9) for a set of values of  $\theta$ . It can be observed that the two quantities are never close. For comprehensiveness, we take our empirical verification another step forward. We see that the sign of  $\nabla \text{RMSE}(\theta)$  changes between  $\theta = 0.00025$  and  $\theta = 0.0003$ . We perform a more granular search within this region to see whether Equation (9) is satisfied in regions around local minima. From Table 4, we can confidently argue that the two quantities do not become the same around the local minima. This can be said because these quantities are more or less stable within this range. However, since these are different quantities, Equation (9) is not satisfied.

Thus, we have justified that our proposed loss function may help the model get out of local minima that could occur with just using RMSE loss.

## D Evaluation Metric

We use three metrics to evaluate the quality of our estimates with respect to the observations. The first metric is Normalized Root Mean Squared Error (NRMSE). Let,  $Y$  be the set of all observations over all timesteps. Let,  $y(t) \in Y$  be the observed value at time  $t$  and  $\tilde{y}(t)$  denote the corresponding forecast value. Thus, the NRMSE is defined as follows:

$$\text{NRMSE} = \frac{1}{\max(Y) - \min(Y)} \sqrt{\frac{\sum_t (y(t) - \tilde{y}(t))^2}{|Y|}}$$

While NRMSE can capture the quality of the point prediction, it cannot take into account the uncertainty associated with it. We utilize the concept of the Continuous Ranked Probability Score (CRPS) Winkler *et al.* [1996] to address this. Let us assume that we have  $M$  forecasts  $\hat{y}(t)^1, \hat{y}(t)^2, \dots, \hat{y}(t)^M$  corresponding to observation  $y(t)$ . First, the empirical cumulative distribution function (CDF) at a point  $z$ ,  $\hat{F}_t^M(z)$  from these forecast values is defined as

$$\hat{F}_t^M(z) = \frac{1}{M} \sum_{n=1}^M \mathbb{I}(\hat{y}(t)^n \leq z) \quad (10)$$

where  $\mathbb{I}(\cdot)$  is the indicator function that yields 1 if the input argument holds, 0 otherwise. From this, CRPS between the empirical CDF  $\hat{F}_t^M$  and the observation  $y(t)$  is calculated as

$$\text{CRPS}(\hat{F}_t^M, y(t)) = \int_z \left( \hat{F}_t^M(z) - \mathbb{I}(y(t) \leq z) \right)^2 dz \quad (11)$$

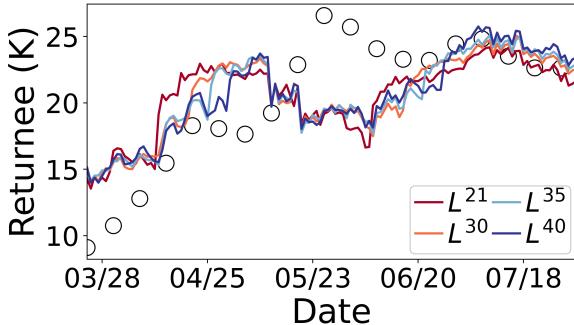
To account for scaling, we extend this metric to calculate the Normalized Average CRPS (NACRPS), defined as

$$\text{NACRPS}(\hat{F}_t^M, Y) = \frac{\sum_t \text{CRPS}(\hat{F}_t^M, y(t))}{\sum_t |y(t)|} \quad (12)$$

The third metric we evaluate is PCC, to understand how well the model can capture the trend in the observed data. The definition of PCC can be found in Section C

## 961 E Performance of Vanilla regression

962 Figure 4 shows the performance of naive baseline regressions  
963 in estimating number of returnees based on past  $D$  days of  
964 conflict event counts and fatalities. It has to be noted that an  
965  $L^D$  regression model has  $2D+1$  trainable parameters. On the  
966 other hand, the highest number of parameters in our proposed  
967 model is 3, which is for any  $C^L P$  model.



968 Figure 4: Return estimation trained with regression analysis using  
969 historical conflict as features

## 968 F Digital Library of ABM-TPB

969 Figure 5 shows the results of 100 simulations of ABM-TPB.  
970 Out of these simulations, the output of the best 30 simulations  
971 in terms of RMSE were used as inputs to our models.

## 972 G Additional Experiments

### 973 G.1 Uncertainty Quantification of Calibration

974 As an alternative way to assess the quality of the calibration  
975 of the models producing probabilistic estimates, we compare  
976 the proportion of observations covered by various  
977 intervals around the central prediction (nominal interval  
978 coverage). Ideally, the central 50% prediction interval should  
979 contain around 50% of the observed values, the 90% inter-  
980 val should contain around 90% of observations, and so on,  
981 given the model is well calibrated. Figure 6 compares the  
982 nominal interval coverages of each ABM and its correspond-  
983 ing surrogate across various quantiles. Apart from BASE, all  
984 the models are in the upper diagonal region, indicating that  
985 the probabilistic predictions are too widespread (i.e., under-  
986 confident); they cover more observations than they should.  
987 Aside from  $C^{21}P$  model, all the ABMs are in general simi-  
988 lar or closer to the central diagonal line than their surrogate  
989 counterpart, indicating that they are better calibrated for prob-  
990 abilistic forecasts. Finally, BASE seems to be best calibrated  
991 to perform probabilistic forecasts. This is expected, as it is  
992 the simplest model with only one parameter to calibrate.

### 993 G.2 Ablation Study of Loss Function

994 We start by choosing a fixed value of the error coefficient  
995 ( $\lambda_e = 0.8$ ). Then we calibrate a set of surrogate models with  
996 different lags of conflict contexts upto 300 iterations while  
997 varying  $\lambda_c$  uniformly within 0 and 1. After the calibration

998 process, we examine the NRMSE and PCC of the calibrated  
999 models in Figures 7a and 7b to make several key insights.

1000 *First,,* We see that for very small non-negative weights of  
1001  $\lambda_c$  ( $\lambda_c \leq 0.02$ ), the calibrated model produces better results  
1002 in terms of NRMSE than  $\lambda_c = 0$ . Simultaneously, the PCC of  
1003 the calibrated model increases as  $\lambda_c$  increases. When  $\lambda_c = 0$ ,  
1004 the entire weight is given to minimize the MSE loss. Thus,  
1005 the model may not explore parameter regions beyond a lo-  
1006 cal minima. On the contrary, putting a small weight on the  
1007 correlation coefficient may make the model explore a broader  
1008 parameter space since it can get penalized due to the corre-  
1009 lation coefficient term in the loss function, allowing it to find  
1010 other regions where NRMSE is lower. However, it can be  
1011 very well the case that with more iterations, the model would  
1012 have found these regions even with  $\lambda_c = 0$ . Nonetheless,  
1013 this insight highlights that a small weight on the loss function  
1014 makes the model find better solutions with fewer calibration  
1015 steps. This can also be observed in the case of the ABM (Fig-  
1016 ure 7c) where we see that  $0 < \lambda_c \leq 0.2$  actually improves  
1017 the model NRMSE.

1018 *Second,* in terms of the accuracy-correlation tradeoff, we  
1019 find that up to  $\lambda_c \leq 0.2$ , the NRMSE of the surrogate models  
1020 slowly degrades while the PCC improves. For example, for  
1021 the model with 50 days lag period, as  $\lambda_c$  increases, compared  
1022 to the case of  $\lambda_c = 0$ , an additional 9.4% improvement can  
1023 be obtained in terms of PCC while the NRMSE declines by  
1024 2.62%. As another example, an additional improvement of  
1025 4.5% is gained in PCC without any penalty to the NRMSE  
1026 for the model with 45 days lag period. This highlights another  
1027 policy-relevant aspect of tuning our loss function based on the  
1028 specific policy use-cases. When policymakers want to get a  
1029 good estimate about the number of return migrants, the model  
1030 can be optimized with very small or no correlation coefficient.  
1031 On the other hand, if the policymakers want to detect surges  
1032 in the daily inflow of the return migrants, small or moderate  
1033  $\lambda_c$  can provide them with a model well calibrated to capture  
1034 such trends. Thus, our proposed loss function can provide the  
1035 necessary flexibility to policymakers based on their specific  
1036 use-cases.

1037 *Finally,* for the surrogate models, it is worth noting that  
1038 when the value of  $\lambda_c$  exceeds a certain point, the accuracy-  
1039 correlation tradeoff may not be feasible any more. In the case  
1040 of surrogate models, we try to highlight this by two dashed  
1041 lines. The left side of the first dashed line represents regions  
1042 where choosing a  $\lambda_c > 0$  can help the model improve both  
1043 its NRMSE and PCC. Region within the two dashed lines is  
1044 where the model can learn to fit the correlation better, with  
1045 a small penalty to accuracy. Finally, at the right side of the  
1046 second dashed line, the improvement of a PCC may come  
1047 with a heavy penalty in the accuracy. While we do not find a  
1048 similar region in the case of the ABM since we do not eval-  
1049 uate its learning against many values of  $\lambda_c$  like the surrogates  
1050 due to the computational expense, we can still see that above  
1051 the value of  $\lambda_c > 0.3$ , there is a significant increase in the  
1052 NRMSE which may not be feasible compared to the improve-  
1053 ment in PCC; although notably the NRMSE is always bet-  
1054 ter compared to the case when learning was conducted with  
1055  $\lambda_c = 0$ .

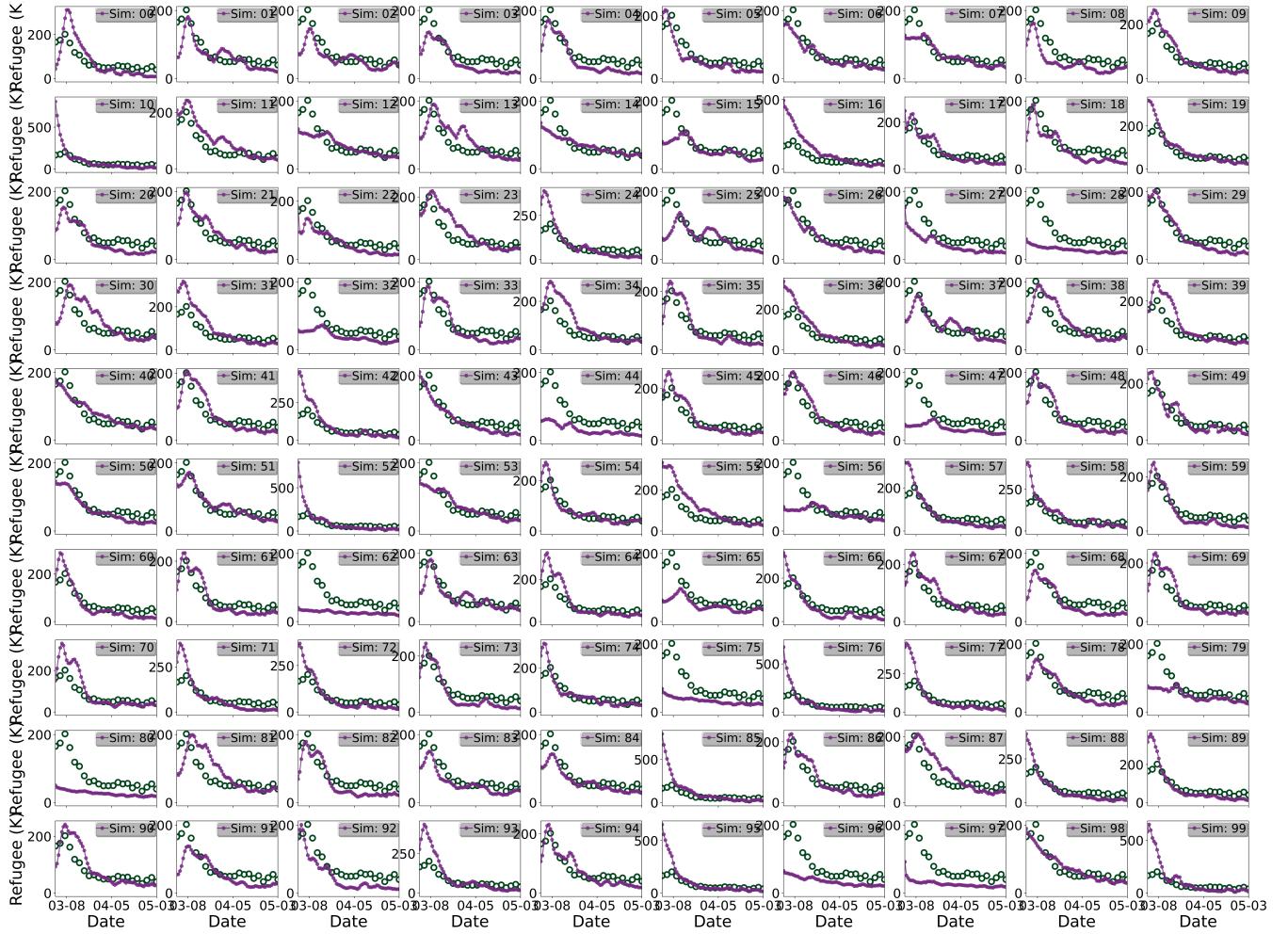


Figure 5: Digital Library of simulations from the ABM-TPB model. The green dots represent the ground truth corresponding to total migrant outflow from Ukraine. Figures only show simulation up to May 06, 2022 although for our purpose, the simulation was continued until end of July 2022.

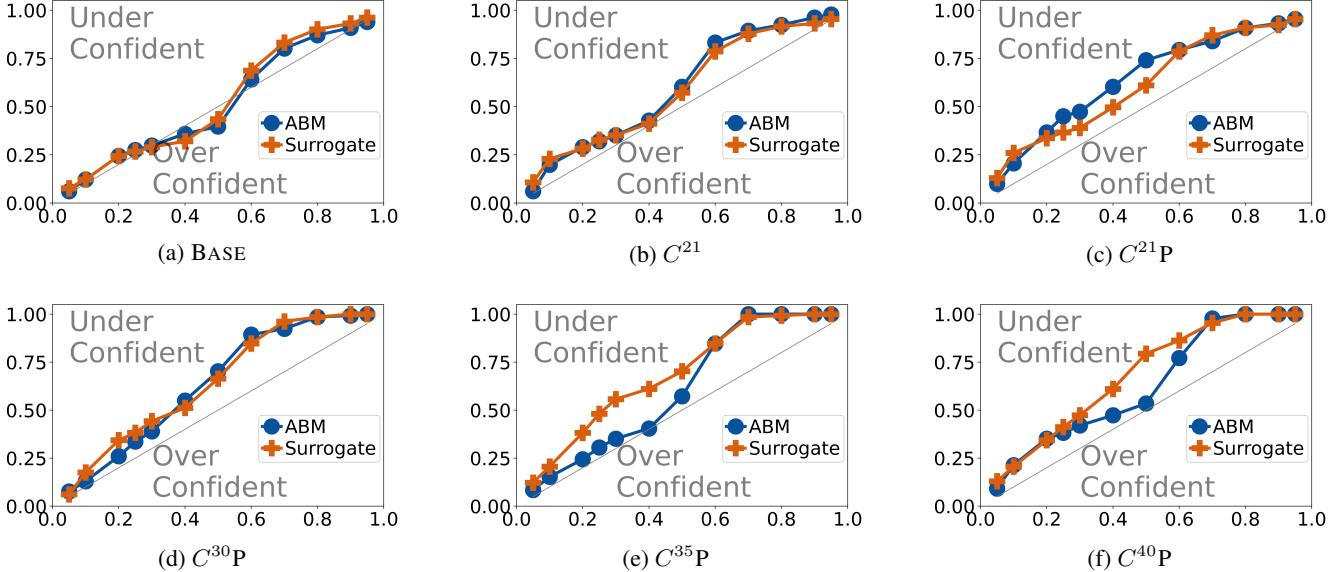


Figure 6: Nominal Interval Coverages of various models.

### G.3 Ablation Study of Model Components

This section studies the effect of introducing different layers of complexity into the model. The BASE only applies a constant hazard rate over time across all agents uniformly, with the only difference between agents being their time of migration. From Figure 8a, we see that the bell-shaped return estimated by this simplest model captures the general migration pattern reasonably well with an NRMSE of 0.1203 and a PCC of 0.947. However, this model is not able to capture the more intricate dynamics of return migration (e.g., the two peaks around Late April and Late May). Once conflict context is accommodated, we see the more intricate dynamics in Figure 8b. Adding peer influence improves the NRMSE slightly (from 0.112 to 0.106) because the CI of estimation is able to better capture the later points in the observations.

Notably, although adding layers of complexity makes the model appear visually convincing, there are drawbacks which has caused the PCC to drop (Figure 8a vs Figure 8b and Figure 8c). Careful analysis reveals that the primary reason behind this is due to the underestimation of the first peak of returnees around late May. We address this further in the subsequent section when we analyze the effect of  $\mathcal{L}$ .

This highlights another policy-relevant aspect of tuning our loss function based on the specific policy use-cases. When policymakers want to get a good estimate of the number of return migrants, the model can be optimized with a very small  $\lambda_c$ . On the other hand, if the policymakers want to detect surges in the daily inflow of the return migrants, moderate  $\lambda_c$  can be used. Thus, our proposed loss function can provide the necessary flexibility to policymakers based on their specific use-cases.

### G.4 Effect of Lag $\mathcal{L}$

The lag value  $\mathcal{L}$  is one of the key parameters in the hazard-based model. A journey like migration requires planning

and time, be it migration out of country or return migration. Therefore, this lag parameter is insightful in identifying when the decision for return migration is actually taken and when the migration actually took place. Previous experiments were conducted with  $\mathcal{L} = 21$  days. We empirically conducted experiments with other values of  $\mathcal{L}$ . Figure 9a, 9b, 9c show the return estimates for lag values of  $\mathcal{L} = 30, 35, 40$  days, respectively. We find that the model performs best when conflict context is assessed with a 30 day period of lag. For higher lag values, the performance starts declining.

These observations indicate that there is likely a lag period of a month that goes into the planning of return migration and conducting the migration. This is high compared to the lag value reported for forward migration in earlier research Mehrab *et al.* [2024b]. Since return migration is more proactive than reactive forward migration, it makes sense that more time would go into planning Al Husein and Wagner [2023]. We acknowledge that during this lag, people's decisions may change and some of the migrants who planned for return migration may not end up returning. However, such analysis is out of the scope of our current study and requires a greater understanding of human cognition and potentially other political contexts.

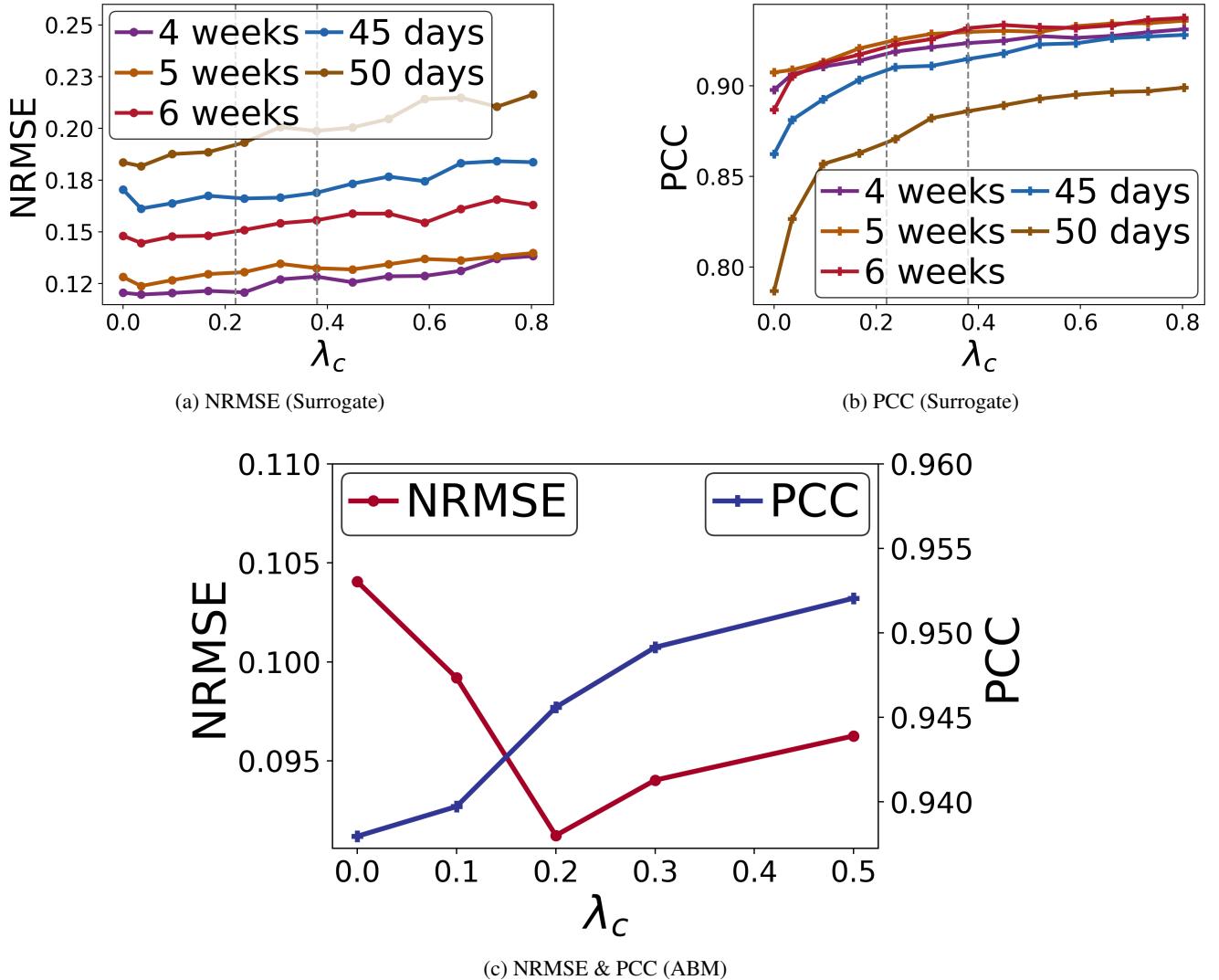


Figure 7: Ablation study of correlation coefficient ( $\lambda_c$ ) for different surrogate models (Figure 7a and 7b) and C<sup>30</sup>P ABM (Figure 7c). For the surrogate models, 300 iterations is used for calibration,  $\lambda_e$  is set to 0.5 and  $\lambda_c$  is varied uniformly in [0,1]. The different lines are for surrogate model with different lags in the conflict context (Lags mentioned in legend). The ABM is calibrated for 100 iterations, with  $\lambda_e$  set to 0.8 and  $\lambda_c \in \{0, 0.1, 0.2, 0.3, 0.5\}$ .

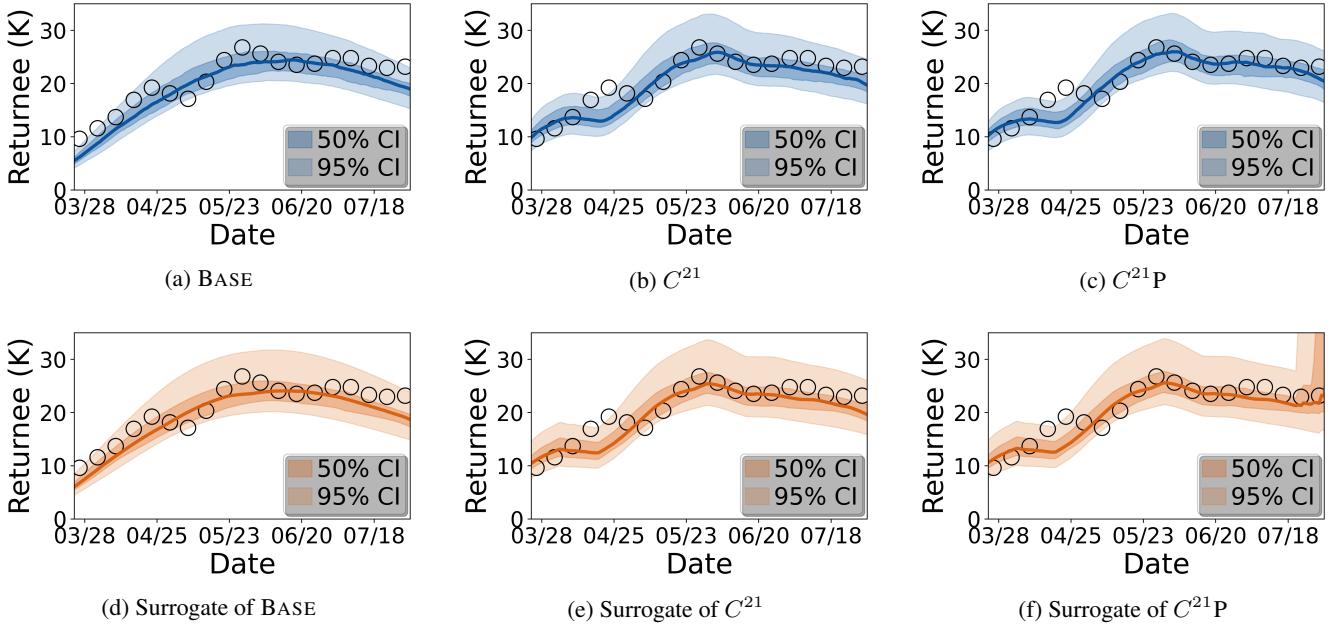


Figure 8: Ablation Study of adding different layers of sociopolitical contexts to the models. Top row corresponds to the ABMs and bottom row corresponds to their surrogate counterparts. Circles denote ground truth data. Solid line corresponds to the median estimate of the model.

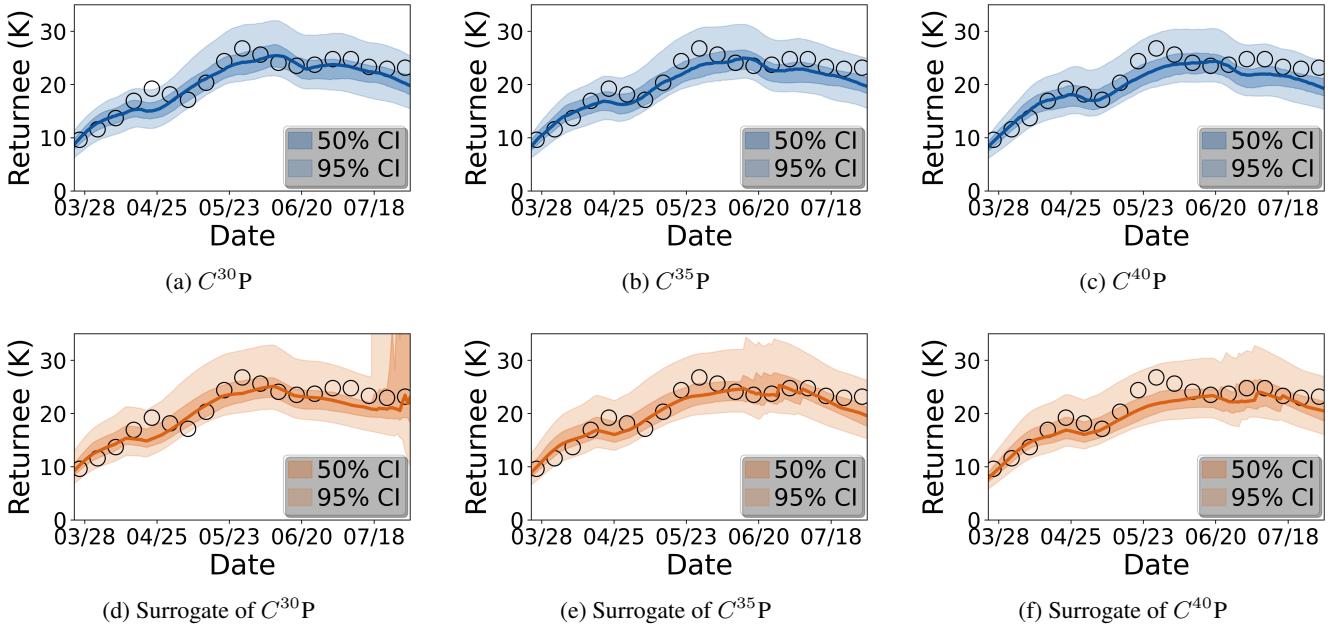


Figure 9: Performance of our models for different  $\mathcal{L}$  parameters.