POLITECNICO DI TORINO

Master's Degree in Mathematical Engineering

Modeling Urban Butterflies in Barcelona: An ODE-Based Spatial Graph Approach



Supervisors

Prof. Luigi Preziosi
Prof. Giovanni Dalmasso
Prof. Yolanda Melero

Candidate

Francesca Pozzo

A.y. 2025/2026
Graduation Session July 2026

Abstract

Butterflies are widely recognised as sensitive bioindicators of ecosystem health, whose distribution reflects both climatic and urban changes. Understanding how butterflies move across fragmented urban landscapes is therefore an important challenge for biodiversity conservation.

This thesis presents an ongoing research project aimed at studying butterfly movement between urban parks in Barcelona through a mathematical modelling framework. Data are provided by the Urban Butterfly Monitoring Scheme (uBMS), which monitored 24 parks between 2018 and 2024. Population abundances are reconstructed using a neural-network-based pipeline, producing continuous daily time series from field observations.

Since direct data on butterfly movements are not available, the main objective is to infer inter-park flows from observed population dynamics. To this end, a system of ordinary differential equations is coupled with a graph representation of the urban landscape, where nodes correspond to parks and edges describe potential movement based on distance, connectivity, and habitat characteristics. The model is developed progressively, from a single-park formulation to a fully connected network that can also account for external inflows from surrounding natural areas, such as the Collserola Natural Park. Model parameters are estimated through numerical optimisation.

The inferred flow network is then used to investigate the role of different parks and identify potential ecological traps. Preliminary results suggest that incorporating spatial movement improves the model's ability to reproduce observed dynamics and highlights the importance of landscape structure in shaping butterfly movements across the city. Overall, the framework provides a quantitative tool for assessing movement patterns and habitat functionality in urban environments, supporting biodiversity-oriented management of urban green spaces.

Table of Contents

List of Figures 4

1 Introduction 8
- 1.1 Biological background . . . 9
- 1.2 Mathematical foundations: literature and reference . . . 13
- 1.3 The Project . . . 14
  - 1.3.1 Thesis structure . . . 16

2 Data 17
- 2.1 Source of information . . . 17
  - 2.1.1 Counting methodology . . . 19
- 2.2 Butterfly: observations and categorization . . . 20
- 2.3 Space level: patches . . . 28
- 2.4 Detectability and Data Uncertainty . . . 45
- 2.5 Neural Network Data . . . 47

3 Mathematical Model 50
- 3.1 Procedural idea . . . 50
- 3.2 Population dynamics . . . 52
- 3.3 Connections . . . 55
  - 3.3.1 Graph . . . 56
- 3.4 Mathematical analysis . . . 61
  - 3.4.1 Evaluation . . . 61
  - 3.4.2 Trap index . . . 83
- 3.5 Summary . . . 86
  - 3.5.1 Constraints . . . 89

4 Implementation 91
- 4.1 Python . . . 92

5 Results and Discussion 94
5.1 Phase 0 . . . 95
5.2 Phase 1 . . . 101
5.3 Phase 2 . . . 107
5.4 Comparison of Model Phases . . . 113
5.5 Flux analysis . . . 117

6 Conclusions 121
6.1 Final consideration . . . 124
6.2 Future development . . . 126

A Mathematical Background 128
A.1 Graph theory . . . 128
A.2 Softmax function . . . 130
A.3 Existence and uniqueness of ODE solutions . . . 131
A.4 Perron–Frobenius theorem . . . 131
A.5 The L-BFGS-B Optimization Algorithm . . . 132
A.6 Mathematical formulation of the goodness-of-fit metrics . . . 133
A.6.1 Mean Squared Error (MSE) . . . 133
A.6.2 Normalized Mean Squared Error (nMSE) . . . 134

B Additional results 135
B.1 Phase 0 . . . 135
B.2 Phase 1 . . . 137
B.3 Phase 2 . . . 140

List of Figures

1.1 Butterfly life cycle. 10

2.1 Map of Barcelona showing the monitoring sites in yellow and the city's green areas in green. 18
2.2 Visual example of the counting methodology (uBMs). 19
2.3 Illustration of voltinism: (1) univoltine, (2) multivoltine (AI-generated image). . 20
2.4 *Pieris rapae*. Photo by Jordi Sánchez. 27
2.5 Scatter plot of Pieris rapae counts in 2022 in Parc de la Trinitat. . . 27
2.6 Scatter plot of Pieris rapae counts in Parc de la Trinitat over all the years studied. 27
2.7 Scatter plot of Pieris rapae counts 2023. 28
2.8 Map of the 24 Barcelona parks considered [1]. 30
2.9 Barcelona land cover map used for urbanised distance computation [1]. Green: vegetated areas (low impedance = 1); yellow: open/cleared surfaces (medium impedance = 5); red: built surfaces and roads (high impedance = 10). 34
2.10 Urbanisation map of Barcelona [2]. 36
2.11 Urbanisation evolution in Parc de la Trinitat from 2018 to 2024. . . 37
2.12 The increase in urbanization in parks through the growth rate of urbanization. 37
2.13 Urbanisation trend in all 24 parks in Barcelona over the years. . . . 38
2.14 Example of temperature in a Barcelona park over years 2018-2024 in Parc de la Trinitat. 39
2.15 Calculation of isolation between parks based on euclidean distance. 42
2.16 Map of the Collserola park territory. 43
2.17 Collserola polygon representation with Geojson.io. 43
2.18 Graphic representation of distances parks-Collserola (outside Barcelona). 45
2.19 Neural network layers. 47
2.20 Basic demonstration of neural network results [3]. 48
2.21 Neural network data (blue) from the Pieris rapae species in Parc de la Trinitat with background temperature (orange) [3]. 48

2.22 Neural network data (blue) from the Pieris rapae species in all 24 parks with background temperature (orange)[3]. 49

3.1 Schematic representation of Barcelona as a graph. 56

3.2 Synthetic-data demonstration with phase 0 of the upper bound on ω^s. As ω^s approaches the stability threshold of 2.785, the model's ability to reproduce a biologically meaningful trajectory degrades visibly; beyond this value, the numerical scheme becomes unstable. 88

5.1 Frequency distribution of the Phase 0 parameter estimates obtained from the optimization procedure over 100 independent runs (the marked value corresponds to the solution producing the lowest loss). 96

5.2 Phase 0 loss values obtained from the optimization procedure over the 100 independent runs. 97

5.3 Phase 0 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile. 98

5.4 Fit quality with mse phase 0. 99

5.5 Phase 0. On the left, Jardí del Laberint is shown, a coastal park in Barcelona. On the right, Parc de Diagonal Mar is displayed; as indicated by the inset map, it is located at the northern edge of Barcelona, near the boundary with Collserola Park. 100

5.6 Illustration of the relationship between transport and isolation based on connectivity in phase 0. The x-axis shows connectivity values, ranging from more isolated parks (lower values, left) to more connected parks (higher values, right). 101

5.7 Frequency distribution of Phase 1 parameter estimates over 100 optimization runs (the highlighted value corresponds to the minimum-loss solution). 102

5.8 Phase 1 loss values obtained from the optimization procedure over the 100 independent runs. 103

5.9 Phase 1 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile. 104

5.10 Fit quality with mse phase 1. 105

5.11 Left: Phase 1 representation of Jardins de Joan Maragall, an inland park in Barcelona situated in close proximity to several other green areas. Right: Jardins de la Maternitat, located within Barcelona's urban network in a non-central area of the city and closer to Collserola Park, as indicated by the inset map. 106

5.12 Relationship between transport and connectivity-based isolation in phase 1. The x-axis shows connectivity values, from more isolated (left) to more connected (right) parks. 107

5.13 Frequency distribution of Phase 2 parameter estimates over 100 independent runs; the highlighted value corresponds to the minimum-loss solution. 108

5.14 Phase 2 loss values obtained from the optimization procedure over the 100 independent runs. 109

5.15 Phase 2 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile. 110

5.16 Fit quality with mse phase 2. 111

5.17 Phase 2 fitting. Left: Jardins de Joan Maragall, an inland park surrounded by several neighbouring green areas. Right: Jardins de la Maternitat, located in an off-centre position and in closer proximity to Collserola Park (see inset map). 112

5.18 Illustration of the relationship between transport and isolation based on connectivity in phase 2. The x-axis shows connectivity values, ranging from more isolated parks (lower values, left) to more connected parks (higher values, right). 113

5.19 Evolution of the nMSE across the three model phases for each of the 24 parks considered. 115

5.20 Ranking of parks by net flow: sites experiencing net immigration are shown in red, while those experiencing net emigration are shown in blue. 117

5.21 Attractiveness ranking of the parks: sites with high immigration, as identified in figure 5.20, are shown in red, while those with high emigration are shown in blue. 118

5.22 Sites ranked by attraction (share of network movement flow converging on each site, from the column sums of W). 119

5.23 Trap index and sink severity by park, ranked from lowest to highest trap index. The solid colored bar shows the trap index, the final metric quantifying how strongly a park behaves as an ecological trap, while the dashed black outline shows the sink severity, the raw local-deficit component before being weighted by the park's attractiveness in the movement network. Colors indicate the final source/sink/trap classification. 120

6.1 Map of the Barcelona coastal territory (orange boundary) and the surrounding natural areas (red and green for Collserola Park), which represent a major source of individuals for the system dynamics beyond the city limits [4]. 126

B.1 Phase 0, initial values of the biological parameters a, T^*, σ_T and ω^s across the 100 optimisation runs. 135
B.2 Phase 0, distribution of the initial parameter values sampled across the 100 optimisation runs. 136
B.3 Phase 0, optimised values of the biological parameters a, T^*, σ_T and ω^s across the 100 optimisation runs. Bar colour encodes the normalised loss of the corresponding run, and the star marks the run achieving the global best fit. 137
B.4 Phase 1, initial values of the biological parameters a, T^*, σ_T, ω^s, μ and δ across the 100 optimisation runs. 138
B.5 Phase 1, distribution of the initial parameter values sampled across the 100 optimisation runs. 139
B.6 Phase 1, optimised values of the biological parameters a, T^*, σ_T, ω^s, μ and δ across the 100 optimisation runs. 140
B.7 Phase 2, initial values of the biological parameters a, T^*, σ_T, ω^s, μ, γ and ρ across the 100 optimisation runs. 141
B.8 Phase 2, distribution of the initial parameter values sampled across the 100 optimisation runs. 142
B.9 Phase 2, optimised values of the biological parameters a, T^*, σ_T, ω^s, μ, γ and ρ across the 100 optimisation runs. Bar colour encodes the normalised loss of the corresponding run, and the star marks the run achieving the global best fit. 143

Chapter 1

Introduction

In an increasingly urbanised world, where attention to climate change remains insufficient, it is essential to make use of all available sources of ecological information. In this context, butterflies play a crucial role. Despite their small size, they are highly sensitive to even minor environmental changes and are, alongside birds, among the most commonly used taxa for monitoring habitat alteration. For this reason, they are considered powerful bioindicators of ecosystem health [1].
More than ever, butterflies provide an effective means of assessing human-driven habitat modification, as they closely reflect both climatic and urban transformations. This is particularly relevant given that urban expansion is projected to substantially increase global urban cover by at least 2050 [5, 6]. As cities continue to grow, they will play an increasingly central role in biodiversity conservation, with the potential to support a wide range of plant and animal species. Nevertheless, urbanisation generally has negative effects on species richness and abundance across many taxa, including invertebrates, birds, and mammals. From an ecological and evolutionary perspective, urban environments create novel conditions that act as selective filters on species traits. Species whose characteristics are poorly suited to urban habitats may decline, whereas those with traits such as trophic generalism, high dispersal ability, thermophily, and elevated reproductive rates are more likely to persist. This process often leads to functional community homogenisation, in which urban communities become dominated by species sharing similar ecological traits. Consequently, the structure and composition of urban landscapes strongly shape biodiversity patterns and determine which organisms are able to persist in cities [1].
Monitoring butterfly communities and tracking population trends over time can therefore provide important insights into environmental change. Within this framework, increasing attention has been directed toward urban biodiversity and the development of urban planning strategies capable of supporting native species, particularly through the preservation and enhancement of urban green spaces. In

this context, diurnal butterflies are especially valuable not only for their ecological role as pollinators and their aesthetic significance, but also because they are widely recognised as reliable indicators of biodiversity. Monitoring butterfly populations provides insight into broader ecological dynamics, including changes in species richness and abundance, as well as environmental pressures that may also affect less visible organisms. Collecting long-term monthly data on butterfly presence and abundance makes it possible to characterise urban biodiversity patterns, analyse their temporal dynamics, and identify the urban features that promote or constrain biodiversity. These observations can then inform management strategies for urban green spaces and the surrounding landscape.

From this perspective, Barcelona represents a particularly valuable case study for investigating these dynamics, thanks to the availability of long-term ecological data collected within its urban green spaces. Such datasets make it possible to move from general ecological considerations on urban biodiversity to a more detailed analysis of how species respond to urban environmental conditions at the local scale. Building on this, the present project develops a mathematical modelling study based on real ecological data from Barcelona, with the aim of understanding how butterfly populations move across the urban landscape. In particular, the research seeks to investigate the ecological processes underlying butterfly movement between urban green spaces, with the broader goal of identifying how different habitat patches contribute to population persistence. To address this question, butterfly population dynamics are modelled through a system of ordinary differential equations coupled with a graph-based representation of the urban landscape, allowing the spatial structure of habitat patches and the flows between them to be incorporated. This framework provides a basis for exploring metapopulation structure, dispersal dynamics, and the potential presence of source habitats, sink habitats, or ecological traps associated with urbanisation and related environmental change.

1.1 Biological background

As mentioned above, butterflies were selected as model species, due to their well-documented ecological relevance and their ability to capture fine-scale variation in environmental conditions. Their diverse life-history traits make them suitable representatives of many insect groups and other short-lived organisms, enabling robust ecological inference across heterogeneous habitats [1]. In addition, their population dynamics are closely linked to habitat quality, resource distribution, and landscape structure. For this reason, a detailed understanding of their life

cycle is fundamental for interpreting spatial and temporal patterns in population variation and for supporting quantitative analyzes of ecological dynamics in fragmented environments.

Butterflies undergo a complete metamorphosis characterised by four distinct life stages as shown in figure 1.1: egg, larva (caterpillar), pupa (chrysalis), and adult. Each stage plays a fundamental role in the development and survival of the species, and the duration of these phases may vary depending on environmental conditions such as temperature, heat accumulation, humidity and food availability.

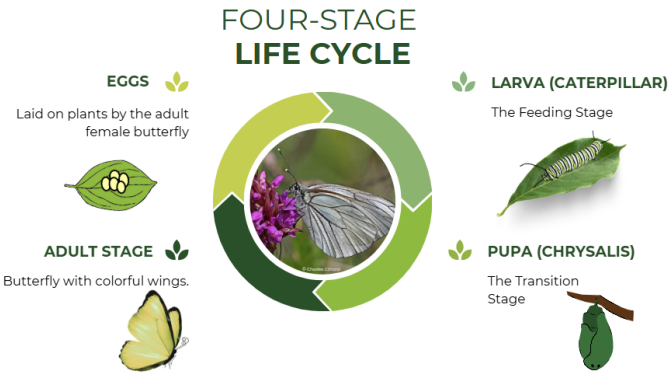


Figure 1.1: Butterfly life cycle.

In this study, the analysis focuses exclusively on adult diurnal butterflies, since the available data are based on field observations of visible adult individuals. Information regarding eggs, larvae and pupae is generally scarce and considerably more difficult to collect because these stages are less detectable in natural environments. Adult butterflies are therefore the most suitable stage for monitoring population abundance, dispersal dynamics and habitat use in urban ecosystems.

This choice of restricting the analysis to adult butterflies is particularly relevant for interpreting population dynamics in a spatially structured context, where movement between suitable but fragmented habitat patches (hereafter, sites) plays a central role. In urban environments, green areas such as parks, gardens and ecological corridors function as discrete habitat sites embedded within a fragmented landscape, connected through butterfly movement across the city. Within this framework, urban butterfly populations can be described using metapopulation theory. A metapopulation consists of a set of spatially separated local populations connected through dispersal and migration processes [7]. While individual local populations may undergo extinction, the metapopulation can persist at a larger

scale through the recolonisation of empty habitat sites by individuals from neighbouring populations [8]. The persistence of the overall system therefore depends not only on the quality of individual sites, but also on their area, the strength and configuration of connectivity between them. The classical metapopulation concept is based on several assumptions:

- Habitat is distributed in discrete patches supporting local populations;
- Local extinctions are recurrent events;
- Different patches show asynchronous dynamics;
- Migration between patches allows recolonisation processes.

In real urban landscapes, however, these conditions are often only partially met, as spatial configuration and anthropogenic structures introduce additional constraints. Some habitat patches may be highly connected, whereas others are partially or strongly isolated due to roads, buildings, or other barriers to movement. As a consequence, butterfly populations may exhibit *partial metapopulation* structures, in which only subsets of patches are effectively linked through dispersal [9, 10]. More generally, metapopulation ecology provides a useful theoretical framework for understanding species persistence in fragmented landscapes, particularly in relation to habitat connectivity, colonisation dynamics and extinction risk [8, 11]. From an eco-biological perspective, studying metapopulations is particularly important because it links patterns of demographic and ecological diversity to processes operating across spatial scales. Indeed, characterizing the diversity of demographic strategies among species can inform research on population trends and stability, trait syndromes, community stability, species coexistence, and ecological succession [12].

Within a metapopulation framework, not all habitat patches contribute equally to long-term population persistence. Urban green areas can differ substantially in habitat quality: some function as productive habitats, whereas others, despite appearing suitable, are unable to sustain stable local populations over time. This outcome is not driven only by environmental conditions, but also by species-specific dispersal abilities, which determine how many individuals reach each patch. Even when habitat quality is sufficient, low immigration can prevent population establishment due to too few arriving individuals to ensure successful reproduction and long-term persistence. This heterogeneity gives rise to the concepts of *source* and *sink* habitat, as well as *ecological traps*. Source–sink dynamics and ecological traps therefore represent a key component in the study of population ecology and conservation within fragmented urban landscapes.
A *source* habitat is a high-quality patch in which local environmental conditions

allow a population to maintain positive growth. In these habitats, reproduction exceeds mortality, generating a surplus of individuals that can disperse to surrounding patches. Source populations therefore act as net exporters of butterflies and play a fundamental role in maintaining metapopulation stability. They are generally characterised by high resource availability, the presence of suitable host plants, and favourable microclimatic conditions. Conversely, a *sink* habitat is a lower-quality patch where local conditions are insufficient to sustain a population in the long term. In these habitats, mortality exceeds reproduction, resulting in a net population decrease. Sink populations therefore persist primarily through continuous immigration from nearby source habitats. Factors such as habitat degradation, limited food availability, pollution, or excessive human disturbance can contribute to the emergence of sink habitats within urban environments. A distinct but related phenomenon is represented by *ecological traps*. An *ecological trap* occurs when individuals are attracted to a habitat that appears suitable based on environmental cues but, in reality, provides poor conditions for survival or reproduction. These traps often arise following rapid environmental change, which alters the cues organisms use to select habitat and leads them to preferentially occupy environments where their fitness is ultimately reduced [13, 14]. In urban landscapes, artificial structures, exotic vegetation, or modified environmental signals may trigger maladaptive behavioural responses that favour the selection of low-quality habitats. Although both traps and sinks reduce population performance, the key difference lies in habitat selection: sink habitats are not necessarily preferred, whereas ecological traps actively attract individuals despite their poor quality [15]. This distinction has important implications for conservation and urban ecological management [13].
Traditional source–sink models that do not explicitly account for habitat selection may overlook the effects of ecological traps, leading to overly optimistic predictions about population persistence. In extreme cases, ecological traps can drive local populations to extinction [15]. Despite their potential ecological importance, robust empirical evidence remains limited, partly because ecological traps are difficult to detect due to incomplete mechanistic understanding, small spatial scales of observation, and a strong taxonomic bias toward birds in the existing literature [13, 14, 15]. Consequently, expanding empirical research and developing robust methodologies are essential to address these knowledge gaps and to explicitly incorporate the relationship between behavioural habitat selection and actual habitat quality into urban conservation planning [13, 14, 15].

In this context, the present thesis aims to investigate both source–sink–trap dynamics and metapopulation structure, providing ecologically relevant insights to support urban biodiversity conservation.

1.2 Mathematical foundations: literature and reference

The study of population dynamics and migration has been extensively investigated across a wide range of species and ecological systems, including butterflies, using diverse mathematical and ecological approaches [7, 16, 17]. The choice of model strongly depends on the biological question under investigation, the spatial and temporal scale considered, and the type of available data. For this reason, the literature presents a variety of complementary frameworks, each designed to capture different aspects of population persistence, dispersal, and habitat interaction.

A central concept in population and landscape ecology is *patch dynamics*, which considers ecological systems as collections of interacting habitat patches. In this context, a *patch* refers to a distinct and relatively homogeneous area of habitat that differs from the surrounding landscape matrix, often as a result of natural or anthropogenic disturbances. Patches vary in size, shape, and structural characteristics, and represent fundamental ecological units for species dispersal, resource use, and the maintenance of biodiversity. Accordingly, the structure, function, and temporal evolution of ecosystems can be studied through the interactions among habitat patches and the movement of organisms between them. Patch dynamics is particularly relevant in fragmented landscapes such as urban environments, where green areas are spatially separated yet remain ecologically connected through dispersal.
One of the most widely used mathematical formulations in this context is the Incidence Function Model (IFM), which describes species occupancy in fragmented habitats as a function of patch size, isolation, and colonisation–extinction processes [7]. A classical theoretical reference is Levins' metapopulation model, which describes the temporal variation in the fraction of occupied habitat patches:

$$\frac{dp}{dt} = cp(1-p) - ep$$

where p represents the proportion of occupied patches, c is the colonisation rate, and e is the extinction rate. Despite its simplicity, this model provides an effective conceptual description of how local extinction and recolonisation can maintain a population over time in fragmented habitats [7].

Beyond compartmental models, other approaches include *agent-based models* and *cellular automata*, which simulate the behaviour of individual organisms moving through space. In these models, each individual can colonise neighbouring habitat cells according to local movement rules, while extinction events occur with a certain probability [7, 18]. These frameworks are particularly useful for explicitly

representing local interactions, stochasticity, and heterogeneous spatial structure. More recently, hybrid approaches combining spatial representations with differential equations have become increasingly relevant. These models integrate continuous-time population dynamics with explicit movement across discrete habitat networks, allowing both local abundance changes and spatial dispersal to be represented simultaneously [19, 20]. Such formulations are particularly suitable for urban ecological systems, where species movement occurs across a fragmented but interconnected network of habitat patches.

A common thread across several of these frameworks is the explicit treatment of stochasticity. *Stochastic differential equations* (SDEs) extend classical deterministic models by incorporating random perturbations directly into population dynamics, thereby capturing environmental variability and demographic noise that deterministic formulations cannot account for. The mathematical theories underlying these stochastic approaches range from standard SDEs to delay differential equations and fractional-order differential equations, and have been systematically reviewed in the context of pollinator population modelling [21]. Complementary to analytical stochastic methods, *Monte Carlo simulations* offer a flexible computational strategy in which population trajectories are generated through repeated random sampling from prescribed probability distributions, enabling the estimation of outcome distributions and uncertainty ranges across a wide parameter space [17]. Together, these stochastic frameworks offer a principled way to quantify variability and assess the robustness of ecological predictions under uncertainty, and prove particularly valuable when the relative contribution of habitat quality, dispersal, and demographic noise to patch occupancy needs to be disentangled [7].

While the previously discussed models provide robust theoretical baselines, none fully captures the specific ecological dynamics to be mapped in this study; therefore, an alternative investigative approach has been adopted.

1.3 The Project

Modelling butterfly dynamics in urban environments is inherently complex and can be approached in multiple ways depending on the ecological processes of interest. As previously noted, an alternative investigative approach has been adopted here to better satisfy the specific biological requirements of the system and achieve the desired mapping of population dynamics. In this work, the selected strategy combines ordinary differential equations with a graph-based spatial representation of the city, allowing local population dynamics to be coupled with movement

between urban green spaces to capture both temporal evolution and spatial connectivity. While this approach is deemed particularly appropriate for addressing the core questions of the present research, it is not advanced as the uniquely correct solution, nor is it presented as inherently superior to alternative methodological frameworks.
In particular, the ordinary differential equations describe the temporal dynamics of adult butterfly populations inferred from real data across the green spaces of Barcelona. The city is represented as a network of habitat patches, where nodes correspond to urban green areas and edges describe potential dispersal pathways between them, defined according to spatial distances, species-specific traits, and characteristics of the urban environment. This graph-based representation naturally reflects the metapopulation perspective introduced above, in which habitat fragmentation and connectivity act as key drivers of population persistence.
Within this modelling framework, abundance estimates are obtained through a neural network prediction pipeline (Section 2.5), which is trained on real butterfly count data collected from field monitoring campaigns. Given the inherent variability and uncertainty associated with observational data, these estimates are directly incorporated as state variables in the mathematical model.

Therefore the main objective of this thesis is to develop a mathematical model capable of both describing and predicting butterfly abundance across the city over time. More specifically, the primary goal of this work is to study the flow and dispersion of butterflies across Barcelona's urban parks in order to identify suitable and unsuitable habitats within the park network. The model therefore aims to:

- reconstruct butterfly population dynamics in Barcelona's urban green spaces;
- describe movement and connectivity between habitat patches;
- identify possible metapopulation structures and dispersal pathways;
- detect potential ecological traps associated with urbanisation.

By combining ecological theory, empirical monitoring data, and mathematical modelling, this project provides a quantitative framework for understanding how urban landscapes shape butterfly populations and for supporting biodiversity-oriented management of urban green spaces.

It must be acknowledged, however, that the proposed model retains an experimental character and relies on a set of approximations (section 2.4), which inevitably introduce a degree of uncertainty into the results. Despite these limitations, the conceptual approach proposed here represents a methodologically innovative contribution to the field, offering a valid basis for further refinements and empirical

validation.

1.3.1 Thesis structure

The remainder of the thesis is organised as follows.

- **Chapter 2** presents the datasets used in this study, including butterfly observations, habitat information, data uncertainty and the neural-network-based abundance estimates adopted as model inputs.
- **Chapter 3** introduces the mathematical framework, describing the population dynamics model, the graph representation of habitat connectivity and the methods used to analyse dispersal and ecological dynamics.
- **Chapter 4** describes the computational implementation of the proposed work and the numerical procedures adopted for parameter estimation and simulation.
- **Chapter 5** presents and discusses the results obtained from the proposed model, with particular attention to its predictive performance and the ecological interpretation of the inferred movement patterns.
- **Chapter 6** summarises the main conclusions of the study, discusses its limitations and outlines possible directions for future research.

Chapter 2

Data

As mentioned in section 1.3, the main objective of this project is to develop a model capable of meaningfully representing real data collected from the Barcelona study sites. However, this information cannot be interpreted directly, as it is affected by methodological constraints, uneven sampling frequency, and substantial heterogeneity in uncertainty. In addition, strong variability in species-specific behaviour further complicates data organisation and classification, making it difficult to capture all dynamics within a single simplified setup. As a result, the mathematical model must be sufficiently flexible to account for a wide range of heterogeneous and partially unpredictable observations.
For this reason, the modelling approach relies exclusively on neural network predictions as state variables, which provide a more consistent and structured representation of the underlying observational data, while also ensuring a complete daily count series without gaps, a property that considerably simplifies the estimation procedure.

In the following chapter, a detailed analysis is presented to systematically explore the available datasets, including daily butterfly counts, species composition and habitat characteristics, in order to better understand the structure and variability of the ecological information used in this study.

2.1 Source of information



Data were collected in collaboration with the University of Barcelona (UB) and CREAF research center.

The Butterflies counts, essential for mapping population dynamics were provided by the urban Butterfly Monitoring Scheme (uBMS), led by Prof. Yolanda Melero (CREAF) and supported by the city council of Barcelona. The uBMS is a citizen science network that relies on volunteers to monitor species abundance to gather long term spatiotemporal information on species presence and abundances to better understand the effect of urbanisation on biodiversity, for which assessing species population dynamics is key[22]. This local initiative feeds into the broader European Butterfly Monitoring Scheme (eBMS). Managed by the UK Centre for Ecology & Hydrology (UKCEH) alongside numerous research institutions, the eBMS integrates datasets from over 25 countries to facilitate large-scale ecological research [23].

This international network originated in the UK in 1970 and expanded across Europe in the 1990s, with Spain launching its first schemes in Catalonia in 1994. Ultimately, the goal of the uBMS is to translate volunteer-collected observations into actionable management strategies that enhance urban biodiversity. The impact of this approach is highly tangible: in 2022, 41 volunteers surveyed 27 parks and gardens in Barcelona (figure 2.1) on a weekly or fortnightly basis, successfully recording over 7,000 butterflies across 33 different species.

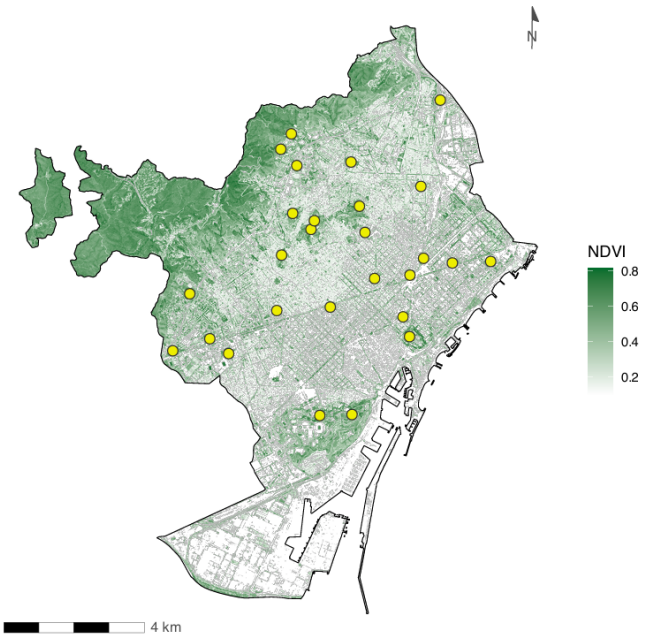


Figure 2.1: Map of Barcelona showing the monitoring sites in yellow and the city's green areas in green.

2.1.1 Counting methodology

To guarantee the reliability and consistency of the generated count data, the project relies on a standardized monitoring protocol across all participating urban areas. Specifically, uBMS volunteers count butterflies weekly or biweekly within specific gardens and parks across Barcelona, Madrid and Sabadell from March to the end of October. The monitoring of each garden begins with a transect, where individuals are counted along a fixed path of 300 meters with a constant width. This is followed by a timed walk, during which butterflies are recorded along an additional route for a duration ranging from 15 to 90 minutes, depending on the size of the park. While the area for this walk is fixed and primarily covers the vegetated sections, the movement within it is random and free. In smaller gardens, observations may even be conducted from a stationary position for the allotted time. To ensure the comparability of results over time, both segments must remain consistent during every visit; however, while the transect follows a strictly fixed path, the timed walk allows for greater freedom of movement within the selected area. To maintain methodological coherence during both phases, the observer walks slowly and counts only those butterflies seen within a distance of 5 meters ahead and 2.5 meters to each side; individuals observed beyond these limits are excluded (visual example in figure 2.2). While it is permissible to pause or momentarily leave the path to identify or photograph a butterfly, the observer must always return to the exact point where the route was interrupted. Surveys should be conducted between 10:00 and 16:00 under favorable weather conditions, as butterflies require solar heat to maintain activity. Optimal conditions include a minimum temperature of 13°C on sunny days or 17°C on cloudy days (with up to 50% cloud cover), while temperatures exceeding 35°C should be avoided. Surveys must not be carried out during rain, in heavy cloud cover, or when wind speeds reach level 5 on the Beaufort scale (approximately > 29 km/h), as butterfly activity decreases significantly under adverse conditions.



Figure 2.2: Visual example of the counting methodology (uBMs).

2.2 Butterfly: observations and categorization

In total, 152 distinct butterfly species are documented within the region of Catalonia [1], showcasing a notable ecological diversity. Each of these species exhibits unique morphological and behavioral traits.
For the purpose of this study, the subsequent sections delineate the most fundamental characteristics required for the analysis, focusing exclusively on the adult life stage and its relevance to population monitoring.

- Voltinism: in the world of Lepidoptera, voltinism refers to the number of generations a butterfly species completes within a single year. This biological rhythm is primarily driven by environmental temperature and the availability of host plants. In this sense butterflies are divided into two macro categories based on their voltinity: multivoltine and univoltine, as shown in the example illustration in figure 2.3. Univoltine species, such as the *Brintesia circe*, produce only one generation per year, a strategy common in cold or high-altitude habitats where the growing season is short. In contrast, multivoltine species like the *Pieris rapae* can complete three or more generations, allowing them to rapidly increase their population during warm seasons. Between these cycles, many butterflies enter diapause, a state of suspended development, to survive unfavorable conditions like winter.

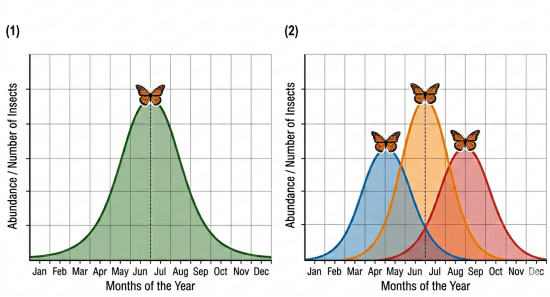


Figure 2.3: Illustration of voltinism: (1) univoltine, (2) multivoltine (AI-generated image).

This rhythm is not always fixed; many species exhibit plasticity, transitioning from univoltine (one generation) in colder northern regions to multivoltine (multiple generations) in warmer southern climates. This flexibility allows them to exploit longer growing seasons. Consequently, multivoltine populations often show distinct seasonal peaks in abundance, with numbers surging during the hot summer months when rapid development is possible.
These seasonal peaks in abundance, driven by the interplay between temperature and generation timing, make voltinism a central feature to capture in the mathematical model developed. Since the number of generations is not fixed but emerges from the cumulative effect of temperature over time, it enters the model as a temperature-driven quantity rather than a fixed parameter.

- Wing Index: the Wing Index (WI) is a morphological trait based on the size and shape of the anterior wing, describing the relationship between body size and wing dimensions in Lepidoptera. It is commonly used in landscape ecology as a proxy for flight efficiency, dispersal capacity, and sensitivity to habitat fragmentation. Species with relatively larger wings compared to body mass generally exhibit higher mobility and greater dispersal ability, whereas species with smaller relative wings tend to be more sedentary. Although WI is often associated with increased mobility, it represents a morphological proxy rather than a direct measure of movement. As such, it does not necessarily reflect realised dispersal, since morphology and behaviour are not always coupled. For this reason, WI is treated as a simplifying indicator of potential mobility and is not included as a direct variable in the quantitative model.

- Generalism: the concept of generalism provides a framework for classifying butterfly species into two broad ecological categories: generalists and specialists. Generalist species, such as *Pieris rapae* or the migratory *Vanessa cardui*, exhibit high adaptability and can easily navigate between fragmented urban parks and gardens to forage. These highly mobile species often maintain larger population counts as they are not constrained to a single location. Conversely, specialist species, including *Charaxes jasius* and *Pararge aegeria*, are predominantly sedentary and strictly dependent on specific host plants or microclimates, such as shaded woodlands. While a generalist can traverse an entire urban matrix to locate resources, a specialist remains tightly bound to its original habitat, rendering it significantly more vulnerable to environmental changes and habitat loss.

  Because this breadth of the ecological niche plays a fundamental role in shaping butterfly population dynamics within urban parks, it is crucial to characterize each species with the highest possible accuracy. Following the methodology established by [1], seven specific ecological traits were selected to quantify this intrinsic generalism. These traits, widely recognized as critical to the behavior and ecology of butterflies and other invertebrates [24, 25, 26, 27], include:

  – species specialisation index (SSI): adult habitat specialisation, i.e. quantified as the coefficient of the variations in the average density in the available habitat; higher the value greater the specialization [28];

  – larval specialisation (LS): from L1 (extreme specialists) to L3 (extreme generalists) [29, 30]:

  – voltinism: number of generations per year categorised as 'Univoltine' (1 generation/year) or 'Multivoltine' (> 1 generations/year) [29, 30];

  – overwintering stage: categorised as 'Egg', 'Larva', 'Pupa', 'Adult', 'Any' (for species

that depending on the year can overwinter in any stage) or 'None' (for migratory specie);

- species temperature index (STI): medium temperature the species live in [31];
- mobility: dispersion capacity, from 'D1' (few dispersion capacity) to 'D4' (high dispersion capacity) [32].
- habitat openness: from −1 (totally forest species) to +1 (cropland species) [33].

By analyzing the co-occurrence of these seven traits, species were separated into distinct functional groups [34, 35]. The clustering analysis encompassed the 152 butterfly species known to inhabit the Mediterranean area of Catalonia, which includes the present study area. This baseline figure excludes only exceptionally rare species, defined as those with fewer than 20 observations over the past 24 years [36]. To define these functional groups, a hierarchical clustering approach was applied based on similarities in species-specific trait values. Specifically, a Hill & Smith principal component analysis was utilized, as it effectively handles multivariate datasets containing both categorical and continuous variables [37]. The resulting dendrogram was used to partition species into clusters, which were then interpreted as functional groups. The clusters derived from this regional baseline were subsequently applied to the butterfly species observed in the sampled urban gardens for the modelling analysis. This procedure resulted in four distinct functional groups, effectively categorising the 152 butterfly species into macro-categories based on shared ecological and behavioural traits. These groups, labelled Cluster 1 through Cluster 4, provide a structured setup for evaluating the diverse responses of butterfly communities within urban habitats. Accordingly, the following sections first describe the ecological characterisation of each macro-category summarised in table 2.1, followed by the assignment of individual species to the corresponding clusters reported in table 2.2.

Species trait	**Cluster1-sedentary** specialists	**Cluster2-sedentary** interm.	**Cluster3-forest** specialists	**Cluster4-mobile** generalists
Spec. index	Specialist	Intermediate	Med. Generalist	Generalist
Larval spec.	Specialist	Intermediate	Intermediate	Generalist
Voltinism	Univoltine	Both	Multivoltine	Multivoltine
Overwintering	Larva, Egg / (Pupa)	Larva / (Pupa, Egg)	Adult, Pupa / (Any, Larva)	None, Pupa / (Larva)
Temp. index	Low	Interm. / High	Intermediate	High
Mobility	Very low	Low	Medium	High
Habitat	Open	Interm. / Closed	Closed	Interm. / Open

Table 2.1: Summary of species trait clusters.

Cluster	Species Name	Voltinism
1	Aphantopus hyperantus	Univoltine
1	Argynnis aglaja	Univoltine
1	Argynnis niobe	Univoltine
1	Boloria eunomia	Univoltine
1	Boloria selene	Univoltine
1	Brenthis daphne	Univoltine
1	Brenthis ino	Univoltine
1	Carcharodus baeticus	Multivoltine
1	Carcharodus floccifera	Univoltine
1	Chazara briseis	Univoltine
1	Coenonympha glycerion	Univoltine
1	Cupido minimus	Univoltine
1	Cupido osiris	Univoltine
1	Cyaniris semiargus	Univoltine
1	Erebia cassioides	Univoltine
1	Erebia epiphron	Univoltine
1	Erebia euryale	Univoltine
1	Erebia meolans	Univoltine
1	Erebia neoridas	Univoltine
1	Erebia oeme	Univoltine
1	Erebia triaria	Univoltine
1	Euphydryas desfontainii	Univoltine
1	Hamearis lucina	Univoltine
1	Hesperia comma	Univoltine
1	Hyponephele lycaon	Univoltine
1	Laeosopis roboris	Univoltine
1	Lycaena alciphron	Univoltine
1	Lycaena tityrus	Multivoltine
1	Lycaena virgaureae	Univoltine
1	Lysandra coridon	Univoltine
1	Melanargia occitanica	Univoltine
1	Melanargia russiae	Univoltine
1	Melitaea cinxia	Univoltine
1	Melitaea diamina	Univoltine
1	Melitaea nevadensis	Univoltine
1	Melitaea parthenoides	Univoltine
1	Melitaea trivia	Multivoltine
1	Muschampia proto	Univoltine
1	Parnassius apollo	Univoltine
1	Parnassius mnemosyne	Univoltine
1	Phengaris arion	Univoltine
1	Plebejus argus	Univoltine
1	Plebejus idas	Univoltine
1	Polyommatus amandus	Univoltine
1	Polyommatus damon	Univoltine
1	Polyommatus daphnis	Univoltine
1	Polyommatus dorylas	Univoltine
1	Polyommatus escheri	Univoltine
1	Polyommatus fulgens	Univoltine
1	Polyommatus ripartii	Univoltine
1	Pyrgus alveus	Univoltine
1	Pyrgus carthami	Univoltine
1	Pyrgus cirsii	Univoltine
1	Pyrgus serratulae	Univoltine
1	Satyrium acaciae	Univoltine
1	Satyrium ilicis	Univoltine
1	Satyrium spini	Univoltine
1	Satyrium w-album	Univoltine
1	Satyrus actaea	Univoltine
1	Scolitantides orion	Univoltine
1	Thecla betulae	Univoltine
1	Thymelicus lineola	Univoltine

Cluster	Species Name	Voltinism
2	Anthocharis euphenoides	Univoltine
2	Apatura ilia	Multivoltine
2	Aporia crataegi	Univoltine
2	Arethusana arethusa	Univoltine
2	Argynnis adippe	Univoltine
2	Argynnis pandora	Univoltine
2	Argynnis paphia	Univoltine
2	Aricia agestis	Multivoltine
2	Boloria dia	Multivoltine
2	Brintesia circe	Univoltine
2	Callophrys avis	Univoltine
2	Callophrys rubi	Univoltine
2	Carcharodus alceae	Multivoltine
2	Carcharodus lavatherae	Univoltine
2	Coenonympha arcania	Univoltine
2	Coenonympha dorus	Univoltine
2	Coenonympha pamphilus	Multivoltine
2	Colias alfacariensis	Multivoltine
2	Cupido alcetas	Multivoltine
2	Cupido argiades	Multivoltine
2	Erynnis tages	Multivoltine
2	Euphydryas aurinia	Univoltine
2	Favonius quercus	Univoltine
2	Glaucopsyche alexis	Univoltine
2	Glaucopsyche melanops	Univoltine
2	Hipparchia fagi	Univoltine
2	Hipparchia fidia	Univoltine
2	Hipparchia hermione	Univoltine
2	Hipparchia semele	Univoltine
2	Hipparchia statilinus	Univoltine
2	Iolana debilitata	Univoltine
2	Lasiommata maera	Multivoltine
2	Limenitis camilla	Multivoltine
2	Limenitis reducta	Multivoltine
2	Lycaena phlaeas	Multivoltine
2	Lysandra bellargus	Multivoltine
2	Lysandra hispana	Multivoltine
2	Maniola jurtina	Univoltine
2	Melanargia lachesis	Univoltine
2	Melitaea deione	Multivoltine
2	Melitaea didyma	Multivoltine
2	Melitaea phoebe	Multivoltine
2	Ochlodes sylvanus	Multivoltine
2	Pieris mannii	Multivoltine
2	Polyommatus icarus	Multivoltine
2	Polyommatus thersites	Multivoltine
2	Pseudophilotes panoptes	Univoltine
2	Pyrgus armoricanus	Multivoltine
2	Pyrgus malvoides	Multivoltine
2	Pyronia bathseba	Univoltine
2	Pyronia cecilia	Univoltine
2	Pyronia tithonus	Univoltine
2	Satyrium esculi	Univoltine
2	Spialia sertorius	Multivoltine
2	Thymelicus acteon	Univoltine
2	Thymelicus sylvestris	Univoltine
2	Tomares ballus	Univoltine
2	Zerynthia rumina	Univoltine

Cluster	Species Name	Voltinism
3	Aglais io	Multivoltine
3	Aglais urticae	Multivoltine
3	Anthocharis cardamines	Univoltine
3	Araschnia levana	Multivoltine
3	Celastrina argiolus	Multivoltine
3	Charaxes jasius	Multivoltine
3	Euchloe crameri	Multivoltine
3	Gonepteryx cleopatra	Multivoltine
3	Gonepteryx rhamni	Univoltine
3	Iphiclides feisthamelii	Multivoltine
3	Issoria lathonia	Multivoltine
3	Lasiommata megera	Multivoltine
3	Leptidea sinapis	Multivoltine
3	Libythea celtis	Univoltine
3	Nymphalis antiopa	Univoltine
3	Nymphalis polychloros	Univoltine
3	Pararge aegeria	Multivoltine
3	Pieris napi	Multivoltine
3	Polygonia c-album	Multivoltine
3	Vanessa cardui	Multivoltine

Cluster	Species Name	Voltinism
4	Aricia cramera	Multivoltine
4	Cacyr eus marshalli	Multivoltine
4	Colias crocea	Multivoltine
4	Danaus chrysippus	Multivoltine
4	Gegenes nostradamus	Multivoltine
4	Lampides boeticus	Multivoltine
4	Leptotes pirithous	Multivoltine
4	Papilio machaon	Multivoltine
4	Pieris brassicae	Multivoltine
4	Pieris rapae	Multivoltine
4	Pontia daplidice	Multivoltine
4	Vanessa atalanta	Multivoltine

Table 2.2: 152 Butterfly species divided by cluster [1].

Among the 152 butterfly species documented across Catalonia, a total of 44 distinct species were identified across the Barcelona parks during the study period from 2018 to 2024. The data presented in this section constitute raw counts collected by volunteers through the uBMS monitoring protocol described in section 2.1.1, and should be understood as a preliminary descriptive analysis aimed at characterising the structure of the dataset before any mathematical modelling is applied. Each species is reported below together with its voltinism and functional cluster classification:

- *Anthocharis cardamines* (Uni, Cl. 3)
- *Apatura ilia* (Multi, Cl. 2)
- *Aricia cramera* (Multi, Cl. 4)
- *Brintesia circe* (Uni, Cl. 2)
- *Cacyreus marshalli* (Multi, Cl. 4)
- *Carcharodus alceae* (Multi, Cl. 2)
- *Celastrina argiolus* (Multi, Cl. 3)
- *Charaxes jasius* (Multi, Cl. 3)
- *Colias crocea* (Multi, Cl. 4)
- *Danaus chrysippus* (Multi, Cl. 4)
- *Euchloe crameri* (Multi, Cl. 3)
- *Euphydryas aurinia* (Uni, Cl. 2)
- *Favonius quercus* (Uni, Cl. 2)
- *Gonepteryx cleopatra* (Multi, Cl. 3)
- *Gonepteryx rhamni* (Uni, Cl. 3)
- *Hipparchia fidia* (Uni, Cl. 2)
- *Hipparchia semele* (Uni, Cl. 2)
- *Hipparchia statilinus* (Uni, Cl. 2)
- *Iphiclides feisthamelii* (Multi, Cl. 3)
- *Lampides boeticus* (Multi, Cl. 4)
- *Lasiommata megera* (Multi, Cl. 3)
- *Leptotes pirithous* (Multi, Cl. 4)
- *Libythea celtis* (Uni, Cl. 3)
- *Limenitis reducta* (Multi, Cl. 2)
- *Lycaena phlaeas* (Multi, Cl. 2)
- *Maniola jurtina* (Uni, Cl. 2)
- *Melanargia lachesis* (Uni, Cl. 2)
- *Melitaea didyma* (Multi, Cl. 2)
- *Ochlodes sylvanus* (Multi, Cl. 2)
- *Papilio machaon* (Multi, Cl. 4)
- *Pararge aegeria* (Multi, Cl. 3)
- *Pieris brassicae* (Multi, Cl. 4)
- *Pieris napi* (Multi, Cl. 3)
- *Pieris rapae* (Multi, Cl. 4)
- *Polygonia c-album* (Multi, Cl. 3)
- *Polyommatus icarus* (Multi, Cl. 2)
- *Pontia daplidice* (Multi, Cl. 4)
- *Pyrgus malvoides* (Multi, Cl. 2)
- *Pyronia bathseba* (Uni, Cl. 2)
- *Pyronia cecilia* (Uni, Cl. 2)
- *Satyrium esculi* (Uni, Cl. 2)
- *Thymelicus acteon* (Uni, Cl. 2)
- *Vanessa atalanta* (Multi, Cl. 4)
- *Vanessa cardui* (Multi, Cl. 3)

Of these, approximately 65.9% are multivoltine and 34.1% univoltine, reflecting the predominance of adaptable species with rapid reproductive cycles in the urban environments analysed, as summarised in the following table.

Voltinism	**Counts**	**Percentage**
Multivoltine	29	∼65.9%
Univoltine	15	∼34.1%
Total	**44**	**100%**

This distribution already suggests a first structural feature of the dataset: the community recorded in the Barcelona parks is dominated by ecologically flexible, high-mobility species. Following this observation, the presence frequency of each species across the 24 monitored transects was examined. Only five species were recorded across all 24 parks:

- *Cacyreus marshalli* (Multivoltine, Cluster 4)
- *Pararge aegeria* (Multivoltine, Cluster 3)
- *Pieris brassicae* (Multivoltine, Cluster 4)
- *Pieris rapae* (Multivoltine, Cluster 4)
- *Vanessa atalanta* (Multivoltine, Cluster 4)

Notably, all five belong to the multivoltine category and to the clusters of highest mobility and generalism (Clusters 3 and 4), confirming that urban ubiquity is strongly associated with high dispersal capacity and broad ecological plasticity.

To further explore this pattern, the complete presence frequency of each species across the 24 monitored transects is reported in the table 2.3 below, ordered from most to least widespread. This distribution is particularly relevant in view of the mathematical modelling developed in the following chapters: the varying presence counts across species and parks provide a basis for assessing to what extent the model can be considered representative, and for which species or ecological profiles its application is most appropriate.

Species	**Presence**
Cacyreus marshalli (Multivoltine, Cluster 4)	24 / 24
Pararge aegeria (Multivoltine, Cluster 3)	24 / 24
Pieris brassicae (Multivoltine, Cluster 4)	24 / 24
Pieris rapae (Multivoltine, Cluster 4)	24 / 24
Vanessa atalanta (Multivoltine, Cluster 4)	24 / 24
Lampides boeticus (Multivoltine, Cluster 4)	23 / 24
Leptotes pirithous (Multivoltine, Cluster 4)	23 / 24
Lasiommata megera (Multivoltine, Cluster 3)	22 / 24
Polyommatus icarus (Multivoltine, Cluster 2)	22 / 24
Iphiclides feisthamelii (Multivoltine, Cluster 3)	21 / 24
Vanessa cardui (Multivoltine, Cluster 3)	21 / 24
Colias crocea (Multivoltine, Cluster 4)	20 / 24
Gonepteryx cleopatra (Multivoltine, Cluster 3)	20 / 24
Papilio machaon (Multivoltine, Cluster 4)	19 / 24
Celastrina argiolus (Multivoltine, Cluster 3)	18 / 24
Aricia cramera (Multivoltine, Cluster 4)	16 / 24
Charaxes jasius (Multivoltine, Cluster 3)	15 / 24
Gonepteryx rhamni (Univoltine, Cluster 3)	13 / 24
Pontia daplidice (Multivoltine, Cluster 4)	13 / 24
Carcharodus alceae (Multivoltine, Cluster 2)	12 / 24
Lycaena phlaeas (Multivoltine, Cluster 2)	12 / 24
Brintesia circe (Univoltine, Cluster 2)	11 / 24
Euchloe crameri (Multivoltine, Cluster 3)	10 / 24
Libythea celtis (Univoltine, Cluster 3)	9 / 24
Maniola jurtina (Univoltine, Cluster 2)	7 / 24
Pyronia bathseba (Univoltine, Cluster 2)	7 / 24
Satyrium esculi (Univoltine, Cluster 2)	7 / 24
Pyronia cecilia (Univoltine, Cluster 2)	6 / 24
Thymelicus acteon (Univoltine, Cluster 2)	6 / 24
Hipparchia fidia (Univoltine, Cluster 2)	4 / 24
Melanargia lachesis (Univoltine, Cluster 2)	4 / 24
Apatura ilia (Multivoltine, Cluster 2)	3 / 24
Danaus chrysippus (Multivoltine, Cluster 4)	3 / 24
Favonius quercus (Univoltine, Cluster 2)	3 / 24
Limenitis reducta (Multivoltine, Cluster 2)	3 / 24
Pieris napi (Multivoltine, Cluster 3)	3 / 24
Anthocharis cardamines (Univoltine, Cluster 3)	2 / 24
Hipparchia semele (Univoltine, Cluster 2)	2 / 24
Hipparchia statilinus (Univoltine, Cluster 2)	2 / 24
Polygonia c-album (Multivoltine, Cluster 3)	2 / 24
Euphydryas aurinia (Univoltine, Cluster 2)	1 / 24
Melitaea didyma (Multivoltine, Cluster 2)	1 / 24
Ochlodes sylvanus (Multivoltine, Cluster 2)	1 / 24
Pyrgus malvoides (Multivoltine, Cluster 2)	1 / 24

Table 2.3: Presence of the different 44 species of the 24 parks of Barcelona

The raw count data thus show a coherent pattern, where the most widespread species consistently belong to the higher mobility and generalism clusters.

To illustrate the count data, the specie *Pieris rapae* is examined here as a representative case study, embodying the ecological profile that characterises the dominant component of the observed community (figure 2.4).



Figure 2.4: *Pieris rapae*. Photo by Jordi Sánchez.

Classified within Cluster 4 (mobile generalists), this species displays all the traits that favour its success in fragmented urban landscapes: high generalism at both the adult (SSI = 0.707) and larval level (LS3), a multivoltine reproductive strategy, preference for open habitats (openness = 0.28), and maximum dispersal capacity (D4). Its species temperature index (STI = 9.628) further confirms its affinity with warm thermal conditions, consistent with the phenology visible in the raw count data collected across the parks, where abundance peaks concentrate during the summer months [1]. In the following, figure 2.7 displays the raw counts of *Pieris rapae* recorded by volunteers across the 24 Barcelona parks over the period 2018 – 2024. A single-year breakdown is also shown in figure 2.5 to better illustrate the temporal pattern of observations within a single year, whereas figure 2.6 presents the corresponding plot aggregated across all years..

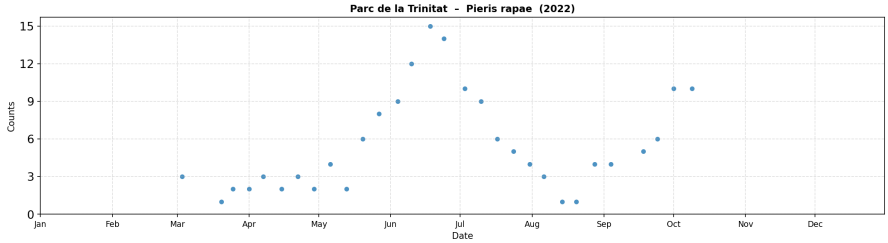


Figure 2.5: Scatter plot of Pieris rapae counts in 2022 in Parc de la Trinitat.

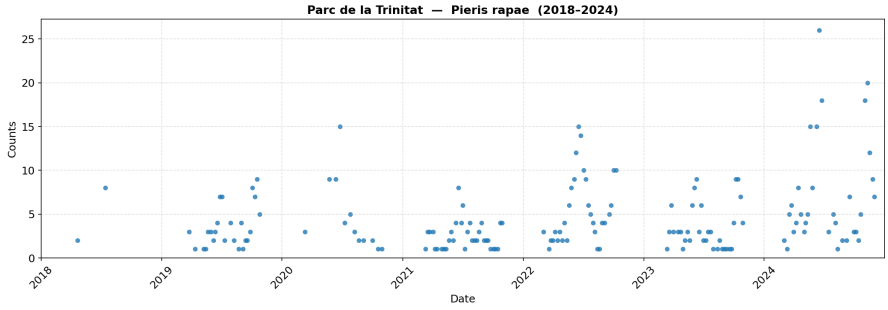


Figure 2.6: Scatter plot of Pieris rapae counts in Parc de la Trinitat over all the years studied.

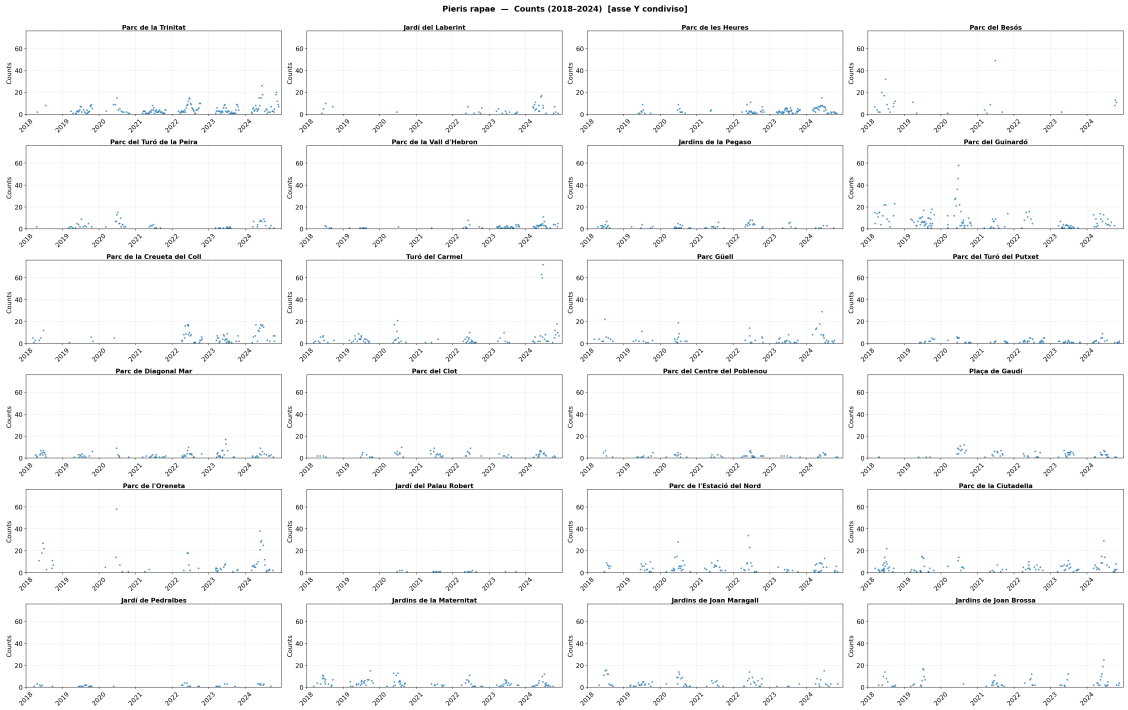


Figure 2.7: Scatter plot of Pieris rapae counts 2023.

2.3 Space level: patches

As introduced in the previous sections, the concept of *patch* refers here to a distinct habitat area embedded within the surrounding matrix. In the context of this study, patches correspond to urban green areas that, despite being immersed in a predominantly built-up environment, can provide resources, shelter and movement corridors for butterfly communities.

The study focuses on the city of Barcelona and its immediate metropolitan conurbation, covering a total area of approximately 12,600 ha. This landscape includes urban, natural and semi-natural environments. About 6,600 ha, corresponding to 52.4% of the study area, are occupied by a dense urban matrix. The remaining area is divided between the south-eastern sector of the Collserola Natural Park, which extends over 3,500 ha (27.8%) and is recognised for its role as a species source area, and approximately 2,500 ha (19.8%) of built-up areas belonging to neighbouring cities and suburban zones [1].
This heterogeneous spatial configuration is particularly relevant from an ecological

perspective. The urban landscape is largely dominated by built-up surfaces, which tend to reduce both the availability and connectivity of suitable habitat patches. Compared with matrices composed of natural or semi-natural land, urban matrices exert a stronger negative influence on species distribution and abundance. As a consequence, urban green areas, which represent the potentially suitable patches for butterflies, are often relatively small, spatially isolated, and strongly influenced by the composition and extent of the surrounding built environment.
The spatial delineation of the study patches was carried out by digitizing their boundaries on up-to-date (2018) orthoimages generated by the Cartographic Institute of Catalonia. Land cover composition within each garden was subsequently inferred from a land-cover map with a 5-metre cell resolution, obtained through the combination and reclassification of three complementary data sources: the Copernicus Urban Atlas, the land cover map of Catalonia (MCSC 2015), and the Spanish SIOSE Land Cover Map [1]. In parallel with this spatial characterisation, a filtering procedure was applied to the butterfly occurrence data to ensure adequate and consistent sampling effort across sites. Gardens with insufficient survey frequency were excluded, retaining only those monitored weekly, biweekly in two non-consecutive weeks, or at least once per month during at least one year.
This selection resulted in a final dataset of 24 gardens, covering 318 ha out of the 335 ha of public and private gardens originally included in the monitoring program, with garden sizes ranging from 0.49 ha to 68.69 ha. From the full dataset of 54 parks, including all monitored sites in Barcelona (29) and Madrid (25), with observations collected between 2018 and 2024, only the 24 Barcelona parks meeting both the sampling requirements and the completeness criteria across all model variables were therefore retained for the subsequent mathematical and computational analyses [1].

Therefore of the 29 Barcelona parks constituting the complete set of patches considered in the spatial mapping model, namely Parc del Guinardó, Parc del Carmel, Parc del Turó del Putxet, Parc de la Trinitat, Jardins de Joan Maragall, Parc de Diagonal Mar, Parc de la Ciutadella, Parc del Castell de l'Oreneta, Jardins del Doctor Pla i Armengol, Parc de la Creueta del Coll, Parc Güell, Parc de les Glòries, Parc del Turó de la Peira, Parc de Cervantes, Jardins de Joan Brossa, Jardins del Palau de les Heures, Jardins de la Maternitat, Parc de l'Estació del Nord, Parc del Laberint d'Horta, Plaça de Gaudí, Parc del Centre del Poblenou, Parc Fluvial del Besòs, Parc de la Pegaso, Parc del Clot, Parc de les Aigües, Jardins del Palau de Pedralbes, Casal de Gent Gran Vall d'Hebron, Turó Parc, and Palau Robert, only 24 were retained for the mathematical and computational analyses, specifically those for which complete data were available across all variables required by the model as shown in figure 2.8:

1. Parc de la Trinitat
2. Jardí del Laberint
3. Parc de les Heures
4. Parc del Besós
5. Parc del Turó de la Peira
6. Parc de la Vall d'Hebron
7. Jardins de la Pegaso
8. Parc del Guinardó
9. Parc de la Creueta del Coll
10. Turó del Carmel
11. Parc Güell
12. Parc del Turó del Putxet
13. Parc de Diagonal Mar
14. Parc del Clot
15. Parc del Centre del Poblenou
16. Plaça de Gaudí
17. Parc de l'Oreneta
18. Jardí del Palau Robert
19. Parc de l'Estació del Nord
20. Parc de la Ciutadella
21. Jardí de Pedralbes
22. Jardins de la Maternitat
23. Jardins de Joan Maragall
24. Jardins de Joan Brossa

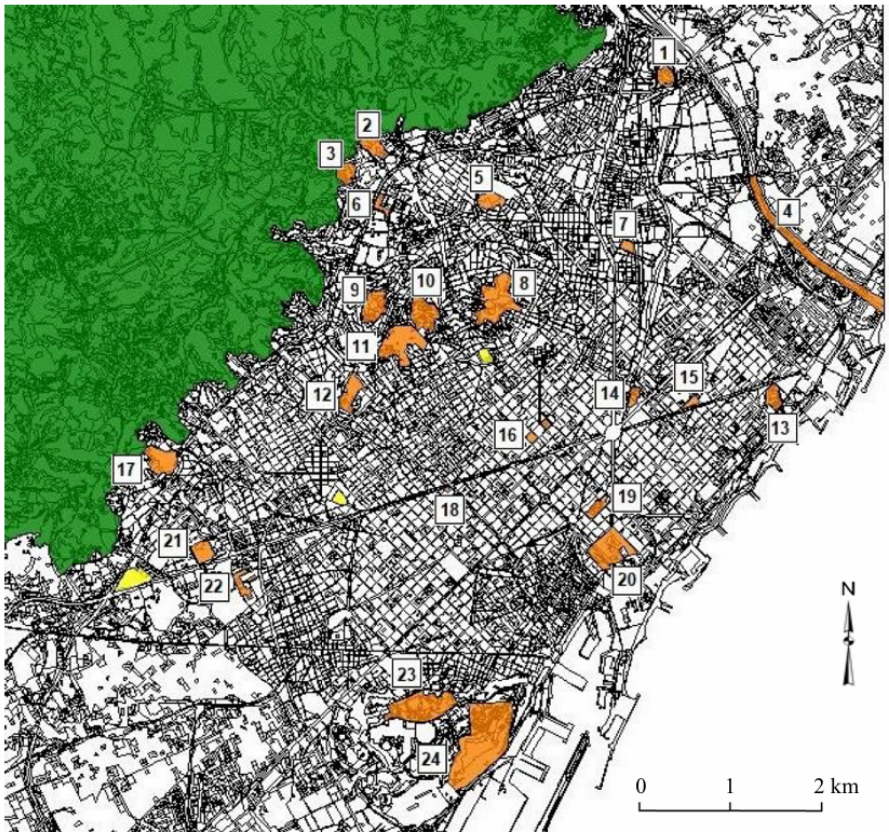


Figure 2.8: Map of the 24 Barcelona parks considered [1].

For each of the 24 patches selected for mathematical and computational analyses, the relevant characteristics required by the model are reported below.

- **Distance between parks**:
  - The Euclidean distances between transects were calculated starting from the centroid coordinates expressed in the projected reference system [38].

    Since these are Cartesian coordinates in meters, it was possible to directly apply the distance formula in the plane, which measures the length of the segment between two points in two-dimensional space.
    In particular, for each pair of transects i and j, the distance was obtained as the square root of the sum of the squares of the differences between the respective coordinates along the x (projected longitude) and y (projected latitude) axes; the results are visible in the table 2.4. This approach allows for distances expressed in meters, consistent with the metric of the reference system used, and is appropriate as it avoids the distortions typical of non-projected geographical systems.

Park	Bes	Bro	Car	Ciu	Clo	Col	Dia	Gau	Gue	Gui	Heb	Heu	Lab	Mar	Mat	Nor	Pob	Ore	Ped	Peg	Pei	Put	Rob	Tri
Parc del Besós	0	8085	5953	5090	3363	6598	2062	4833	6204	4764	6543	7096	6867	8515	9523	4806	2772	9930	9824	2955	5021	7112	6337	3649
Jardins de Joan Brossa	8085	0	5849	3000	5076	6151	6086	4154	5570	6065	7490	8048	8386	938	4014	3380	5370	5922	4715	7006	7399	5119	3338	9495
Turó del Carmel	5953	5849	0	4285	3266	681	5182	2363	387	1204	1660	2276	2548	5551	4548	3725	4083	4181	4534	3127	1913	1440	2512	4881
Parc Ciutadella	5090	3000	4285	0	2208	4846	3124	1948	4183	3929	5875	6528	6702	3461	5247	600	2377	6474	5776	4273	5238	4347	2458	6749
Parc del Clot	3363	5076	3266	2208	0	3943	1976	1488	3371	2375	4516	5178	5200	5333	6174	1707	833	6763	6510	2086	3454	4055	2980	4543
P. Creueta del Coll	6598	6151	681	4846	3943	0	5864	2966	663	1834	1406	1901	2287	5758	4364	4307	4762	3714	4248	3726	2218	1249	2847	5277
P. Diagonal Mar	2062	6086	5182	3124	1976	5864	0	3392	5322	4156	6250	6890	6816	6585	7972	2950	1149	8713	8370	2992	4948	6030	4800	4854
Plaça de Gaudí	4833	4154	2363	1948	1488	2966	3392	0	2313	2029	3928	4583	4755	4199	4692	1369	2254	5325	5022	2927	3371	2737	1507	5380
Parc Güell	6204	5570	387	4183	3371	663	5322	2313	0	1511	1920	2491	2817	5233	4161	3646	4200	3842	4153	3435	2300	1062	2240	5258
Parc del Guinardó	4764	6065	1204	3929	2375	1834	4156	2029	1511	0	2141	2804	2853	5948	5499	3331	3128	5352	5580	1927	1348	2541	2879	3881
P. Vall d'Hebron	6543	7490	1660	5875	4516	1406	6250	3928	1920	2141	0	667	897	7142	5700	5295	5260	4757	5510	3591	1556	2643	4156	4463
P. de les Heures	7096	8048	2276	6528	5178	1901	6890	4583	2491	2804	667	0	540	7656	5999	5952	5915	4841	5736	4162	2075	3068	4729	4756
J. del Laberint	6867	8386	2548	6702	5200	2287	6816	4755	2817	2853	897	540	0	8036	6496	6114	5896	5379	6253	3979	1888	3504	5051	4325
J. Joan Maragall	8515	938	5551	3461	5333	5758	6585	4199	5233	5948	7142	7656	8036	0	3124	3714	5751	5091	3832	7121	7243	4630	3089	9577
J. la Maternitat	9523	4014	4548	5247	6174	4364	7972	4692	4161	5499	5700	5999	6496	3124	0	5129	6886	2064	709	7255	6460	3140	3195	9372
P. Estació del Nord	4806	3380	3725	600	1707	4307	2950	1369	3646	3331	5295	5952	6114	3714	5129	0	2038	6187	5605	3730	4639	3897	2127	6218
P. Centre Poblenou	2772	5370	4083	2377	833	4762	1149	2254	4200	3128	5260	5915	5896	5751	6886	2038	0	7568	7258	2377	4081	4883	3696	4648
Parc de l'Oreneta	9930	5922	4181	6474	6763	3714	8713	5325	3842	5352	4757	4841	5379	5091	2064	6187	7568	0	1405	7270	5927	2835	4061	8991
Jardí de Pedralbes	9824	4715	4534	5776	6510	4248	8370	5022	4153	5580	5510	5736	6253	3832	709	5605	7258	1405	0	7412	6422	3095	3571	9410
J. de la Pegaso	2955	7006	3127	4273	2086	3726	2992	2927	3435	1927	3591	4162	3979	7121	7255	3730	2377	7270	7412	0	2098	4444	4290	2494
P. Turó de la Peira	5021	7399	1913	5238	3454	2218	4948	3371	2300	1348	1556	2075	1888	7243	6460	4639	4081	5927	6422	2098	0	3335	4155	3069
P. Turó del Putxet	7112	5119	1440	4347	4055	1249	6030	2737	1062	2541	2643	3068	3504	4630	3140	3897	4883	2835	3095	4444	3335	0	2001	6320
Jardí Palau Robert	6337	3338	2512	2458	2980	2847	4800	1507	2240	2879	4156	4729	5051	3089	3195	2127	3696	4061	3571	4290	4155	2001	0	6640
P. de la Trinitat	3649	9495	4881	6749	4543	5277	4854	5380	5258	3881	4463	4756	4325	9577	9372	6218	4648	8991	9410	2494	3069	6320	6640	0

Table 2.4: Complete matrix of Euclidean distances calculated (m) between the 24 parks.

  - Impedance: a distance weighted on urbanisation
    $\Longrightarrow$included calculated in the connectivity term 2.2.

- **Surface**: total area of patches in hectare (ha). Total garden area, including vegetated and non-vegetated areas. Large habitat patches are generally more favourable because they reduce the risk of local butterfly extinction, as they provide greater availability of resources such as food and space [1].

- **Quality**: habitat quality was assessed through a proxy based on the similarity of each garden's vegetation to the semi-natural areas where butterfly species richness is greatest.
  Given that previous research has highlighted the importance of native flora for butterfly communities [39], habitat quality was computed for each garden as the area-weighted mean across four vegetation categories, defined as the sum of each category score multiplied by its corresponding area, divided by the total garden area.

$$Q_x = \frac{\sum_j (v_{x,j} \cdot S_{x,j})}{S_x} \tag{2.1}$$

  - Q_x = quality of patch $y \in [0,\ 3]$;
  - $v_{x,j}$ = vegetation category $j \in \{0,\ 1,\ 2,\ 3\}$:
    vegetation per garden was surveyed in situ and categorized into values ranging from zero to three according to the following criteria
    - 0 = no vegetation cover
    - 1 = non-native spontaneous and ornamental vegetation
    - 2 = isolated native plants embeddedwithin non-native vegetation
    - 3 = continuous native vegetation
  - $S_{x,j}$ = surface (area) of category j for all patch x;
  - S_x = surface (area) of patch x.

  This formula was derived from [1] as well as the data related shown in the table 2.5. Although the habitat quality data employed in this study do not represent the most recent available estimates, they constitute the only quantitative assessment currently applicable to the parks under consideration.
  These values have therefore been assumed to remain valid for the purposes of the present analysis, while an updated version is currently being developed by the biological research team involved in the project.

Park Name	Habitat quality	Total area (ha)
Parc de la Trinitat	1.79	5.27
Jardí del Laberint	2.35	8.52
Parc de les Heures	2.52	6.71
Parc del Besós	1.94	33.76
Parc del Turó de la Peira	2.06	7.25
Parc de la Vall d'Hebron	1.63	2.86
Jardins de la Pegaso	1.83	2.83
Parc del Guinardó	2.62	26.16
Parc de la Creueta del Coll	2.26	12.39
Turó del Carmel	2.55	14.54
Parc Güell	2.27	22.60
Parc del Turó del Putxet	2.12	10.06
Parc de Diagonal Mar	1.45	6.04
Parc del Clot	1.57	3.51
Parc del Centre del Poblenou	1.50	2.08
Plaça de Gaudí	1.94	1.35
Parc de l'Oreneta	2.83	13.83
Jardí del Palau Robert	1.70	0.49
Parc de l'Estació del Nord	1.82	4.58
Parc de la Ciutadella	1.48	21.84
Jardí de Pedralbes	1.57	7.14
Jardins de la Maternitat	1.64	5.52
Jardins de Joan Maragall	1.58	30.11
Jardins de Joan Brossa	2.05	68.72

Table 2.5: Habitat quality and total Area of the 24 Barcelona parks [1].

- **Connectivity**: connectivity was estimated for each study garden with respect to all vegetated areas located outside its boundaries, encompassing both study and non-study gardens as well as adjacent natural areas within the city.

  The analysis was based on a 5-metre resolution land cover grid, in which each cell was classified as either vegetated, belonging to open or closed vegetation categories, or non-vegetated. Connectivity was then computed by means of a modified version of the Hanski index, which quantifies the total surrounding vegetated area weighted by the cost distance from the edge of each study garden [1, 40]:

$$C_x = \sum_{y \neq x} S_y \cdot e^{(-\upsilon \cdot \Upsilon_{xy})} \qquad \text{[hanski index]} \tag{2.2}$$

- C_x = connectivity patch x;
- S_y = is each vegetated cell y outside the garden x in area units (0.0025ha);
- Υ_{xy} = "urbanised distance".
  The cost distance from the edge of the garden x to the centre of each vegetated cell y: the cost distance d_{ij} was calculated using the euclidean distance (in m) to each vegetated cell j, weighted by a resistance value assigned to each land cover category (visible in the figure 2.9) based on previous studies [1, 30, 41, 42]:
  * 1: Land cover, Grasslands, Crops, Ornamental vegetated areas, Shrubs, Forest, Urban trees;
  * 5: Beach, Clear areas:
  * 10: Roads,City highways,Buildings,High buildings.

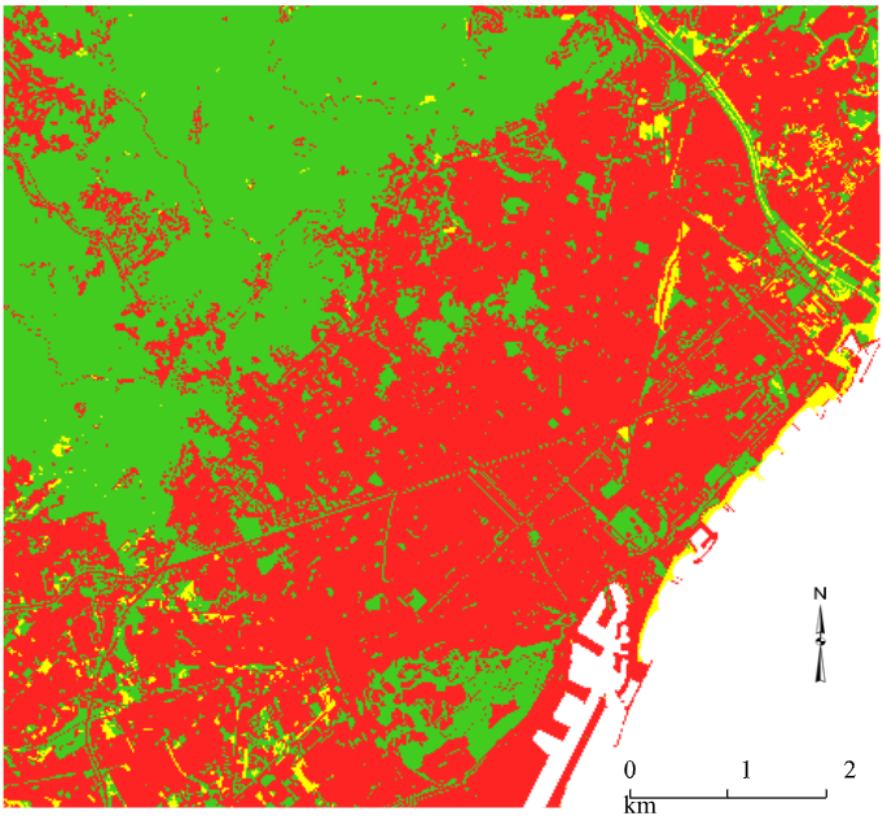


Figure 2.9: Barcelona land cover map used for urbanised distance computation [1]. Green: vegetated areas (low impedance = 1); yellow: open/cleared surfaces (medium impedance = 5); red: built surfaces and roads (high impedance = 10).

– υ the slope of the exponential function of C_x^s to generate a value near zero (0.001) in the exponential function of the model using three maximum dispersal distances (200, 500 and 1000 m).

Connectivity was calculated in 3 different ways because the park is physically one, but it is "perceived" in three different ways depending on who is looking at it.

In agreement with the reference biologists it was decided to consider connectivity C500 connectivity at 500 m buffer because it is quite representative for any species considered.

Below table 2.6 gives the relevant numerical references

Park name	**Connectivity C500**
Parc de la Trinitat	14.03
Jardí del Laberint	140.84
Parc de les Heures	129.23
Parc del Besós	34.15
Parc del Turó de la Peira	2.04
Parc de la Vall d'Hebron	29.22
Jardins de la Pegaso	0.53
Parc del Guinardó	18.12
Parc de la Creueta del Coll	3.50
Turó del Carmel	24.69
Parc Güell	19.64
Parc del Turó del Putxet	3.93
Parc de Diagonal Mar	13.95
Parc del Clot	2.65
Parc del Centre del Poblenou	8.02
Plaça de Gaudí	0.60
Parc de l'Oreneta	0.40
Jardí del Palau Robert	0.48
Parc de l'Estació del Nord	4.43
Parc de la Ciutadella	1.03
Jardí de Pedralbes	4.94
Jardins de la Maternitat	19.96
Jardins de Joan Maragall	41.61
Jardins de Joan Brossa	6.54

Table 2.6: C500 connectivity data per park [1].

- **Urbanisation**

Urbanisation represents a key factor in understanding butterfly spatial dynamics, and is explicitly accounted for in the connectivity term of the model through a cost distance weighted by the degree of built-up cover.
The urbanisation variable is quantified via the Built-Up Fraction (BUFRAC), defined as the proportion of each raster cell covered by built-up surfaces:

$$U = \text{BUFRAC} = \frac{\text{Total area occupied by buildings}}{\text{Total area of the cell}}$$

Barcelona data (figure 2.10) were extracted from the Global Human Settlement Layer (GHSL)[43], which maps built-up surfaces worldwide at a resolution of 100 × 100 m, expressing coverage in square metres of total built-up area. For the temporal anchor point of 2018, data are available at 10 m resolution derived from Sentinel-2 imagery. The total built-up surface was computed across multiple spatial scales around the centroid of each transect: 100 × 100 m, 500 × 500 m, 1000 × 1000 m, 2000 × 2000 m.

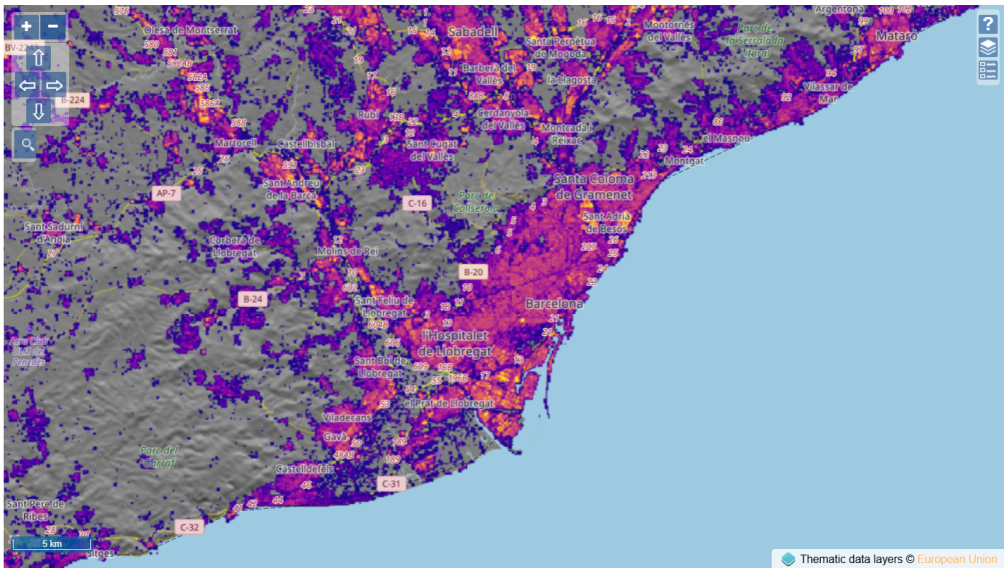


Figure 2.10: Urbanisation map of Barcelona [2].

For the purposes of this study, urbanisation is incorporated into the connectivity term of the model as a distance-weighting factor, computed at fine spatial resolution using Catalonia-specific land cover data [44].

Plots 2.11, 2.12 and 2.13 show a distinct, albeit varied, increase in urbanization across all 24 Barcelona parks from 2018 to 2024, demonstrating the urgency of the analysis being carried out in light of what has already been discussed.

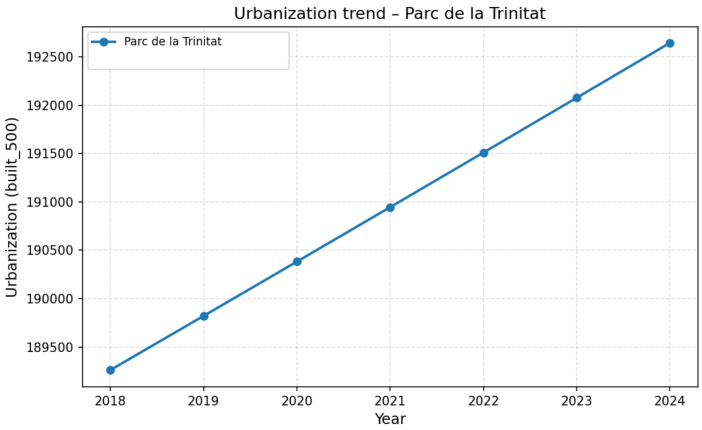


Figure 2.11: Urbanisation evolution in Parc de la Trinitat from 2018 to 2024.

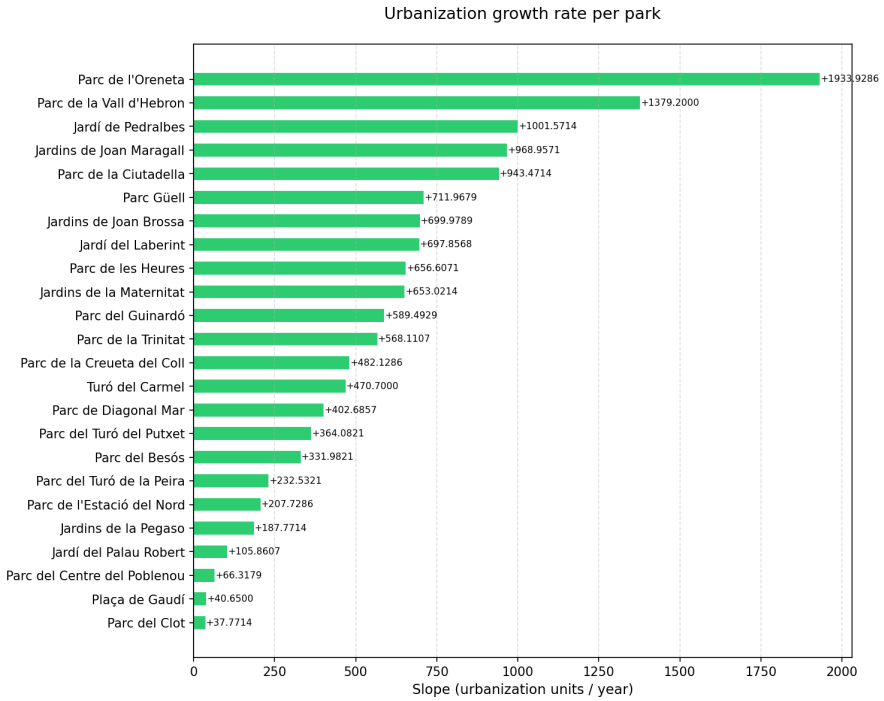


Figure 2.12: The increase in urbanization in parks through the growth rate of urbanization.

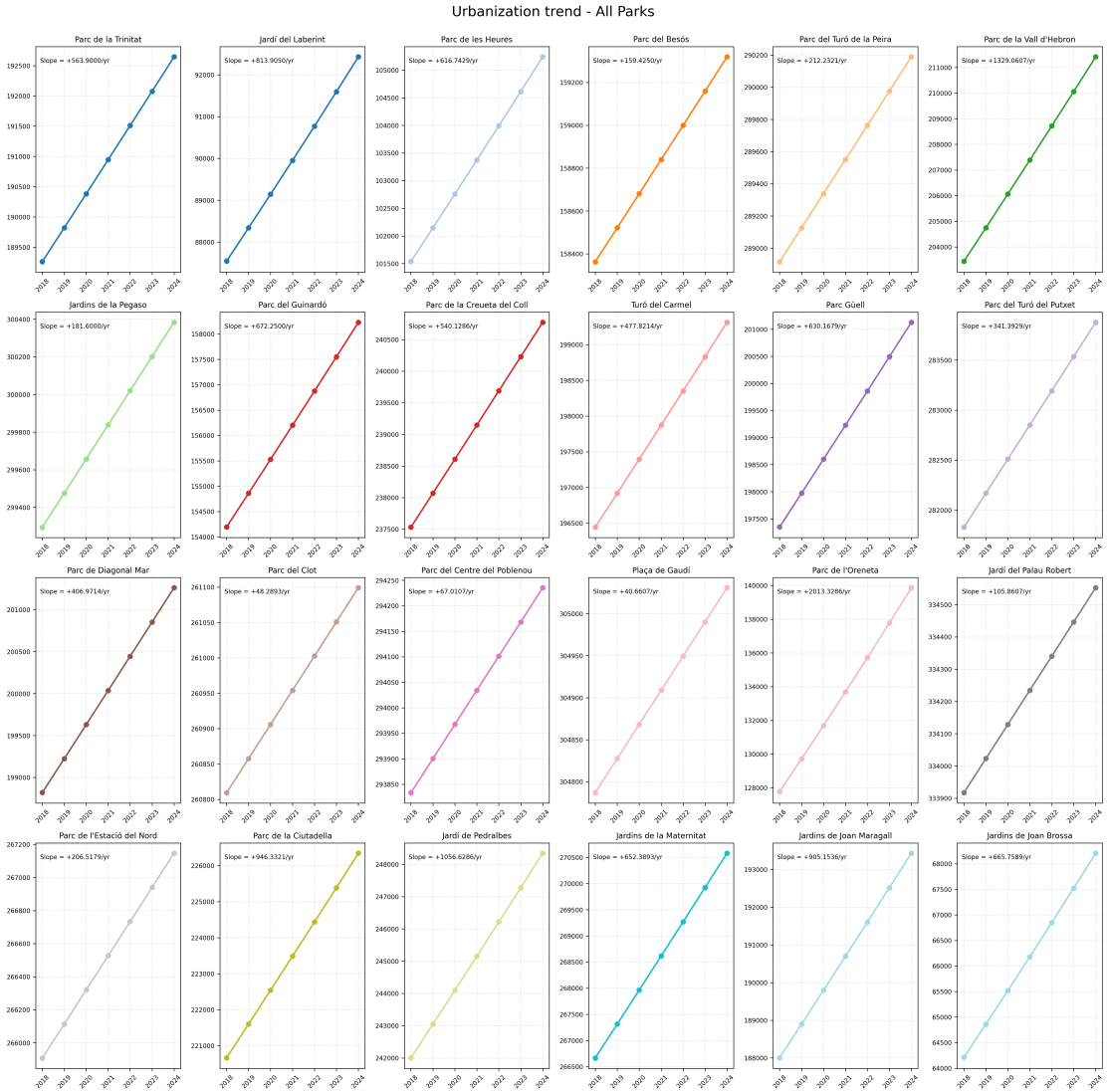


Figure 2.13: Urbanisation trend in all 24 parks in Barcelona over the years.

- **Temperature**: temperature is a key factor in the population dynamics of adult butterflies, as their appearance is closely linked to it. There are specific temperature data for each park each year.
  In the data received, each park was divided into two sections called transects A and T, and it was checked that the temperature of each was the same so that only one of the two temperatures could be used to study the park.

  Note from figure 2.14 that the temperature increase over time as the urbanisation.

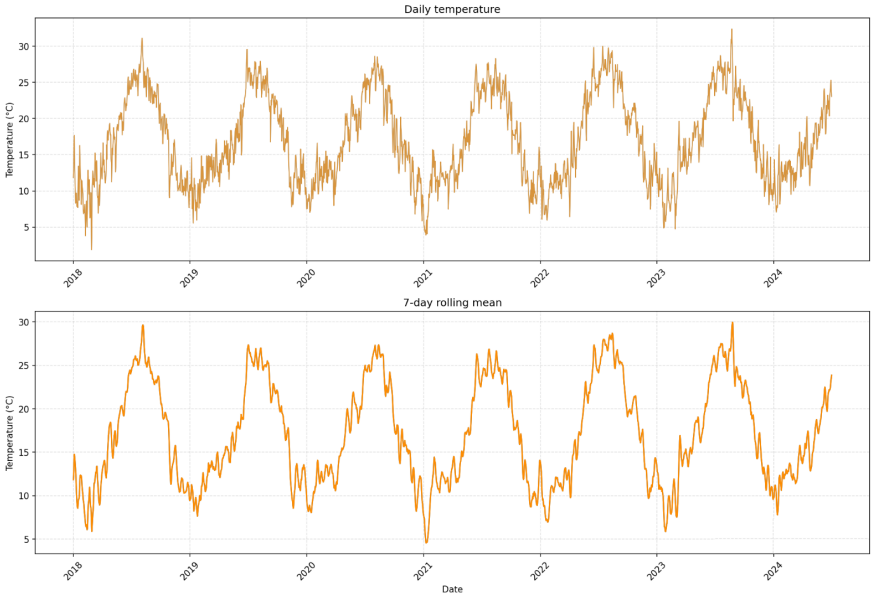


Figure 2.14: Example of temperature in a Barcelona park over years 2018-2024 in Parc de la Trinitat.

- **Isolation**: another key factor in analysing the results and understanding spatial dynamics is the degree of isolation of each park. In this study, isolation is quantified using two different approaches. In general, it reflects the idea that areas located farther from other suitable habitats are more difficult to reach and therefore less likely to be colonised, although this strongly depends on the species' dispersal ability and mobility.
  - Isolation between parks: how isolated a park is from the others.
    - Isolation based on connectivity 2.2 (inversely proportional relationship), from the most isolated to the least.
      - Parc de l'Oreneta [17] (0.40)
      - Jardì del Palau Robert [18] (0.48)
      - Jardins de la Pegaso [7] (0.53)
      - Plaça de Gaudí [16] (0.60)
      - Parc de la Ciutadella [20] (1.03)
      - Parc del Turó de la Peira [5] (2.04)
      - Parc del Clot [14] (2.65)
      - Parc de la Creueta del Coll [9] (3.50)
      - Parc del Turó del Putxet [12] (3.93)
      - Parc de l'Estació del Nord [19] (4.43)

- Jardì de Pedralbes [21] (4.94)
- Jardins de Joan Brossa [24] (6.54)
- Parc del Centre del Poblenou [15] (8.02)
- Parc de la Trinitat [1] (14.03)
- Parc de Diagonal Mar [13] (13.95)
- Parc del Guinardó [8] (18.12)
- Parc Güell [11] (19.64)
- Jardins de la Maternitat [22] (19.96)
- Turó del Carmel [10] (24.69)
- Parc de la Vall d'Hebron [6] (29.22)
- Parc del Besós [4] (34.15)
- Jardins de Joan Maragall [23] (41.61)
- Parc de les Heures [3] (129.23)
- Jardì del Laberint [2] (140.84)

* Isolation based on Euclidean distance: this metric is based on the distance to the nearest neighbouring patch, providing a simple measure of how spatially isolated a park is in its surroundings. It can be interpreted as how "surrounded" a park is by other suitable habitats in all directions. In practice, it is computed either as the distance to the closest neighbour or as the average distance to the 3 nearest neighbours. A park is considered highly isolated when even its closest neighbouring patch is located at a large distance (results obtained in the table 2.7).

Rank	Park	Nearest neighbor	Average 3 neighbors
1	Parc de la Trinitat [1]	2.494 m	3.071 m
2	Parc del Besós [4]	2.062 m	2.596 m
3	Jardins de la Pegaso [7]	1.927 m	2.037 m
4	Jardí del Palau Robert [18]	1.507 m	1.878 m
5	Parc de l'Oreneta [17]	1.405 m	2.101 m
6	Plaça de Gaudí [16]	1.369 m	1.455 m
7	Parc del Turó de la Peira [5]	1.348 m	1.597 m
8	Parc del Guinardó [8]	1.204 m	1.354 m
9	Parc de Diagonal Mar [13]	1.149 m	1.729 m
10	Parc del Turó del Putxet [12]	1.062 m	1.250 m
11	Jardins de Joan Brossa [24]	938 m	2.425 m
12	Jardins de la Maternitat [22]	938 m	2.384 m
13	Parc del Clot [14]	833 m	1.343 m
14	Parc del Centre del Poblenou [15]	833 m	1.340 m
15	Jardins de Joan Maragall [23]	709 m	1.966 m
16	Jardí de Pedralbes [21]	709 m	1.737 m
17	Parc de la Vall d'Hebron [6]	667 m	990 m
18	Parc de la Creueta del Coll [9]	664 m	865 m
19	Parc de la Ciutadella [20]	600 m	1.585 m
20	Parc de l'Estació del Nord [19]	600 m	1.225 m
21	Parc de les Heures [3]	540 m	1.036 m
22	Jardí del Laberint [2]	540 m	1.108 m
23	Turó del Carmel [10]	387 m	758 m
24	Parc Güell [11]	387 m	704 m

Table 2.7: Ranking of parks by distance from neighbors based on euclidean distance.

For each park, it was computed the straight-line (euclidean) distance to every other park using the standard formula

$$d(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2},$$

where coordinates were taken from the Barcelona Municipal Services matrix. Two complementary metrics summarise how "surrounded" a park is. The nearest-neighbor distance d_1 is obtained by taking all distances to the other 23 parks, excluding the zero distance to itself, and keeping the minimum value. This metric measures how close the nearest park is: if d_1 is large, the park is truly isolated, since no other park lies nearby in any direction. The second metric, the average distance of the three nearest neighbors,

$$\frac{d_1 + d_2 + d_3}{3},$$

is computed by sorting the same distances from smallest to largest, selecting the three closest parks, and averaging their distances. This provides

a broader measure of "circular isolation": a park may have one neighbor quite close, but if its second and third neighbors are already far away, the park will still appear relatively isolated according to this metric.

The graph below 2.15 illustrates the methodology used: Parc de la Trinitat, the most isolated park, shows a wide radius before reaching its nearest neighbor (Jardins de la Pegaso, $\approx 2\,494$ m), while Parc Güell and Turó del Carmel sit almost on top of each other (≈ 387 m), resulting in the smallest isolation scores in the dataset.

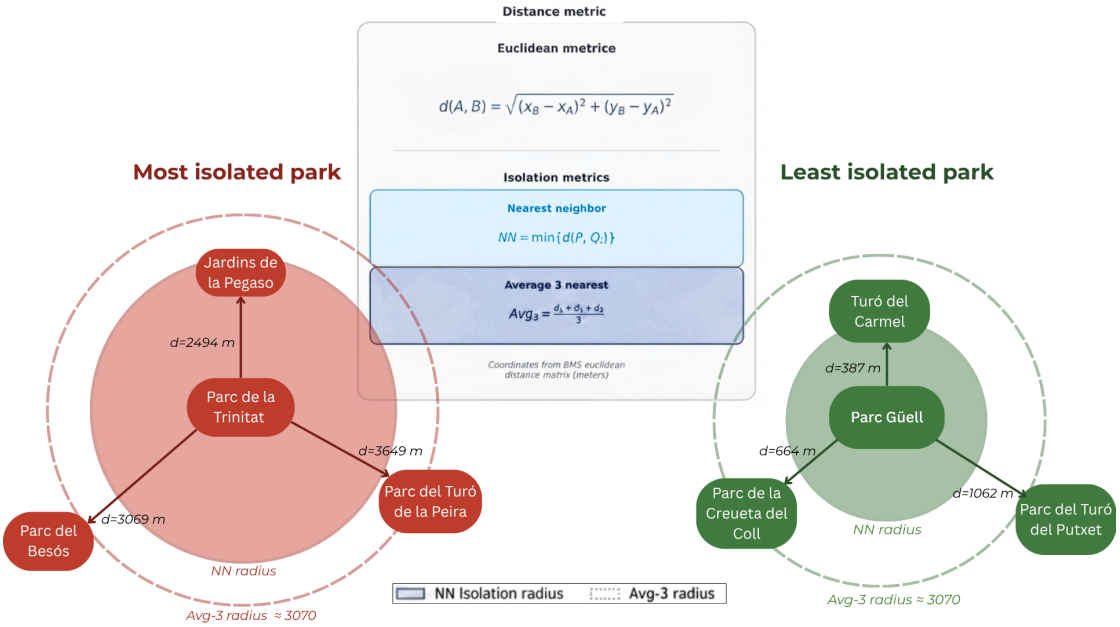


Figure 2.15: Calculation of isolation between parks based on euclidean distance.

– Isolation from outside: it describes how far a park is located from what can be considered the "outside" of the urban system, i.e., areas beyond the internal network of Barcelona's parks. Urban parks are often influenced not only by local dynamics but also by external ecological sources, and quantifying this influence requires measuring their distance from the most significant external green infrastructure.
In this context, the reference area is the Serra de Collserola Natural Park (map in the figure 2.16), which represents a major reservoir of biodiversity and an important source of species dispersal for the entire metropolitan region. Although external ecological influence may originate from multiple natural areas surrounding the city, Collserola was selected as the

primary reference because it constitutes the largest and most clearly defined source of ecological influence, making it the most straightforward to quantify in a first approximation. The distance from each park to Collserola is therefore used as a proxy for external ecological influence: parks closer to Collserola are expected to experience stronger ecological exchange with surrounding natural habitats, while more distant parks are generally more embedded in the urban matrix.

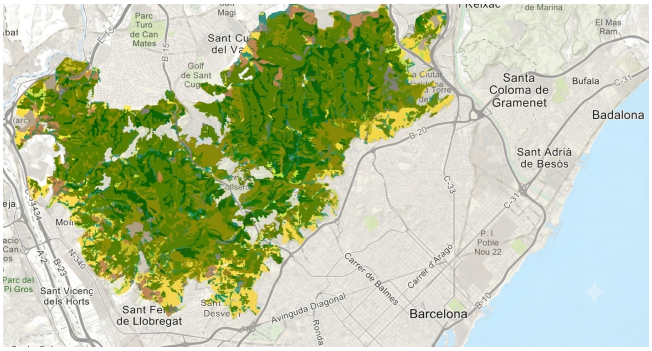


Figure 2.16: Map of the Collserola park territory.

In the figure 2.17, the perimeter of the Serra de Collserola Natural Park is represented as a simplified GeoJSON polygon with coordinates in WGS84 (longitude, latitude) format. All coordinates fall within the expected geographical range for Barcelona (longitude approximately 2.05–2.17°E, latitude approximately 41.40–41.47°N). The polygon contains approximately 43,000 vertices, making it highly detailed and suitable for accurate spatial representation.[4]

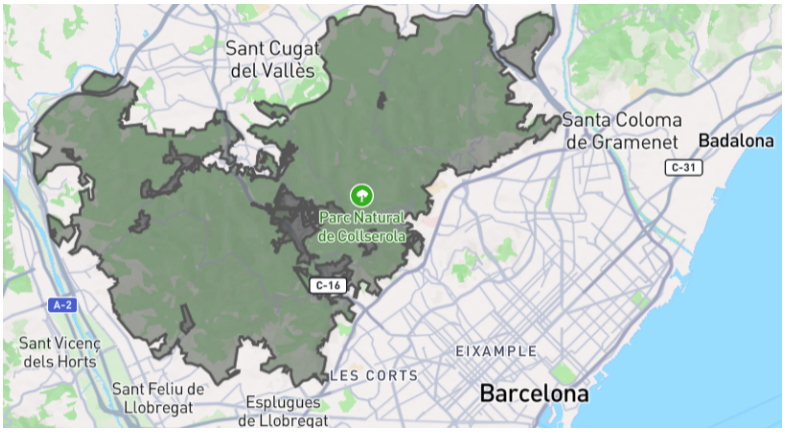


Figure 2.17: Collserola polygon representation with Geojson.io.

The distance between each park and the "outside" reference (Collserola), ranging from the closest to the most distant, was then calculated after performing all necessary coordinate transformations and the resulting values are listed below (visually in the figure 2.18):

* Jardí del Laberint [2] (167 m)
* Parc de les Heures [3] (175 m)
* Parc de l'Oreneta [17] (343 m)
* Parc de la Vall d'Hebron [6] (493 m)
* Parc de la Creueta del Coll [9] (687 m)
* Parc de la Trinitat [1] (920 m)
* Jardí di Pedralbes [21] (1.063 m)
* Parc del Turó de la Peira [5] (1.108 m)
* Parc del Turó del Putxet [12] (1.262 m)
* Parc Güell [11] (1.313 m)
* Turó del Carmel [10] (1.361 m)
* Jardins de la Maternitat [22] (1.646 m)
* Parc del Guinardó [8] (2.448 m)
* Jardins de la Pegaso [7] (2.627 m)
* Jardí del Palau Robert [18] (3.191 m)
* Plaça de Gaudí [16] (3.611 m)
* Parc del Besós [4] (4.534 m)
* Parc del Clot [14] (4.634 m)
* Parc de l'Estació del Nord [19] (4.863 m)
* Parc del Centre del Poblenou [15] (5.000 m)
* Jardins de Joan Maragall [23] (5.289 m)
* Parc de la Ciutadella [20] (5.372 m)
* Parc de Diagonal Mar [13] (5.483 m)
* Jardins de Joan Brossa [24] (6.385 m)

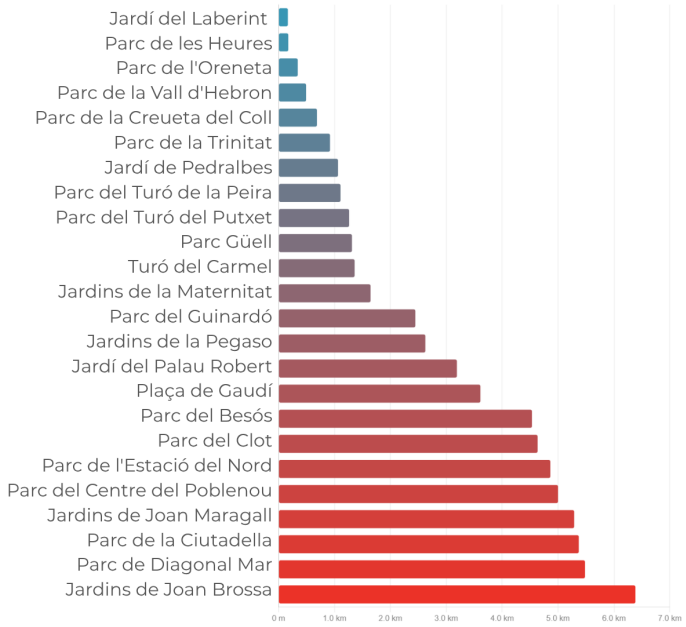


Figure 2.18: Graphic representation of distances parks-Collserola (outside Barcelona).

2.4 Detectability and Data Uncertainty

The term *detectability* (or detection probability) refers to the probability of observing or capturing an individual of a given species, assuming that the individual is actually present in the study area during the sampling survey. When an observer (a researcher or a volunteer) conducts field monitoring and fails to record a specific butterfly species, this zero-count output can be attributed to two distinct factors:

1. true absence: the species is genuinely absent from the surveyed site;
2. false negative (imperfect detection): the species is present within the area (e.g., concealed beneath vegetation, resting, or out of the observer's immediate line of sight) but goes undetected.

Relying solely on raw count data without accounting for detectability poses a significant risk of introducing systematic bias into ecological datasets. For instance, a species might be erroneously classified as rare due to its cryptic behavior or elusiveness, whereas its actual abundance might be substantially higher. In the context of lepidoptera monitoring, detection probability is never constant; rather, it fluctuates as a function of several intersecting covariates [45]:

- Species-specific traits: morphological and phenological characteristics heavily influence visibility. Large, brightly colored butterflies exhibit inherently

higher detectability compared to smaller or mimetic species. Furthermore, univoltine species (producing only one generation per year) present narrower flight windows, rendering them more challenging to detect.

- Observer expertise: sampling efficiency varies significantly with surveyor experience. A seasoned volunteer with a decade of field monitoring experience will display a significantly higher detection capacity than a novice tracker.

- Environmental and meteorological conditions: adverse weather conditions, such as high wind speeds or dense cloud cover/shade, severely suppress the flight activity of lepidoptera, thereby drastically reducing their detectability.

To correct for these observation errors and derive robust abundance estimates, contemporary ecological studies employ Bayesian hierarchical models. These frameworks are specifically designed to adjust raw counts for imperfect detection and to mitigate the stochastic observation noise inherent to field sampling [45].
Empirical evidence indicates that detection probability can exhibit a linear increase over long-term monitoring periods. For instance, studies spanning two decades of data collection have observed this upward trend across all volatility types (univoltine, bivoltine, and multivoltine species) [45]. This systematic improvement is primarily attributed to a "learning-by-doing" effect, reflecting the progressive skill acquisition and data-collection proficiency of citizen science volunteers. Nevertheless, interspecific variations remain pronounced, with the detection probability of univoltine species consistently lagging behind that of multivoltine counterparts.
Accounting for this variable is paramount for the validity of demographic trends. Models that overlook temporal shifts in detectability tend to underestimate negative population trends (declines) and overestimate positive ones [45]. This distortion occurs because an apparent increase in population density over time may simply be an artifact of enhanced observer efficiency rather than a genuine ecological recovery.

In conclusion, the intrinsic uncertainty embedding field-collected ecological data underscores a fundamental methodological caveat: the model implemented in this thesis inevitably operates as an approximation of a system that is already partially approximated at its source. Quantifying the precise error margin attributable to imperfect detection remains highly challenging within the current framework, given the data acquisition modalities and the heterogeneous nature of the available raw datasets.

2.5 Neural Network Data

Given the inherent uncertainty, sparsity, and observational gaps typically associated with raw ecological field counts, this thesis does not rely directly on raw observational data. Instead, it utilizes a robust, continuous baseline dataset generated through a specialized neural network (reference to figures 2.21 and 2.22), developed in collaboration with ecologists [3].

A neural network is a computational framework composed of interconnected layers of artificial neurons designed to process multidimensional information. Figure 2.19 provides a schematic representation of this architecture, where information flows from the input layer through one or more hidden layers before reaching the output layer.

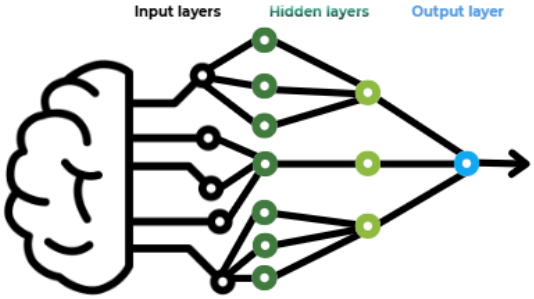


Figure 2.19: Neural network layers.

As input data propagates through the network, these hidden layers hierarchically extract features, allowing the model to uncover highly complex, non-linear relationships between variables. This machine learning methodology is utilized in this research because traditional statistical approaches often struggle with the intricate, multidimensional dynamics of ecological datasets; neural networks, by contrast, can accurately map these interactions without imposing restrictive, predefined functional forms.
Structurally, the neural network adopted in this study consists of multiple fully connected hidden layers designed to capture the complex, non-linear relationships between environmental variables and butterfly abundance. The model is trained to perform a regression task, predicting seasonal abundance from a set of climatic, temporal, spatial, and urbanisation-related predictors. Further details on the network architecture, input variables, and training procedure are provided in [3].

A crucial operational advantage of this neural network is its capacity to effectively predict missing survey observations and reconstruct complete, continuous seasonal abundance curves across highly fragmented time series as illustrated in the figure 2.20.

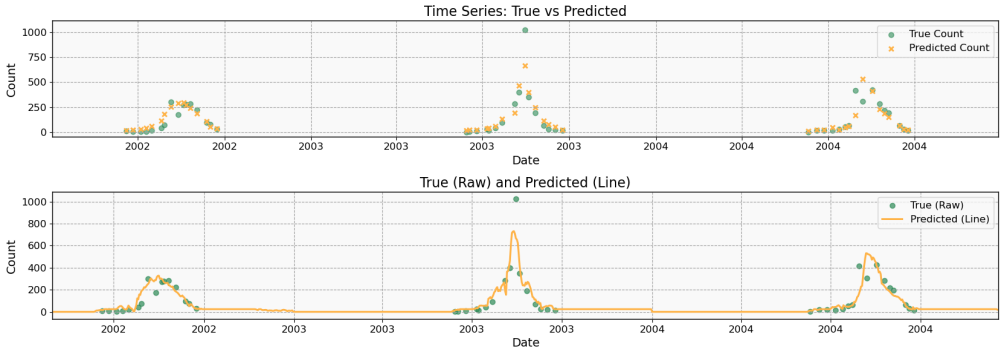


Figure 2.20: Basic demonstration of neural network results [3].

Furthermore, the ecological interpretations derived from this framework demonstrated that urbanisation significantly alters fundamental population dynamics, specifically impacting phenological timing, flight periods, and voltinism. Consequently, overcoming the limitations of discontinuous and potentially biased raw counts, the continuous abundance curves predicted by this neural network serve as an empirical, high-fidelity "ground truth."

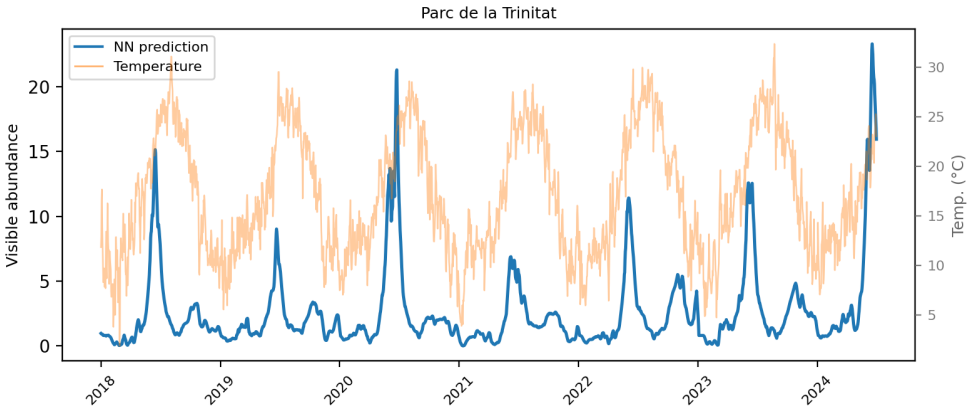


Figure 2.21: Neural network data (blue) from the Pieris rapae species in Parc de la Trinitat with background temperature (orange) [3].

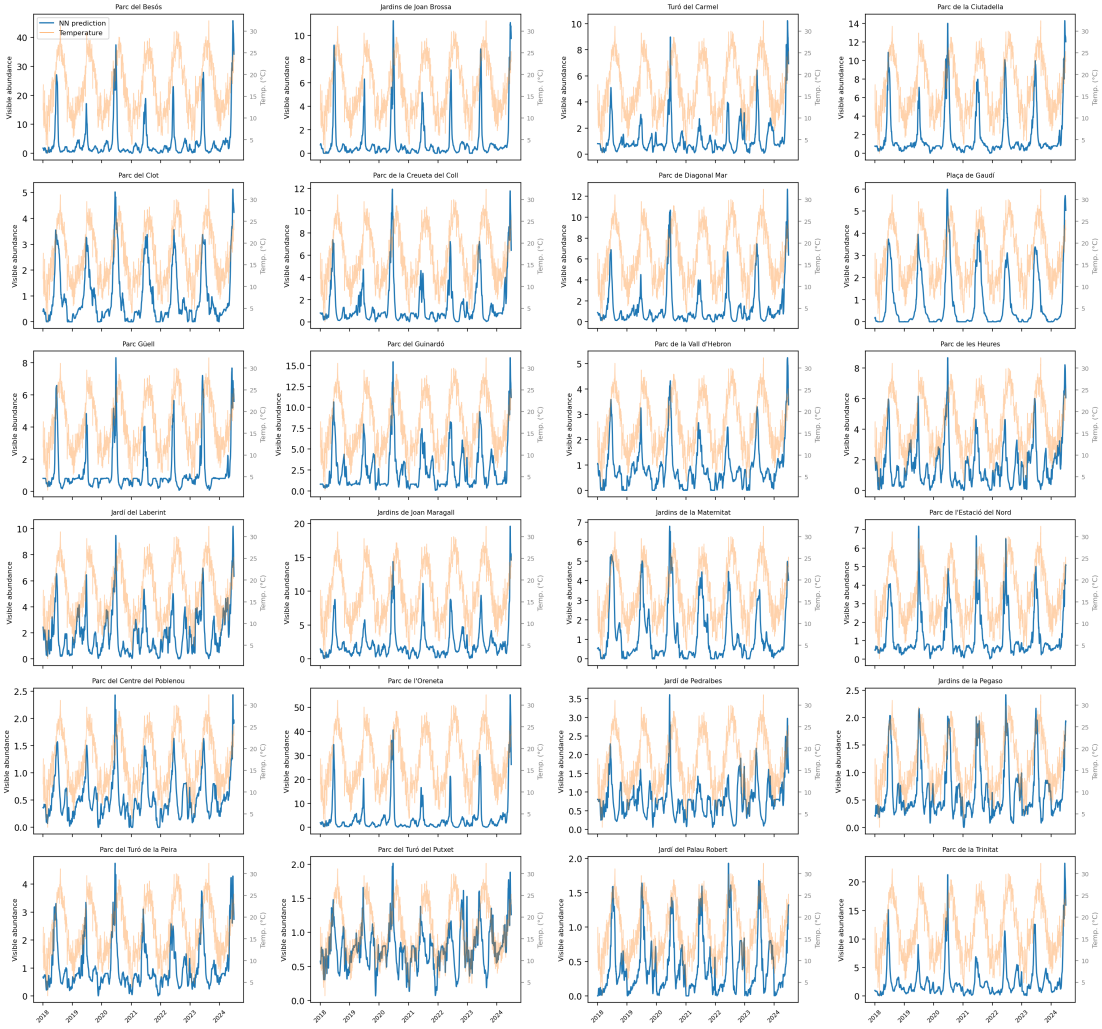


Figure 2.22: Neural network data (blue) from the Pieris rapae species in all 24 parks with background temperature (orange)[3].

These refined, data-driven predictions form the authoritative foundational dataset upon which the mathematical model proposed in this thesis will be implemented, evaluated, and comprehensively compared.

Chapter 3

Mathematical Model

As discussed in section 1.2, the models available in the literature provide robust theoretical frameworks for describing ecological dynamics. However, it is important to emphasize that, in mathematical modelling, there are no intrinsically right or wrong models; rather, the suitability of a model depends on the specific research question being addressed, the characteristics of the available data, and the ecological processes of interest. Although the existing models constitute valuable references, none fully captures the ecological dynamics of the urban butterfly community investigated in this study. For this reason, and to better address the objectives of the present project using the available dataset, the mathematical formulation described in section 1.3 has been adopted. This chapter is therefore devoted to present the proposed model in detail, grounding its mathematical structure in the biological framework introduced in the preceding chapters.

The modelling strategy adopted is built around two complementary mathematical structures: ordinary differential equations (ODEs), which govern the temporal evolution of butterfly population abundance at each habitat patch, and a weighted graph, which encodes the spatial relationships and dispersal pathways between parks. This combination allows the model to capture both the local population dynamics within each site and the movement of individuals across the urban landscape, thereby representing the two fundamental spatial scales at which butterfly dynamics operate.

3.1 Procedural idea

The model was developed incrementally through three phases of increasing complexity, each introducing an additional ecological mechanism while preserving the structure established in the previous step. This progressive modelling strategy was

intentionally adopted to assess whether increasingly complex ecological dynamics provide a more realistic representation of butterfly populations. The development started from the simplest formulation, based on biologically plausible assumptions regarding butterfly movement and the mechanisms expected to drive dispersal. These assumptions reflect ecological processes for which empirical data are available or whose effects can be reasonably estimated. Successive model extensions were then introduced to investigate whether incorporating additional ecological mechanisms leads to a more accurate and biologically meaningful description of the observed dynamics.
The main modelling phases are outlined below, while their mathematical formulation and implementation are presented in detail in the following sections.
In all three phases, Barcelona is modelled as a closed ecological system, i.e., no immigration from or emigration to the surrounding environment is considered. The differences between the phases concern exclusively the representation of dispersal within the urban park network.

- **Phase 0: Closed system without inter-patch dispersal.**
  Each park evolves independently, with no exchange of individuals between sites. Population dynamics are governed exclusively by local demographic processes, namely population growth and mortality. This baseline formulation is first applied to the most abundant butterfly species to establish a reference model.

- **Phase 1: Closed system with distance-based inter-patch dispersal.**
  Inter-patch dispersal is introduced through directed flows along the edges of the graph while preserving the closed-system assumption. Edge weights are defined solely by the Euclidean distance between park centroids, providing a first approximation of spatial connectivity.

- **Phase 2: Closed system with attractiveness-based habitat connectivity.**
  The representation of connectivity is further refined by incorporating habitat attractiveness and the degree of urbanisation of the surrounding matrix. In this framework, habitat quality is also introduced as an additional weighting factor, together with urbanisation, in shaping dispersal pathways between patches. Edge weights are computed using a cost-distance approach, in which Euclidean distance is modulated by the built-up cover and other landscape-quality components. As a result, movement depends not only on geometric proximity but also on the ecological permeability and overall quality of the urban landscape.

Several alternative formulations were therefore explored throughout the modelling process, as will be described in detail in section 3.4.

3.2 Population dynamics

The objective of the model is to describe the spatial distribution of observable butterfly populations within the urban environment. This requires a number of simplifying assumptions and modelling compromises, as the observed adult population represents the final outcome of a sequence of underlying life-history stages that are not explicitly resolved due to data limitations. The main assumptions underlying the model are the following:

- interspecific interactions are neglected;
- population sources arising from unobserved life-history stages are not explicitly modelled;
- the mortality term is assumed to be independent of temperature;
- fluxes between habitat patches are assumed to be independent of temperature.;
- for all species, the probability of an individual butterfly being present in a garden increases with habitat quality, garden size, and landscape connectivity [1];
- the influence of landscape variables is assumed to strengthen along a gradient from wide- to narrow-breadth niche traits (i.e., from generalist, highly mobile, and highly reproductive species to specialists with limited mobility and low reproductive rates), implying that species at the extreme of these restrictive trait combinations are expected to be rare or absent in the urban environment [1].

Under these assumptions, the temporal dynamics of the population of a butterfly species s at site x are modelled as the combination of three main processes.

$$\begin{array}{c}\text{Variation over time}\\ \text{of the population}\\ \text{of species } s \text{ at site } x\end{array} = \underbrace{\begin{array}{c}\text{Adult butterfly}\\ \textit{("Source")}\\ \hookrightarrow\text{Temperature}\end{array}}_{\text{Fly-period}} - \underbrace{\text{Dead} + \text{No Detection} + \text{Other stages}}_{\text{Non-Activity}} + \underbrace{\begin{array}{c}\text{Graph}\\ \textit{("Shift")}\end{array}}_{\text{Flux}}$$

The first is a source term, which represents an effective input of observable butterflies into the system. This term should not be interpreted as a classical biological birth rate, since the model does not explicitly resolve all life-cycle stages. Instead, it describes the transition of individuals into a detectable (i.e., observable) adult state and will therefore be referred to as the *fly-period term*. Unobserved

processes are implicitly subsumed within this formulation, contributing in aggregate to the observed adult population. Accordingly, reproduction is not modelled mechanistically but is incorporated implicitly through this external input term, which depends on environmental drivers such as temperature. In particular, it is well known that the emergence and detectability of butterflies is strongly linked to temperature conditions and to the accumulation of thermal resources, commonly quantified through growing degree days (GDD), also referred to as heat units, which measure accumulated heat above a given base temperature.
The second contribution is a removal term, which accounts not only for actual mortality but also for all processes leading individuals to exit the observable system. These include transitions to unobserved life stages as well as conditions under which individuals become undetectable within the observational framework. This term will be referred to as the *non-activity term*. Due to the lack of specific data or further biological constraints, this component is assumed to be independent of temperature.
Finally, the model includes a *flux term* describing the movement and dispersal of butterflies among habitat patches. This component is the central focus of the modelling framework, as the primary objective of the model is to characterise and analyse inter-patch dispersal dynamics within the urban landscape. This term will be implemented by incorporating species-specific biological traits through the introduction of a graph-based mathematical tool, which is particularly well-suited to representing spatial movement and connectivity patterns in a rigorous and flexible way.

This formulation can be naturally translated into a system of ordinary differential equations (ODEs), where each of the three contributions is expressed as a mathematically well-defined term acting on the temporal evolution of the population:

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))}_{Flux}$$

Species: $s = 1, ..., 44$ (details in 2.2).
Sites: $x = 1, ..., 24$ (details in 2.3).
Time: t =day in $\mathcal{T}$ =Year.

- P_x^s = Popolation of visible adults of species s in the site x ($P_x^s \geq 0$);

- $T_x(t)$ = Temperature at site x at time t ($T_x(t) > 0$ [$^\circ C$]).

Importantly, each term is biologically motivated in the sense that it corresponds to an interpretable ecological mechanism underlying the observed dynamics. To clarify this interpretation, the parameter legend is provided below:

- $\alpha^s(T_x(t))$: temperature-dependent fly-period (activity) rate $[\frac{1}{t}]$ as a function of temperature $T\,[^\circ C]$.
  This term represents the rate at which new adults become observable at site x, independently of the current adult population size, thereby capturing emergence from juvenile stages. It is modelled using a Gaussian functional form, consistent with empirical evidence suggesting a unimodal dependence on temperature:
  $$\alpha^s(T_x(t)) = a\, e^{-\frac{(T-T^*)^2}{2\sigma_T^2}}$$
  where:
  - $a > 0$: maximum emergence (peak daily activation into visibility),
  - $T^* > 0$: optimal temperature,
  - $\sigma_T > 0$: thermal breadth (width of the response curve).
- ω^s: non-activity rate $[\frac{1}{t}]$, representing the rate at which individuals leave the observable class (due to mortality, inactivity, or transitions to unobserved stages). This term is assumed to be temperature-independent.
  The inverse quantity,
  $$\frac{1}{\omega^s},$$
  corresponds to the mean duration of observability. Once active, an individual remains detectable for an average time $\frac{1}{\omega^s}$, which can be interpreted as the characteristic decay or memory timescale of the system.
- $\sum_y \left(\underbrace{\mu^s W_{yx} P_y^s}_{\text{immigration}} - \underbrace{\mu^s W_{xy} P_x^s}_{\text{emigration}} \right)$: Dispersal flux between patches.
  The difference $(P_y^s - P_x^s)$ acts as the population gradient driving movement, while W_{xy} denotes the weighted adjacency matrix of the graph. A detailed characterization of this term is provided in Section 3.3.

The modelling framework is designed to capture the observable structure of butterfly communities while maintaining a tractable representation of the underlying ecological processes. It should therefore be interpreted as an observational rather than a fully mechanistic demographic model, in which several biological processes are subsumed into effective terms.

3.3 Connections

As previously mentioned, the most important aspect of the survey is the assessment of flux, namely the connections between habitat patches through butterfly dispersal. Understanding these dispersal patterns will provide valuable insights into the ecological quality of the parks by identifying metapopulation components and distinguishing habitats that act as ecological traps from those that support viable populations.

$$\underbrace{\sum_{y} \mu^s (W_{yx} P_y^s(t) - W_{xy} P_x^s(t))}_{Flux} = \sum_{y} \left(\underbrace{\mu^s W_{yx} P_y^s}_{\text{immigration}} - \underbrace{\mu^s W_{xy} P_x^s}_{\text{emigration}} \right)$$

This term, which represents immigration and emigration across all habitat patches, was incorporated into the analysis through a graph-based approach. Graphs provide a powerful and appropriate framework for representing the spatial structure of Barcelona and, in particular, the network of urban parks, allowing the connectivity among habitat patches to be explicitly described.

Before introducing the graph in detail, it is important to explain the rationale behind the inclusion of the term μ^s in the mathematical model:

- μ^s: mobility (or diffusivity) coefficient $\left[\frac{1}{t}\right]$.
  This parameter quantifies the intrinsic mobility of species s, representing its tendency to move between habitat patches. As discussed in section 2.2, mobility is a species-specific characteristic closely related to habitat selectivity and dispersal ability, and it provides a natural distinction between specialist and generalist species. Lower values of μ^s are associated with highly specialized species, which exhibit limited dispersal and stronger habitat preferences, whereas higher values correspond to more generalist species with greater dispersal capacity. Consequently, μ^s directly regulates the flux term of the model by determining the intensity of movement among patches.
  Since mobility is assumed to increase with the degree of generalism (see section 2.2), the mobility coefficient is assigned at the cluster level for each species class ($s \in \{D1, \ldots, D4\}$) according to the following normalized scale:
  - $s \in$ Cluster 1 $\Longrightarrow \mu^s = 0.25$
  - $s \in$ Cluster 2 $\Longrightarrow \mu^s = 0.50$
  - $s \in$ Cluster 3 $\Longrightarrow \mu^s = 0.75$
  - $s \in$ Cluster 4 $\Longrightarrow \mu^s = 1.00$

Therefore, the mobility coefficient is assumed to vary within the normalized interval $\mu^s \in [0,1]$, with increasing values corresponding to progressively greater dispersal ability.

3.3.1 Graph

To represent the spatial component of the model, Barcelona is modeled as a graph in which each habitat patch is represented by a node (figure 3.1). The possible movement paths between parks are described by weighted edges, whose weights characterize the distance or movement cost associated with dispersal from one patch to another.

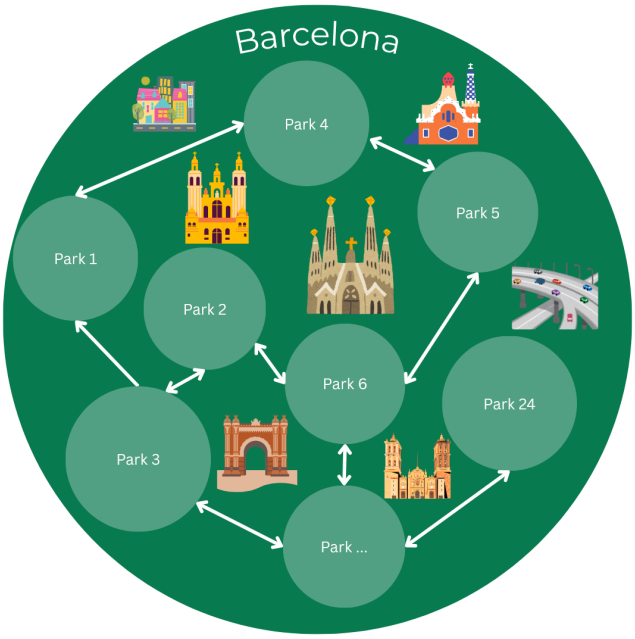


Figure 3.1: Schematic representation of Barcelona as a graph.

From a mathematical point of view:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$$

- $\mathcal{V}$ = nodes = patches = 24 parks in Barcelona
- $\mathcal{E}$ = arches = connections between parks (dispersal pathways)
- $\mathcal{W}$ = weights = f cost $\forall$ arch = determinant of butterfly movement

(Details in appendix A, A.1).

The weights and the associated matrix have been implemented as follows.

Weighted adjacency matrix: this matrix represents the strength of connections between parks in the network. Each element W_{xy} quantifies the cost or weight associated with movement from park x to park y. In other words, it encodes how easily individuals can move between two nodes: higher values correspond to stronger or more favorable connections, while lower values indicate weaker or less likely dispersal pathways. The matrix therefore combines both the structure of the graph and the functional intensity of movement across it.

$$W_{xy} = A_{xy}\, w_{xy} \qquad \mathbf{W} = \mathbf{A} \odot \mathbf{w}$$

$$\mathbf{W} = \begin{bmatrix} W_{11} & W_{12} & \cdots \\ W_{21} & W_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}^{24\times 24}$$

- $A_{xy} \in \{0,1\}$: adjacency matrix. The adjacency matrix $\mathbf{A}$ encodes the connectivity between parks and determines whether movement between two locations is allowed. In general:

$$\begin{cases} A_{xy} = 1 \;\Rightarrow\; \text{movement from patch } x \text{ to patch } y \text{ is possible} \\ \qquad \longrightarrow \exists \text{ arch between } x - y \\ A_{xy} = 0 \;\Rightarrow\; \text{movement from patch } x \text{ to patch } y \text{ is not possible} \\ \qquad \longrightarrow \nexists \text{ arch between } x - y \end{cases}$$

  Typically, $\mathbf{A}$ is treated as a fixed component of the system's structure rather than as a parameter to be estimated. If all entries of $\mathbf{A}$ are set to 1, then every park is connected to every other park, and movement is influenced solely by distance (and any additional modeled effects). At this stage of the project, it is reasonable to assume a fully connected network, i.e., $A_{xy} = 1$ for all park pairs.
  The goal is to first evaluate whether incorporating dispersal and attraction improves model performance. With $\mathbf{A}$ fixed to ones, Phase 1 focuses on distance-driven movement only, while Phase 2 extends the model by adding attraction effects in addition to distance. Allowing $\mathbf{A}$ to be estimated would introduce a large number of additional parameters, making the model harder to interpret and increasing the risk of overfitting.
  If $\mathbf{A}$ is fixed to all ones, the model cannot explicitly represent complete disconnection between parks. However, it can effectively approximate this situation by assigning very small values to W_{xy}, making movement between certain pairs practically negligible. In many cases, this soft inhibition of movement

is more biologically plausible than enforcing a strict zero probability of transition.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{24\times 24}$$

Then it is assumed that every route is impossible $A_{xy} = 1 \; \forall x, y$, it will then be the optimization with real data that will give information on the real routes traveled by the butterfly.

- w_{xy} : edge weight

$$w_{xy} \in \mathbb{R}^{+} \begin{cases} w_{xy} \geq 0 \Longrightarrow attractiveness \; + cost \\ w_{xx} \geq 0 \Longrightarrow attractiveness \; (no \; cost) \end{cases} \quad \exists \text{ self-loop.}$$

The edge weights represent the factors that drive butterfly movement between parks and were defined in consultation with biological experts. Butterfly dispersal may occur because certain species have an intrinsic tendency to move or because individuals perceive another park as a suitable and attractive habitat. The decision to move is therefore influenced by both the attractiveness of the destination and the cost associated with reaching it. Although butterfly movement is affected by many ecological and behavioral factors that are not fully understood or for which no data are available, the present work focuses on the variables that can be reliably quantified.
Specifically, movement is assumed to depend on three key characteristics: the distance between parks, their connectivity and the quality of the destination park (see section 2.3).

$$w_{xy} = f(\underbrace{d_{xy}}_{distance} , \underbrace{C_y}_{connectivity} , \underbrace{Q_y}_{quality}) \geq 0$$

  - d_{xy} = euclidean distance.
    Butterfly movement is strongly influenced by the distance between parks. As previously noted, different species exhibit different dispersal capacities, with some individuals being able to travel significantly further than others. For this reason, distance represents a fundamental component in determining movement flows within the system. This term was computed as described in section 2.3 and the numerical values used are found in the table 2.4.
  - C_y = connectivity.
    This term accounts not only for the influence of geographic distance on butterfly movement, but also for the fact that the effective cost of dispersal may vary depending on the level of urbanization along the path.

In particular, highly urbanized areas can increase movement resistance, while more natural landscapes facilitate dispersal. As described in section 2.3, this effect is incorporated through the Haski index 2.2, which provides a measure of landscape-mediated movement difficulty.

– Q_y = quality.
Last but not least, park quality, defined as in section 2.1 and computed based on vegetation characteristics, plays a crucial role in butterfly movement decisions. In fact, higher-quality habitats are assumed to be more attractive and therefore more likely to attract dispersing individuals.

The weights w_{xy} are defined as a probability distribution over all possible destinations y reachable from a given site x. The choice of a *Softmax* (or Gibbs) formulation (details in appendix A, A.2) is motivated by the need to map ecological movement costs into a valid probabilistic framework. In reality, butterfly movement is not strictly deterministic: individuals tend to prefer low-cost or high-quality habitats, but they may still occasionally select suboptimal destinations, reflecting realistic stochastic dispersal behavior.
Moreover, movement decisions are inherently relative, since each destination competes with all other available options in the network. For this reason, transition probabilities are naturally expressed as a normalized function of relative ecological costs.

Formally, the weights satisfy:

$$0 \leq w_{xy} \leq 1, \qquad \sum_y w_{xy} = 1$$

$$w_{xy} \propto e^{-cost_{xy}}$$

$$\begin{cases} w_{xy} = \dfrac{e^{-cost_{xy}}}{\sum\limits_z e^{-cost_{xz}}} & y \neq x \\ w_{xx} = \dfrac{e^{-cost_{xx}}}{\sum\limits_z e^{-cost_{xz}}} & y = x \end{cases}$$

The movement cost is defined as:

$$cost_{xy} = \delta d_{xy} - \gamma C_y - \rho Q_y \tag{3.1}$$

where:

- d_{xy} is the euclidean distance between parks, representing the energetic and spatial cost of movement. Increasing distance leads to higher cost and therefore lower transition probability:

$$d_{xy} \uparrow \Rightarrow cost_{xy} \uparrow \Rightarrow w_{xy} \downarrow, \qquad d_{xx} = 0$$

- C_y is the connectivity of the destination patch, interpreted as a measure of accessibility within the ecological network. Highly connected parks are easier to reach and therefore more attractive:

$$C_y \uparrow \Rightarrow cost_{xy} \downarrow \Rightarrow w_{xy} \uparrow$$

- Q_y is the quality of the destination park, which reflects habitat suitability. Higher-quality patches are more attractive to individuals:

$$Q_y \uparrow \Rightarrow cost_{xy} \downarrow \Rightarrow w_{xy} \uparrow$$

- δ, γ, and ρ are positive parameters that control the relative contribution of distance, connectivity and habitat quality to the movement decision.

Substituting the cost definition into the Softmax yields:

$$w_{xy} = \frac{e^{-\delta d_{xy} + \gamma C_y + \rho Q_y}}{\sum_z e^{-\delta d_{xz} + \gamma C_z + \rho Q_z}}$$

The Softmax function ensures two key properties: (i) normalization of transition probabilities, and (ii) sensitivity to differences in ecological cost, as the exponential transformation amplifies even small variations. In this framework, low-cost or high-quality destinations are favored, while still allowing stochastic movement toward less optimal patches, preserving biological realism.
To illustrate the limiting behavior:

- If $cost_{xy} \to 0$, then $e^{-cost_{xy}} \to 1$, leading to higher probability.

- If $cost_{xy} \to \infty$, then $e^{-cost_{xy}} \to 0$, leading to negligible probability.

Finally, different modeling phases are defined by specifying different forms of the cost function 3.1, allowing the progressive incorporation of additional ecological mechanisms into the movement model:

- **Phase 1** (distance-only model):

$$cost_{xy} = \delta d_{xy} \tag{3.2}$$

  In this first stage, movement is driven exclusively by euclidean distance, providing a baseline description of dispersal without additional ecological structure.

- **Phase 2** (landscape-driven model):

$$cost_{xy} = -\gamma C_y - \rho Q_y \tag{3.3}$$

In this phase, distance is not explicitly included, as its effect, together with urbanization and spatial heterogeneity, is assumed to be embedded within the connectivity term C_y. This allows the model to focus on landscape-mediated effects, combining connectivity and habitat quality as the main drivers of movement decisions.

3.4 Mathematical analysis

In this section, the successive steps in the formulation of the final model are presented. For each phase, a mathematical analysis is provided to examine in detail the structure of the corresponding equations.

3.4.1 Evaluation

The original formulations of the model equations are presented below, together with the rationale for their exclusion, which ultimately led to the final specification of the model 3.4.

The first candidate formulation for the population dynamics was:

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))P_x^s(t)}_{Fly-period} - \underbrace{\omega^s(T_x(t))P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))}_{Flux}$$

$\Longrightarrow$ This approach was discarded because of the term $P_x^s(t)$ multiplying the emergence rate α^s: this formulation implicitly assumes that the emergence of newly observed adult butterflies is directly generated by the currently observed adult population $P_x^s(t)$. However, adult individuals actually originate from earlier life-history stages (e.g., eggs, larvae, and pupae), which are not explicitly represented in the model. Consequently, even if no adults are observed at a given time, new individuals may still appear in subsequent observations. Since the hidden stages of the butterfly life cycle are not monitored, their dynamics cannot be modelled explicitly. For this reason, the emergence term α^s is introduced as an external forcing term that summarizes the contribution of this unobserved latent population. The objective is therefore not to construct a fully mechanistic population model, but rather a phenomenological model capable of describing the observed dynamics of adult butterflies with the available data.

These considerations led to the following revised model:

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s(T_x(t))P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))}_{Flux}$$

$\Longrightarrow$ This formulation was also discarded because of the dependence $(T_x(t))$ in the non-activity term ω^s: the mortality/non-activity rate was initially assumed to depend on temperature, since extreme temperatures can significantly affect butterfly survival and activity patterns. However, this approach was rejected due to the lack of available data to properly characterize the relationship between temperature and mortality. As a consequence, such effects are implicitly excluded from the counting framework and cannot be reliably estimated within the present study.

For this reason, a simplified formulation was adopted, in which the parameter ω^s is assumed to be independent of temperature, leading to the final version of the model equation:

$$\boxed{\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{\text{Fly-period}} - \underbrace{\omega^s P_x^s(t)}_{\text{Non-Activity}} + \underbrace{\sum_y \mu^s\Big(W_{yx}P_y^s(t) - W_{xy}P_x^s(t)\Big)}_{\text{Flux}}} \tag{3.4}$$

The flow term redistributes the population across habitat patches without creating or destroying total mass. This property can be formalized through a conservation argument.

$$\Longrightarrow \sum_x \sum_y (\mu^s W_{yx} P_y^s - \mu^s W_{xy} P_x^s) = 0$$

Hence, the total population is modified only by the source term α^s and the decay term ω^s.

Proof. The double sum is decomposed into two contributions:

$$F \;=\; \underbrace{\sum_x \sum_y \mu^s W_{yx} P_y^s}_{F_1} \;-\; \underbrace{\sum_x \sum_y \mu^s W_{xy} P_x^s}_{F_2}$$

In the first term F_1, a relabelling of dummy indices $x \leftrightarrow y$ is performed. Since both indices range over the same set, the value of the sum is invariant under this exchange:

$$F_1 = \sum_x \sum_y \mu^s W_{yx} P_y^s \xrightarrow{x \leftrightarrow y} \sum_x \sum_y \mu^s W_{xy} P_x^s = F_2$$

It follows immediately that

$$F = F_1 - F_2 = 0$$

□

The flow term is antisymmetric: each contribution $+\mu^s W_{yx} P_y^s$, representing transfer from node y to node x, is exactly compensated by the opposite contribution $-\mu^s W_{xy} P_x^s$, corresponding to the reverse flow. As a result, all pairwise contributions cancel upon summation over nodes, yielding zero net flux.
Consequently, the total mass $\sum_x P_x^s$ evolves solely under the effect of the source term α^s and the decay term ω^s:

$$\frac{d}{dt}\sum_x P_x^s = \sum_x \alpha^s(T_x(t)) \; - \; \omega^s \sum_x P_x^s(t)$$

The mathematical properties are now examined through an in-depth analysis of each phase introduced in section 3.1.

PHASE 0

In the first phase, as described in Section 3.1, Barcelona is modelled as a closed system in which no movement of butterflies between parks is assumed. As a result, the flux term is neglected, and the dynamics are initially reduced to a simplified birth–death process.

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity}$$

This choice is primarily motivated by modelling simplicity, providing a baseline framework before introducing additional levels of complexity. However, this reduction is not expected to fully capture the real dynamics of the system, since butterfly movement between parks is anticipated to play a relevant role. For this reason, the birth–death-only description is expected to be only partially representative, particularly in capturing isolated or weakly connected parks.
The objective of this simplified phase is therefore to establish a reference model and subsequently assess whether the observed dynamics can be adequately explained without inter-park fluxes, before progressively incorporating more realistic interactions.

From a mathematical perspective, further simplifications can be introduced to clarify the structure of the model. In particular, assuming that ω^s does not depend on temperature leads to $\omega^s = \text{const}$, which yields a linear decay term in P.

This results in a clean and well-structured eigenvalue formulation in the multi-park case.

$\longrightarrow$ Discretization

To obtain a numerical formulation of the model suitable for simulations, the continuous-time system is discretized in time. This step is essential in order to approximate the temporal evolution of the population dynamics and to enable a computational implementation of the model over a finite time horizon $[0, \mathcal{T}]$.

- Approximation [EULER]
  The Explicit Euler scheme approximates the time derivative using a forward finite difference, relying only on information available at the current time step. The time interval $[0, \mathcal{T}]$ is discretized with step Δt, yielding:

$$\frac{dP_x^s}{dt} \approx \frac{P_x^s(t + \Delta t) - P_x^s(t)}{\Delta t}$$

  Substituting into the governing equation gives:

$$\frac{P_x^s(t + \Delta t) - P_x^s(t)}{\Delta t} = \alpha^s(T_x(t)) - \omega^s P_x^s(t)$$

  and therefore the update rule:

$$P_x^s(t + \Delta t) = \alpha^s(T_x(t))\,\Delta t + (1 - \omega^s \Delta t)\,P_x^s(t)$$

  Assuming $\Delta t = 1$ day, this simplifies to:

$$P_x^s(t + 1) = \alpha^s(T_x(t)) + (1 - \omega^s) P_x^s(t)$$

  The factor $(1 - \omega^s)$ can be interpreted biologically as a daily survival rate, i.e., the fraction of individuals that remain detectable from one day to the next.
  More generally, the forward Euler method applied to a system $y'(t) = f(t, y(t))$ reads:

$$y_{n+1} = y_n + \Delta t\, f(t_n, y_n)$$

  and is associated with a global truncation error of order $\mathcal{O}(\Delta t)$.
  In terms of stability, the method is conditionally stable. For linear problems, stability requires:

$$\Delta t \leq \frac{2}{|\lambda_{\max}|}$$

For the present system, stability is governed by the decay term. The discrete update factor $(1-\omega^s\Delta t)$ determines whether perturbations are damped over time. The scheme is asymptotically stable if and only if:

$$|1-\omega^s\Delta t|<1 \quad \Longleftrightarrow \quad 0<\omega^s\Delta t<2$$

Setting $\Delta t=1$ yields:

$$0<\omega^s<2$$

a condition typically satisfied in biologically realistic regimes, consistent with the general stability bound $\Delta t \leq 2/\omega^s$.

- Approximation [RUNGE-KUTTA 4]
  To improve accuracy, a higher-order discretization can also be considered. In particular, the fourth-order Runge–Kutta (RK4) method provides a more accurate approximation of the same continuous system over the interval $[0,\mathcal{T}]$ with step Δt. The right-hand side of the system is defined as:

$$f(P_x^s,t)=\alpha^s(T_x(t))-\omega^s P_x^s(t), \qquad P_x^s(0)=0$$

The RK4 scheme evaluates the slope at four intermediate points:

$$k_1=f(P_x^s(t),t) \quad k_2=f\Big(P_x^s(t)+\tfrac{\Delta t}{2}k_1,t+\tfrac{\Delta t}{2}\Big)$$

$$k_3=f\Big(P_x^s(t)+\tfrac{\Delta t}{2}k_2,t+\tfrac{\Delta t}{2}\Big) \quad k_4=f\Big(P_x^s(t)+\Delta t k_3,t+\tfrac{\Delta t}{2}\Big)$$

leading to the update:

$$P_x^s(t+\Delta t)=P_x^s(t)+\frac{\Delta t}{6}(k_1+2k_2+2k_3+k_4)$$

This formula can be interpreted as a numerical approximation of the exact integral form of the solution:

$$P_x^s(t+\Delta t)=P_x^s(t)+\int_t^{t+\Delta t} f(P_x^s,\tau)\,d\tau$$

where the integral is replaced by a weighted average of slopes evaluated at carefully chosen intermediate points. This results in fourth-order accuracy $\mathcal{O}(\Delta t^4)$, significantly improving over the first-order accuracy of the Euler scheme.
From a stability perspective, focusing on the linear decay component $y'=-\omega^s y$, the RK4 update can be written as:

$$P_x^s(t+\Delta t)=R(-\omega^s\Delta t)\,P_x^s(t)$$

where

$$R(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24}$$

Stability requires $|R(z)| < 1$. For real negative arguments $z = -\omega^s \Delta t$, the stability region is approximately $z \in [-2.785, 0]$, leading to:

$$\omega^s \Delta t \leq 2.785$$

For $\Delta t = 1$ day:

$$\omega^s \leq 2.785 \tag{3.5}$$

which is satisfied for the parameter ranges considered in this model.
Compared to the explicit Euler scheme, which requires $\omega^s < 2$, the RK4 method exhibits a larger stability region and improved numerical robustness, making it more suitable for accurate long-term simulations.

$\longrightarrow$ Assumption: T constant

Under the assumption that the temperature T is constant in time, the model reduces to a linear ordinary differential equation (ODE) of the form:

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s}_{\text{Fly-period}} - \underbrace{\omega^s P_x^s(t)}_{\text{Non-activity}}$$

This equation is well-posed in the sense of existence, uniqueness, and continuous dependence on initial data, since it is linear and satisfies the Lipschitz condition. In particular, well-posedness follows directly from the Cauchy–Lipschitz theorem (appendix A.3, theorem 1). Moreover, the solution remains non-negative provided that $\alpha^s > 0$, $\omega^s > 0$, and $P_x^s(0) \geq 0$.
The equation can be rewritten in standard linear form as:

$$\frac{dP_x^s}{dt} + \omega^s P_x^s(t) = \alpha^s$$

$\longrightarrow$ Continuous-time solution.
The solution is obtained by decomposing the problem into homogeneous and particular contributions.

1) Homogeneous solution

The associated homogeneous equation is obtained by setting the source term to zero:

$$\frac{dP_x^{s*}}{dt} + \omega^s P_x^{s*}(t) = 0$$

This separable equation can be integrated as:

$$\frac{dP_x^{s*}}{P_x^{s*}} = -\omega^s dt$$

leading to:

$$P_x^{s*}(t) = Ce^{-\omega^s t}$$

where C is an integration constant determined by the initial condition.

2) Particular solution

For the non-homogeneous equation, since the forcing term α^s is constant, a constant ansatz is considered:

$$P_x^{s**}(t) = Z$$

Substitution into the ODE yields:

$$\omega^s Z = \alpha^s \quad \Rightarrow \quad Z = \frac{\alpha^s}{\omega^s}$$

Thus, the particular solution is:

$$P_x^{s**}(t) = \frac{\alpha^s}{\omega^s}$$

3) General solution

The general solution is given by the sum of homogeneous and particular contributions:

$$P_x^s(t) = Ce^{-\omega^s t} + \frac{\alpha^s}{\omega^s}$$

This solution provides a clear interpretation of the dynamics in continuous time. The term $Ce^{-\omega^s t}$ represents the transient component, which depends on the initial condition and decays exponentially over time. It quantifies how quickly the initial population is forgotten by the system. The term $\frac{\alpha^s}{\omega^s}$ instead represents the stationary equilibrium level, determined by the balance between emergence (α^s) and decay (ω^s).

$\longrightarrow$ Exact discrete-time formulation

The continuous solution can be used to derive an exact discrete-time update. Solving for C from the current state $P_x^s(t)$ gives:

$$C = \left(P_x^s(t) - \frac{\alpha^s}{\omega^s}\right) e^{\omega^s t}$$

Evaluating the solution at time $t+1$ yields:

$$P_x^s(t+1) = \left(P_x^s(t) - \frac{\alpha^s}{\omega^s}\right) e^{-\omega^s} + \frac{\alpha^s}{\omega^s}$$

Rearranging terms leads to the exact update formula:

$$\boxed{P_x^s(t+1) = e^{-\omega^s} P_x^s(t) + \frac{\alpha^s}{\omega^s}\left(1 - e^{-\omega^s}\right)}$$

$\longrightarrow$ Discrete-time interpretation
This expression admits a clear biological interpretation. The population at the next day is composed of two contributions:

- $e^{-\omega^s} P_x^s(t)$: the fraction of individuals that survive from day t to day $t+1$, representing the decay of the existing population through the survival factor $e^{-\omega^s}$;
- $\frac{\alpha^s}{\omega^s}(1 - e^{-\omega^s})$: the contribution of newly emerging individuals during the day, accounting for both their production and the fact that part of them is lost due to decay within the same time interval. This term interpolates between zero emergence and the stationary level $\frac{\alpha^s}{\omega^s}$.

$\longrightarrow$ Equilibria (constant T)
From a biological perspective, after a sufficiently long time the number of adult butterflies observed in each park is expected to approach a constant value. This equilibrium corresponds to the average abundance sustained by the balance between emergence and loss processes.

The equilibrium is obtained by setting the time derivative to zero:

$$\frac{dP_x^s}{dt} = 0 \quad \Longrightarrow \quad 0 = \alpha^s - \omega^s P_x^s$$

Defining $f(P) = \alpha^s - \omega^s P$, equilibrium points satisfy $f(P) = 0$, which yields:

$$P_x^s = \frac{\alpha^s}{\omega^s}$$

Local stability can be assessed through the derivative of f:

$$f'(P) = -\omega^s$$

which is negative if $\omega^s > 0$. In this case the equilibrium is stable, whereas if $\omega^s < 0$ the equilibrium becomes unstable.

$\longrightarrow$ Stability
The explicit solution derived previously,

$$P_x^s(t) = Ce^{-\omega^s t} + \frac{\alpha^s}{\omega^s}$$

provides a direct interpretation of the stability properties. If $\omega^s > 0$, the exponential term decays to zero as $t \to \infty$, implying:

$$P_x^s(t) \longrightarrow \frac{\alpha^s}{\omega^s}$$

and the equilibrium is asymptotically stable. Conversely, if $\omega^s < 0$, the exponential term diverges, leading to instability.
Since $0 < e^{-\omega^s} < 1$ holds only when $\omega^s > 0$, the system is stable precisely under this condition. The equilibrium value

$$P_* = \frac{\alpha^s}{\omega^s} \tag{3.6}$$

represents the steady-state abundance of visible individuals under constant environmental conditions.
From a biological standpoint, ω^s must therefore be positive. A negative value would imply that individuals spontaneously increase in absence of a source, leading to unbounded growth, which is not biologically meaningful.

$\longrightarrow$ Eigenvalues and eigenvectors
The system can also be written in vector form as:

$$\frac{d\mathbf{P}}{dt} = -\omega^s \mathbf{I}\mathbf{P} + \boldsymbol{\alpha}$$

or equivalently:

$$\frac{d\mathbf{P}}{dt} = \mathbf{B}\mathbf{P} + \mathbf{b}, \qquad \mathbf{B} = -\omega^s \mathbf{I}$$

The spectral properties of $\mathbf{B}$ are particularly simple. The eigenvalue problem

$$\mathbf{B}\mathbf{v} = \lambda \mathbf{v}$$

yields:

$$\lambda = -\omega^s$$

Hence, the system has a single eigenvalue whose sign determines stability:

- $\lambda < 0$ (i.e., $\omega^s > 0$): stable dynamics, trajectories decay toward equilibrium;
- $\lambda > 0$ (i.e., $\omega^s < 0$): unstable dynamics, trajectories diverge.

All non-zero vectors are eigenvectors of $\mathbf{B}$, since $\mathbf{B}$ is proportional to the identity matrix. This reflects the absence of spatial coupling: every component evolves independently and identically.

From a dynamical point of view, this implies that there is no diffusion or interaction between parks. Each park follows the same scalar dynamics, and the spatial structure of the system remains unchanged over time.
Biologically, this corresponds to a regime in which parks do not interact: the evolution of the butterfly population in each park is independent, and differences across parks are neither amplified nor smoothed out by the dynamics.

$\longrightarrow$ No more constant T assumption

The model is now extended to the case in which temperature varies in time, leading to the non-autonomous linear ordinary differential equation

$$\frac{dP_x^s}{dt} = \alpha^s(T_x(t)) - \omega^s P_x^s(t)$$

This formulation explicitly accounts for the dependence of the emergence rate on the instantaneous temperature at each park, while the decay term remains linear in the population.
Rewriting the equation in standard linear form gives

$$\frac{dP_x^s}{dt} + \omega^s P_x^s(t) = \alpha^s(T_x(t))$$

To solve the system, an integrating factor is introduced $e^{\omega^s t}$.
Multiplying the equation by the integrating factor transforms the left-hand side into a total derivative

$$e^{\omega^s t}\frac{dP_x^s}{dt} + \omega^s e^{\omega^s t} P_x^s(t) = e^{\omega^s t}\alpha^s(T_x(t))$$

which can be recognized as

$$\frac{d}{dt}\left(e^{\omega^s t} P_x^s(t)\right) = e^{\omega^s t}\alpha^s(T_x(t))$$

Integrating over the time interval $[t, t+1]$ yields

$$e^{\omega^s (t+1)} P_x^s(t+1) - e^{\omega^s t} P_x^s(t) = \int_t^{t+1} e^{\omega^s \tau}\alpha^s(T_x(\tau))\, d\tau$$

and therefore the exact representation of the solution is

$$P_x^s(t+1) = e^{-\omega^s} P_x^s(t) + e^{-\omega^s (t+1)} \int_t^{t+1} e^{\omega^s \tau}\alpha^s(T_x(\tau))\, d\tau$$

This expression shows that the future state is composed of two contributions: the fraction of the current population that survives one time step and a memory term

that aggregates all temperature-driven activity occurring within the interval.
Assuming a daily time resolution, the temperature-dependent forcing is approximated as constant within each time step

$$\alpha^s(T_x(\tau)) \approx \alpha^s(T_x(t))$$

This assumption reflects the fact that observations and environmental inputs are available at a daily scale, so intra-day variability is averaged out.
Under this approximation, the integral becomes

$$\int_t^{t+1} e^{\omega^s \tau} \alpha^s(T_x(\tau))\, d\tau \approx \alpha^s(T_x(t)) \int_t^{t+1} e^{\omega^s \tau}\, d\tau$$

which evaluates to

$$\int_t^{t+1} e^{\omega^s \tau}\, d\tau = \frac{1}{\omega^s}\left(e^{\omega^s (t+1)} - e^{\omega^s t}\right)$$

Substituting this result yields the exact discrete-time update

$$P_x^s(t+1) = e^{-\omega^s} P_x^s(t) + \frac{1 - e^{-\omega^s}}{\omega^s} \alpha^s(T_x(t))$$

This formulation naturally separates the dynamics into two interpretable mechanisms: persistence of previously existing individuals and contribution from newly generated individuals within the time interval.

$\longrightarrow$ Memory interpretation
The update equation can be rewritten in a more compact form as

$$P_x^s(t+1) = \phi P_x^s(t) + \eta \alpha^s(T_x(t))$$

where

$$\phi = e^{-\omega^s} \qquad \eta = \frac{1 - e^{-\omega^s}}{\omega^s}$$

The parameter ϕ represents the survival fraction from one day to the next and therefore encodes the memory of the system. Large values of ϕ correspond to strong persistence, while small values correspond to rapid loss of memory.
The parameter η quantifies the effective contribution of new emergences during a time step, taking into account that part of the newly generated population decays before the next observation.
Iterating the recursion leads to

$$P_x^s(t) = \phi^t P_x^s(0) + \eta \sum_{k=0}^{t-1} \phi^{t-1-k} \alpha^s(T_x(k))$$

This expression shows that the current population depends on all past environmental conditions, with exponentially decaying weights that reduce the influence of older events.
The characteristic decay timescale of the system is given by

$$\tau = \frac{1}{\omega^s}$$

which can be interpreted as the average time an individual remains in the visible state and therefore as the memory horizon of the system.
After a time τ, the fraction of the initial population that remains is

$$e^{-\omega^s \tau} = e^{-1}$$

corresponding to approximately 37%, which provides a direct interpretation of how quickly the system forgets its initial conditions.
The parameter ω^s therefore governs the temporal behaviour of the system by controlling persistence, memory decay, and the smoothness of the response to environmental forcing. Small values correspond to long-lived visibility and strong temporal inertia, while large values lead to rapid decay and weak memory effects. For small ω^s, the approximation $(1-\omega^s) \approx e^{-\omega^s}$ further highlights the connection between discrete and continuous descriptions of the same decay mechanism.

PHASE 1

In this second phase, Barcelona is still modelled as a closed system, so only the interactions between parks within the city are taken into account. However, unlike the previous phase, a flux term between parks is introduced, allowing butterflies to move from one park to another.

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s (W_{yx} P_y^s(t) - W_{xy} P_x^s(t))}_{Flux}$$

In this setting, movement is governed exclusively by euclidean distance (calculated in section 2.3):

$$w_{xy} = f(d_{xy})$$

the probability of migration depends on how far or close two parks are. In addition, each species is characterised by an intrinsic propensity to migrate through μ^s (details in sections 2.2 and 3.3), which modulates its likelihood of movement.

The resulting framework describes a situation in which butterflies make movement decisions based on both their biological tendency to disperse and the spatial configuration of the parks. In particular, a softmax formulation is used to translate these factors into probabilistic flows. This allows the model to capture a more realistic behaviour, where remaining in the same park has no explicit cost, yet migration still occurs due to intrinsic stochasticity in movement decisions.

$\longrightarrow$ Flux structure
The dynamical system introduced in phase 0 is now complemented by an explicit description of spatial interactions between parks. In particular, the flux term encodes the net exchange of individuals through simultaneous immigration and emigration processes, leading to the following contribution in the governing equation:

$$\sum_y \left(\underbrace{\mu^s W_{yx} P_y^s}_{\text{immigration}} - \underbrace{\mu^s W_{xy} P_x^s}_{\text{emigration}} \right)$$

Each day, a fraction of the visible population attempts to move between parks. The destination of each individual is not deterministic but is instead governed by a probabilistic rule based on transition weights w_{xy} (details in section 3.3).
These weights depend on the spatial configuration of the system, so that movement likelihood is primarily determined by euclidean distance between parks (see 3.2). In this way, closer parks are naturally favoured, while distant parks are exponentially suppressed. At the same time, each species is characterised by an intrinsic dispersal propensity μ^s (sections 2.2 and 3.3), which scales the overall intensity of movement. Although distance is symmetric ($d_{xy} = d_{yx}$), the induced transition probabilities are generally not symmetric due to the normalization introduced by the softmax structure. As a consequence, the matrix $\mathbf{W}$ is typically non-symmetric even if the underlying geometry is symmetric.
The transition probabilities are defined through a softmax over movement costs as previously mentioned in section 3.3 and the movement cost is assumed to scale linearly with euclidean distance::

$$\boxed{\begin{cases} cost_{xy} = \delta d_{xy} \\ cost_{xx} = 0 \implies d_{xx} = 0 \end{cases}}$$

$$w_{xy} \propto e^{-cost_{xy}} \underset{cost_{xy}=\delta d_{xy}}{\Longrightarrow} \begin{cases} w_{xy} = \dfrac{e^{-cost_{xy}}}{\sum\limits_z e^{-cost_{xz}}} & y \neq x \\ w_{xx} = \dfrac{e^{-cost_{xx}}}{\sum\limits_z e^{-cost_{xz}}} \underbrace{=}_{cost_{xx}=0} \dfrac{1}{1 + \sum\limits_{z \neq x} e^{-cost_{xz}}} & y = x \end{cases}$$

ensuring that $0 \leq w_{xy} \leq 1$ and that each row of $\mathbf{W}$ sums to one.
This construction guarantees a biologically consistent mechanism in which migration is more likely between nearby parks, while still allowing self-retention as a stochastic outcome of the softmax normalization.

$\longrightarrow$ Discretization
The continuous system is now discretized in time in order to enable numerical simulations. As it was done for phase 0, both explicit Euler and higher-order schemes are considered.

- Approximation [EULER]
  The explicit Euler method approximates the time derivative by a forward finite difference, yielding:

  $$\frac{P_x^s(t+\Delta t) - P_x^s(t)}{\Delta t} = \alpha^s(T_x(t)) - \omega^s P_x^s(t) + \sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))$$

  which leads to the update rule:

  $$P_x^s(t+\Delta t) = P_x^s(t) + \Delta t\Big[\alpha^s(T_x(t)) - \omega^s P_x^s(t) + \sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))\Big]$$

  Assuming $\Delta t = 1$ day:

  $$P_x^s(t+1) = \alpha^s(T_x(t)) + (1-\omega^s)P_x^s(t) + \sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))$$

- Approximation [RUNGE-KUTTA 4 ORDER]
  To improve accuracy, the fourth-order Runge–Kutta method is applied to the full system. Defining the right-hand side as:

  $$f(P_x^s, t) = \alpha^s(T_x(t)) - \omega^s P_x^s(t) + \sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))$$

  the RK4 scheme is given by:

  $$k_1 = f(P_x^s(t), t), \quad k_2 = f\left(P_x^s(t) + \frac{\Delta t}{2}k_1, t + \frac{\Delta t}{2}\right),$$

  $$k_3 = f\left(P_x^s(t) + \frac{\Delta t}{2}k_2, t + \frac{\Delta t}{2}\right), \quad k_4 = f(P_x^s(t) + \Delta t k_3, t + \Delta t)$$

  leading to:

  $$P_x^s(t+\Delta t) = P_x^s(t) + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

Assuming $\Delta t = 1$ day.
For stability analysis, the system can be written in vector form:

$$\frac{d\mathbf{P}}{dt} = \mathbf{BP} + \boldsymbol{\alpha}(T(t))$$

Since $A_{xy} = 1$, the matrix takes the form

$$\mathbf{B} = \mu^s \mathbf{W}^T - (\omega^s + \mu^s)\mathbf{I}$$

The RK4 stability condition is applied to each eigenvalue λ_j^B:

$$R(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24}$$

with $z_j = \lambda_j^B \Delta t$. Stability requires $|R(z_j)| < 1$ for all j, which for real negative eigenvalues implies $|z_j| \leq 2.785$. Since

$$\lambda_j^B = \mu^s \lambda_j^W - (\omega^s + \mu^s)$$

and $\mathbf{W}$ is row-stochastic, the dominant eigenvalue is

$$\lambda_{\max}^B = -\omega^s$$

so the stability condition reduces to

$$\omega^s \leq 2.785 \tag{3.7}$$

for $\Delta t = 1$ day.
Thus, $\omega^s \leq 2.785$ is sufficient in the fully connected case, while in the general case it is necessary but not sufficient.

$\longrightarrow$ Markov chain
The introduction of a Markov chain framework is motivated by the need for a probabilistic description of butterfly movement across a network of parks. Since dispersal is represented as a sequence of discrete relocation events, where individuals move between locations according to transition probabilities, the dynamics naturally admit a formulation as stochastic processes on a graph.
In this setting, each park corresponds to a state of the system, while transitions between states are governed by a matrix $\mathbf{W}$. The resulting dynamics can be interpreted as a random walk, where the future position depends only on the current one, in accordance with the Markov property.
This perspective is particularly useful because it separates the movement component from demographic processes such as birth and death, allowing one to characterise the intrinsic dispersal structure of the landscape and to analyse how spatial

connectivity shapes long-term patterns.
Moreover, the Markov formulation enables the use of spectral tools: the stationary distribution describes the equilibrium spatial occupation in the absence of demographic forcing, while the spectrum of $\mathbf{W}$ quantifies both the rate of convergence to equilibrium and the connectivity of the underlying network.
The matrix $\mathbf{W}$ is row-stochastic, as its entries are non-negative and each row sums to one. In this framework, since the adjacency matrix satisfies $A_{xy} = 1 \ \forall x, y$, the transition structure is given by $W_{xy} = A_{xy} w_{xy} = w_{xy}$. More precisely,

$$W_{xy} \geq 0 \qquad \sum_{y} W_{xy} = 1$$

so that $\mathbf{W}$ defines a valid transition matrix.
Probabilistically, its entries represent transition probabilities:

$$\mathbb{P}(X_{t+1} = y \mid X_t = x) = w_{xy}$$

Hence, butterfly movement across parks is described as a random walk on a graph, where nodes correspond to parks and edges encode admissible transitions. The matrix $\mathbf{W}$ fully determines the dispersal dynamics.
From a dynamical viewpoint, this defines a Markov chain on the network, whose long-term behaviour (in the absence of demographic processes) is described by the stationary distribution, capturing the asymptotic spatial allocation of individuals.

$\longrightarrow$ Steady state
Let π^t denote the distribution of butterflies across parks at time t. The evolution is given by

$$\pi^{t+1} = \pi^t \mathbf{W}$$

or equivalently,

$$\mathbf{P}(t+1) = \mathbf{W}^T \mathbf{P}(t)$$

A steady state is reached when the distribution is invariant under the dynamics:

$$\mathbf{W}\pi = \pi$$

The vector π_x gives the long-term probability (or proportion) of finding a butterfly in park x and corresponds to the left eigenvector of $\mathbf{W}$ associated with eigenvalue $\lambda^W = 1$. Under standard irreducibility assumptions, this is the dominant eigenvalue by the Perron-Frobenius theorem (see appendix A.4, theorem 2).

$\longrightarrow$ Connectivity
Network connectivity can be quantified via the spectral gap. Let λ_2^W denote the second-largest eigenvalue in magnitude; then

$$\gamma = 1 - |\lambda_2^W|.$$

This quantity controls the rate of convergence to equilibrium:

- if $|\lambda_2^W| \approx 1$, the spectral gap is small, indicating weak connectivity and slow mixing;
- if $|\lambda_2^W| \ll 1$, the spectral gap is large, indicating strong connectivity and rapid dispersal.

$\longrightarrow$ Matrix formulation (flux term)
The spatial diffusion component can be written as a net flux between parks, composed of immigration and emigration contributions:

$$\sum_y \left(\mu^s W_{yx} P_y^s - \mu^s W_{xy} P_x^s\right)$$

Factoring out μ^s, this becomes

$$\mu^s \left(\sum_y W_{yx} P_y^s - \sum_y W_{xy} P_x^s\right) = \mu^s \left(\sum_y W_{yx} P_y^s - P_x^s \sum_y W_{xy}\right)$$

Since $\mathbf{W}$ is row-stochastic (i.e. $\sum_y W_{xy} = 1$), the expression simplifies to

$$\mu^s \left(\sum_y W_{yx} P_y^s - P_x^s\right)$$

In matrix form, letting $\mathbf{P} = [P_1^s, \ldots, P_{24}^s]^T$, his yields

$$\mu^s(\mathbf{W}^T\mathbf{P} - \mathbf{P}) = \mu^s(\mathbf{W}^T - \mathbf{I})\mathbf{P}$$

This operator can be interpreted as a graph Laplacian. In standard form, the (weighted) Laplacian is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}^T$$

where $\mathbf{D}$ is the degree matrix, i.e. a diagonal matrix with entries

$$D_{xx} = \sum_y W_{xy}$$

In the present setting, $\mathbf{W}$ is row-stochastic, hence $D_{xx} = 1$ and $\mathbf{D} = \mathbf{I}$.
Therefore,

$$\mathbf{W}^T - \mathbf{I} = -\mathbf{L}$$

which identifies the flux term as a (negative) normalized weighted Laplacian acting on the population vector.

From an ecological perspective, the Laplacian quantifies deviations of population

density in each park from the weighted average of neighbouring parks. Diffusion acts to reduce these spatial heterogeneities, promoting homogenisation across the network, while the parameter μ^s controls the intensity of this mixing process.
Thus, the flux term can be rewritten as

$$\sum_y \mu^s \left(W_{yx}P_y^s - W_{xy}P_x^s\right) = -\mu^s \sum_y L_{xy}P_y^s$$

showing that dispersal is governed by a Laplacian operator on the ecological network.

$\longrightarrow$ Full system dynamics
Let $\mathbf{P} = [P_1^s, \ldots, P_{24}^s]^T$ (population vector of the species s across the 24 parks) and let

$$\boldsymbol{\alpha}(T(t)) = [\alpha^s(T_1(t)), \ldots, \alpha^s(T_{24}(t))]^T$$

In continuous time, the population dynamics can be written as follows. For an undirected, fully connected graph ($A_{xy} = 1\ \forall x, y$):

$$\boxed{\frac{d\mathbf{P}}{dt} = \boldsymbol{\alpha}(T(t)) - \omega^s\mathbf{P}(t) + \mu^s(\mathbf{W}^T - \mathbf{I})\mathbf{P}(t)}$$

$\longrightarrow$ Eigenvalues and eigenvectors

Consider the linear system

$$\frac{d\mathbf{P}}{dt} = \mathbf{B}\mathbf{P} + \mathbf{b} \qquad \mathbf{B}\mathbf{v} = \lambda\mathbf{v}$$

and focus on the spectral properties of $\mathbf{B}$.
In continuous time, for a fully connected graph ($A_{xy} = 1\ \forall x, y$), the dynamics can be written as

$$\frac{d\mathbf{P}}{dt} = \boldsymbol{\alpha}(T(t)) - \omega^s\mathbf{P}(t) + \mu^s(\mathbf{W}^T - \mathbf{I})\mathbf{P}(t)$$

which can be rearranged into the compact form

$$\frac{d\mathbf{P}}{dt} = \mathbf{B}\mathbf{P}(t) + \mathbf{b}(t)$$

with

$$\mathbf{B} = -\omega^s\mathbf{I} + \mu^s(\mathbf{W}^T - \mathbf{I}), \qquad \mathbf{b}(t) = \boldsymbol{\alpha}(T(t))$$

The matrix $\mathbf{B}$ governs the intrinsic dynamics of the system, determining how the current state propagates in time, while $\mathbf{b}(t)$ acts as an external forcing term driven by temperature, continuously injecting new individuals and shifting the system

toward a moving equilibrium.
A more explicit expression for $\mathbf{B}$ is

$$\mathbf{B} = \mu^s \mathbf{W}^T - (\omega^s + \mu^s)\mathbf{I}$$

To understand its dynamics, we study its eigenvalue problem:

$$\mathbf{B}\mathbf{v} = \lambda \mathbf{v}$$

Substituting the expression of $\mathbf{B}$ gives

$$\mu^s \mathbf{W}^T \mathbf{v} - (\omega^s + \mu^s)\mathbf{v} = \lambda \mathbf{v}$$

Since $\mathbf{W}^T$ and $\mathbf{W}$ share the same spectrum, we use $\mathbf{W}^T\mathbf{v} = \lambda^W \mathbf{v}$, which leads to

$$\left(\mu^s \lambda^W - (\omega^s + \mu^s)\right)\mathbf{v} = \lambda \mathbf{v}$$

and therefore

$$\lambda = \mu^s \lambda^W - (\omega^s + \mu^s)$$

In the fully connected case, $\mathbf{W}$ is row-stochastic, so its largest eigenvalue is $\lambda^W_{\max} = 1$ (Perron–Frobenius theorem 2). This implies

$$\lambda_{\max} = \mu^s - (\omega^s + \mu^s) = -\omega^s$$

The associated eigenvector satisfies $\mathbf{W}^T\mathbf{v} = \mathbf{v}$, hence $\mathbf{v} \propto (1, \ldots, 1)^T$. This corresponds to a spatially homogeneous mode, reflecting that diffusion tends to homogenise the population across parks, while mortality induces exponential decay.

$\longrightarrow$ Stability (continuous time)
Stability is fully determined by the sign of the eigenvalues of $\mathbf{B}$. In particular:

- if $\lambda_j < 0$, the corresponding mode decays in time and the system is stable;
- if $\lambda_j > 0$, the mode grows and the system is unstable.

For the dominant eigenvalue,

$$\lambda_{\max} = -\omega^s$$

so the system is stable whenever $\omega^s > 0$, and instability arises only if $\omega^s < 0$.

$\longrightarrow$ Eigenmode decomposition
Although $\mathbf{B}$ is generally non-symmetric (since $\mathbf{W} \neq \mathbf{W}^T$), it is generically diagonalizable, allowing the decomposition

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$$

Each eigenvalue λ_j defines a temporal mode $e^{\lambda_j t}$, so the solution can be expressed as a superposition of spatial patterns:

$$\mathbf{P}(t) = \sum_j c_j e^{\lambda_j t} \mathbf{v}_j + \mathbf{P}^*$$

The eigenvectors $\mathbf{v}_j$ represent characteristic spatial distributions of the population across parks, while the eigenvalues control their decay rates: strongly negative eigenvalues correspond to fast-decaying modes, whereas eigenvalues close to zero produce long-lived spatial structures (memory effects).

$\longrightarrow$ Solution and memory effects
The full system reads

$$\frac{d\mathbf{P}}{dt} = \mathbf{B}\mathbf{P}(t) + \boldsymbol{\alpha}(T(t))$$

with

$$\mathbf{B} = -\omega^s \mathbf{I} + \mu^s(\mathbf{W}^T - \mathrm{diag}(\mathbf{W}\mathbf{1}))$$

The general solution is given by

$$\mathbf{P}(t) = e^{\mathbf{B}t}\mathbf{P}_0 + \int_0^t e^{\mathbf{B}(t-s)}\boldsymbol{\alpha}(T(s))\, ds$$

For constant forcing $\boldsymbol{\alpha}(T(t)) = \boldsymbol{\alpha}_0$, this simplifies to

$$\mathbf{P}(t) = e^{\mathbf{B}t}\mathbf{P}_0 + \mathbf{B}^{-1}(e^{\mathbf{B}t} - \mathbf{I})\boldsymbol{\alpha}_0$$

The dominant eigenvalue $\lambda^B_{\max} = -\omega^s$ sets the global relaxation timescale, with characteristic memory length

$$\tau = \frac{1}{\omega^s}$$

Thus, small mortality leads to slow decay and long memory, while large mortality induces fast relaxation and short memory. The second eigenvalue λ^B_2 controls the persistence of spatial heterogeneities, and the spectral gap quantifies the rate at which these contrasts vanish, distinguishing between fast-mixing and slow-mixing regimes.

PHASE 2

Finally, as anticipated in Section 3.1, the present formulation represents the final and most comprehensive version of the model, designed to provide the most realistic description of the underlying dynamics. This framework integrates all

available information on butterfly behaviour and their movement-related decision-making processes.
In addition to ecological distance between parks, the model incorporates a distance measure weighted by urbanization, which plays a crucial role in shaping dispersal patterns. In heterogeneous landscapes, urban structures can act as significant barriers to movement: for instance, crossing densely built environments, such as major urban corridors like Barcelona's Avinguda Diagonal, may be considerably more restrictive than moving through natural or sparsely urbanised areas, thereby affecting both route selection and transition probabilities.
Within this formulation, urbanization is explicitly embedded in the transition structure through the connectivity term, alongside a further component representing park quality (as defined in section 2.3).
This additional layer captures habitat suitability and introduces a biologically motivated preference mechanism, whereby movement is biased toward higher-quality patches and less favourable locations are progressively avoided.

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))}_{Flux}$$

$\longrightarrow$ Full weights

As in the previous phases and as already introduced in section 3.3, the transition probabilities are constructed through a softmax formulation, which ensures a probabilistic interpretation of movement choices while incorporating both connectivity and habitat quality:

$$w_{xy} = f(C_y, Q_y)$$

where connectivity C_y and quality Q_y jointly determine the attractiveness of destination y.
The associated movement cost is defined as

$$\boxed{\begin{cases} cost_{xy} = -\gamma C_y - \rho Q_y \\ cost_{xx} = -\gamma C_x - \rho Q_x \end{cases}}$$

and the transition probabilities follow a softmax rule:

$$w_{xy} = \frac{e^{-cost_{xy}}}{\sum_z e^{-cost_{xz}}}$$

Explicitly, this gives

$$w_{xy} = \frac{e^{\gamma C_y + \rho Q_y}}{\sum_z e^{\gamma C_z + \rho Q_z}}$$

Similarly to the previous phase, the parameters γ and ρ control the relative importance of connectivity and habitat quality in shaping movement decisions.

$\longrightarrow$ Mathematical structure
The mathematical structure is identical to that of phase 1; the only difference lies in the weight matrix $\mathbf{W}$. All analytical results obtained previously therefore remain valid.
By construction, $\mathbf{w}$ is row-stochastic due to the softmax normalization, i.e. $\sum_y w_{xy} = 1$. Since $A_{xy} = 1$ for all x, y, the matrix $\mathbf{W} = \mathbf{w}$ is likewise row-stochastic, which implies that its largest eigenvalue is $\lambda^W_{max} = 1$ (Perron-Frobenius theorem), with associated eigenvector $\mathbf{v}_{max} = (1, \ldots, 1)^T$. Consequently, the largest eigenvalue of the matrix $\mathbf{B}$ is simply

$$\lambda_{max}(\mathbf{B}) = -\omega^s$$

The solution of the system retains the same form derived earlier:

$$\mathbf{P}(t) = e^{\mathbf{B}t}\mathbf{P}_0 + \int_0^t e^{\mathbf{B}(t-s)}\boldsymbol{\alpha}(T(s))\, ds$$

and the same holds for the stability of the system: since $\lambda_{max} = -\omega^s < 0$, the system remains stable for any $\omega^s > 0$. When $\mathbf{B}$ is diagonalizable (see below), the modal decomposition is also unchanged:

$$\mathbf{P}(t) = \sum_j c_j e^{\lambda_j t}\mathbf{v}_j + \mathbf{P}^* \qquad \mathbf{P}^* = -\mathbf{B}^{-1}\boldsymbol{\alpha}_0$$

Finally, the relation between the eigenvalues of $\mathbf{B}$ and those of $\mathbf{W}$,

$$\lambda = \mu^s \lambda^W - (\omega^s + \mu^s)$$

continues to hold in this case. The eigenvalues and eigenvectors λ_k, $\mathbf{v}_k$ for $k \neq 1$ must still be computed numerically from the full weight matrix.

Regarding the differences with respect to Phase 1, it is worth noting that $\mathbf{B}$ was already non-symmetric in that case, since $\mathbf{W} \neq \mathbf{W}^T$ even under a purely distance-based formulation. Here, however, this asymmetry becomes more pronounced: since the cost function depends only on the connectivity and quality of the destination node y, the resulting transition matrix $\mathbf{W}$ deviates further from symmetry than in the distance-based case. As in Phase 1, $\mathbf{B}$ remains diagonalizable whenever $\mathbf{W}^T$ has distinct eigenvalues, which is the generic case.
Since $\mathbf{B}$ is not symmetric, its eigenvalues could in principle be complex; however, for biologically realistic parameters they are expected to be real and negative, consistent with the dominant-mode behavior already established for $\lambda_{max} = -\omega^s$.

3.4.2 Trap index

As previously underlined in sections 1.3 and 3.3, the main focus is the flux term of the equation, as it provides key insight into both butterfly dispersal dynamics and the quality of the surrounding habitats. In this sense, analysing the flow component is essential for understanding how movement patterns are shaped by environmental conditions. This perspective is particularly relevant as it directly contributes to achieving the objectives of the project, as outlined in Section 1.3, by linking movement dynamics with habitat structure and ecological interpretation. Consequently, in order to identify ecological traps, namely habitats that attract individuals despite offering lower fitness or survival prospects (details in section 1.1), it is necessary to introduce additional concepts and modelling tools.

Therefore, a flow index is introduced to quantify the net movement of individuals across parks:

$$J_x^s(t) = \sum_y \mu^s \left(W_{yx} P_y^s - W_{xy} P_x^s \right)$$

The two contributions of this expression can be interpreted as follows:

$$\sum_y W_{yx} P_y^s(t) = I_x^s(t) \qquad \sum_y W_{xy} P_x^s(t) = O_x^s(t)$$

where $I_x^s(t)$ represents the total immigration into patch x, while $O_x^s(t)$ corresponds to the total emigration from x.
This allows the net flux to be rewritten in the form:

$$J_x^s(t) = I_x^s(t) - O_x^s(t)$$

so that

$$\begin{cases} J_x^s(t) > 0 \;\Rightarrow\; \text{net inflow (immigration dominates)} \\ J_x^s(t) < 0 \;\Rightarrow\; \text{net outflow (emigration dominates)} \end{cases}$$

Within this framework, habitats can be classified according to both demographic performance and movement behaviour:

$$\begin{cases} \text{SOURCE} \;\Rightarrow\; \text{net export of individuals and high local fitness} \\ \text{SINK} \;\Rightarrow\; \text{net import of individuals but low local fitness} \\ \text{TRAP} \;\Rightarrow\; \text{sink dynamics combined with high attractiveness (high } w_{xy}) \end{cases}$$

As discussed in section 1.1, ecological traps correspond to patches that act as sinks while simultaneously exhibiting strong attractiveness. At equilibrium, denoted by P^*, the stationary flux is given by:

$$J_x^* = \sum_y \mu^s \left(W_{yx} P_y^* - W_{xy} P_x^* \right)$$

which leads to the classification:

$$\begin{cases} J_x^* > 0 & \Rightarrow \text{ sink-like behaviour} \\ J_x^* < 0 & \Rightarrow \text{ source-like behaviour} \end{cases}$$

To further disentangle local demographic effects from movement-driven processes, a local persistence index is defined:

$$L_x^s(t) = \alpha_x^s(T_x(t)) - \omega^s P_x^s(t)$$

This quantity measures the intrinsic ability of a habitat patch to sustain its population independently of immigration, and can therefore be interpreted as a proxy for local ecological suitability:

$$\begin{cases} L_x^s(t) > 0 & \Rightarrow \text{ locally self-sustaining population} \\ L_x^s(t) < 0 & \Rightarrow \text{ local decline in absence of immigration} \end{cases}$$

Complementarily, an immigration dependence index is introduced:

$$D_x^s(t) = \frac{I_x^s(t)}{I_x^s(t) + \alpha_x^s(T_x(t)) + \epsilon}$$

which quantifies the extent to which the observed population is sustained by external inflow. In particular,

$$\begin{cases} D_x^s(t) \approx 0 \Rightarrow \text{locally driven dynamics} \\ D_x^s(t) \approx 1 \Rightarrow \text{immigration-driven population} \end{cases}$$

A sink-type indicator can then be defined as:

$$\text{SINK}_x^s(t) = D_x^s(t)\, \mathbb{1}[L_x^s(t) < 0]$$

highlighting habitats where negative local dynamics are compensated by strong immigration.
To account for temporal variability, a seasonal average is considered:

$$\text{SINK}_x^s = \frac{1}{T} \sum_{t=1}^{T} D_x^s(t)\, \mathbb{1}[L_x^s(t) < 0]$$

Finally, to incorporate the severity of local deficit, a weighted trap index is introduced:

$$\text{SINKSEVERITY}_x^s(t) = D_x^s(t) \frac{\max(0, -L_x^s(t))}{\omega^s P_x^s(t) + \epsilon}$$

with corresponding temporal average:

$$\frac{1}{T}\sum_{t=1}^{T}\text{SINKSEVERITY}_x^s(t)$$

Overall, ecological classification can be summarised as:

$$\begin{cases}\text{SOURCE} \Rightarrow L_x^s(t) > 0 \text{ and low immigration dependence}\\ \text{SINK} \Rightarrow L_x^s(t) < 0 \text{ and high immigration dependence}\\ \text{TRAP} \Rightarrow \text{sink behaviour combined with high attractiveness}\end{cases}$$

To characterise attraction explicitly, the incoming connectivity is considered through:

$$\overline{W}_x = \frac{1}{n-1}\sum_{y\neq x} W_{yx}$$

Leading to the final trap index:

$$\text{TRAP}_x^s = \overline{D_x^s(t)} \cdot \frac{\overline{\max(0, -L_x^s(t))}}{\omega^s P_x^s(t) + \epsilon} \cdot \text{Attraction}_x$$

where

$$\text{Attraction}_x = \frac{\sum_y w_{yx}}{\max_z \sum_y w_{yz}}.$$

This formulation captures ecological traps as habitats combining strong attractiveness, high immigration dependence, and poor local demographic performance.

3.5 Summary

The mathematical model presented in Chapter 3 originates from the need to describe, in a tractable and biologically interpretable way, how the observable adult population of a given butterfly species changes over time and space within the urban park network of Barcelona. This requirement naturally led to an equation

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s(W_{yx}P_y^s(t) - W_{xy}P_x^s(t))}_{Flux}$$

composed of three main terms, each corresponding to a distinct ecological process: a *fly-period* term, representing the temperature-driven emergence of new observable adults; a *non-activity* term, accounting for the loss of individuals from the visible population through mortality, transition to unobserved stages, or non-detection; and a *flux* term, describing the movement of individuals between parks through the dispersal network. Within this flux term, the only unknown parameters are δ, γ, and ρ, which are positive parameters that control the relative contribution of distance, connectivity and habitat quality to the movement decision; since no direct empirical data are available to determine their values, these parameters are estimated by the model through the fitting procedure, in the same way as α^s and ω^s. Starting from this equation, the model is developed incrementally across three phases of increasing complexity (Phase 0, Phase 1, and Phase 2, as introduced in section 3.1), each phase refining the representation of the flux term while leaving the local demographic structure of the equation unchanged.

Among the parameters entering these three terms, particular attention must be given to α^s and ω^s, since no direct empirical data are available to estimate either the emergence rate or the non-activity rate: these quantities must therefore be inferred indirectly, by fitting the model to the observed abundance data. As a consequence, both parameters require carefully justified admissible ranges, derived from biological reasoning and numerical stability considerations, within which the optimisation procedure is allowed to search for plausible values. These bounds, summarised below, will be implemented directly as constraints in the fitting routines described in the following section 4.

- The emergence term $\alpha^s(T_x(t))$ is modelled as a Gaussian function of temperature and is governed by three parameters: the maximum emergence rate a, the optimal temperature T^*, and the thermal breadth σ_T. The amplitude a is bounded above by $\omega^s \cdot P^s_{\max}$, where $P^s_{\max}$ is the maximum observed abundance of species s; this ensures that the equilibrium population implied by the model, α^s/ω^s, remains consistent with realistic abundance levels rather

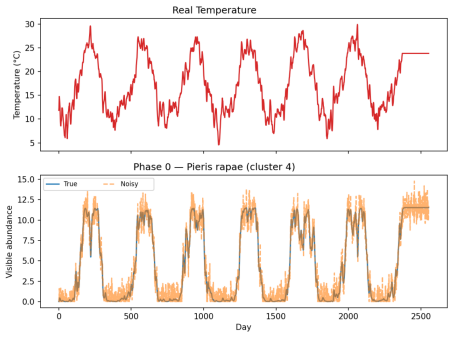
than diverging toward biologically implausible values.
Table 3.1 reports the resulting species-specific bounds $a_{\max} = 2.785 \cdot P^s_{\max}$ for all 44 species considered in the analysis.

Species	$P^s_{\max}$	$a_{\max}$
Leptotes pirithous	120	360
Pieris rapae	72	216
Polyommatus icarus	68	204
Pieris brassicae	46	138
Pararge aegeria	44	132
Lasiommata megera	41	123
Cacyreus marshalli	34	102
Lampides boeticus	34	102
Gonepteryx cleopatra	32	96
Iphiclides feisthamelii	26	78
Vanessa cardui	26	78
Pyronia cecilia	24	72
Colias crocea	20	60
Papilio machaon	17	51
Vanessa atalanta	17	51
Pontia daplidice	16	48
Euchloe crameri	10	30
Carcharodus alceae	10	30
Celastrina argiolus	9	27
Thymelicus acteon	7	21
Pyronia bathseba	7	21
Gonepteryx rhamni	7	21
Charaxes jasius	6	18
Aricia cramera	6	18
Lycaena phlaeas	5	15
Hipparchia statilinus	5	15
Favonius quercus	5	15
Satyrium esculi	5	15
Maniola jurtina	5	15
Libythea celtis	5	15
Pyrgus malvoides	4	12
Hipparchia fidia	4	12
Melanargia lachesis	3	9
Brintesia circe	2	6
Limenitis reducta	2	6
Ochlodes sylvanus	2	6
Euphydryas aurinia	1	3
Danaus chrysippus	1	3
Apatura ilia	1	3
Anthocharis cardamines	1	3
Melitaea didyma	1	3
Hipparchia semele	1	3
Polygonia c-album	1	3
Pieris napi	1	3

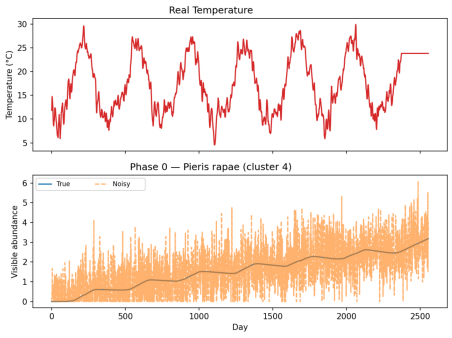
Table 3.1: Species-specific maximum observed abundance $P^s_{\max}$ and corresponding upper bound on the emergence amplitude, $a_{\max} = 2.785 \cdot P^s_{\max}$.

The thermal breadth is constrained to $\sigma_T \in [0.5,\ 10]$, while the optimal temperature is restricted to $T^* \in [15°\text{C},\ 30°\text{C}]$, in line with the physiological range over which butterfly activity is expected to occur.

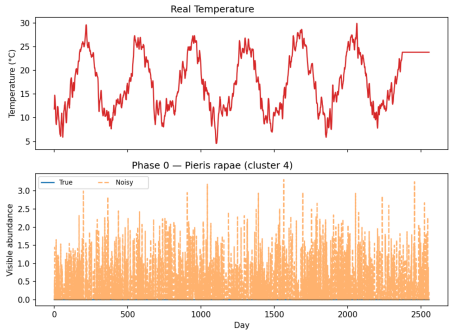
- The non-activity rate ω^s is bounded by the RK4 stability condition (see 3.5, 3.6 and 3.7): as $\omega^s \to 2.785$, the discretisation becomes progressively unstable, as shown in figure 3.2, where synthetic simulations reveal a clear deterioration of the model's behaviour near this threshold.



(a) $\omega^s = 2.6$



(b) $\omega^s = 2.785$



(c) $\omega^s = 3$

Figure 3.2: Synthetic-data demonstration with phase 0 of the upper bound on ω^s. As ω^s approaches the stability threshold of 2.785, the model's ability to reproduce a biologically meaningful trajectory degrades visibly; beyond this value, the numerical scheme becomes unstable.

3.5.1 Constraints

Building on the parameters and bounds introduced above, this subsection brings together, in a single coherent overview, all the boundary and parameter constraints adopted across the model.

- Population.
  The population of visible adults is required to remain non-negative throughout, $P_x^s(t) \geq 0$, with initial condition $P_x^s(0) = 0$ for every species and park, reflecting the assumption that no individuals are observed at the start of the simulated period.

- Phenological window.
  Detection records follow a characteristic seasonal bell-shaped pattern (the so-called *campana*), with essentially no detections occurring outside the March - October window. Rather than imposing an abrupt cutoff, this seasonality is captured smoothly through the emergence term α^s itself, avoiding a sharp interruption in activity.

- Temperature.
  Although no data exist to determine the precise thermal threshold below which eggs or larvae fail to survive, detectability is nonetheless bounded empirically:
  - above $T_{\max} = 35°\text{C}$: no detection (too hot);
  - below $T_{\min} = 13°\text{C}$: no detection (too cold).

- Fly-period rate α^s.
  - $a \geq 0$, bounded above by $\omega^s \cdot P^s_{\max}$;
  - $T^* \geq 0$, restricted to $15°\text{C} \leq T^* \leq 30°\text{C}$;
  - $\sigma_T \geq 0$, bounded as $0 < \sigma_T \leq 10$.

- Non-activity rate ω^s.
  Governed by the Runge–Kutta stability condition (Eqs. 3.5, 3.6, 3.7):
  $$0 < \omega^s < 2.785$$

- Mobility coefficient.
  Fixed rather than estimated, μ^s varies within the normalised interval [0,1] according to the cluster to which each species belongs (details in 2.3, 3.3):
  $$\mu^s = \begin{cases} 0.25 & s \in \text{Cluster 1} \\ 0.50 & s \in \text{Cluster 2} \\ 0.75 & s \in \text{Cluster 3} \\ 1.00 & s \in \text{Cluster 4} \end{cases}$$

- Dispersal weights.
  The weighted adjacency matrix $W_{xy} = A_{xy} \cdot w_{xy}$ satisfies $W_{xy} \geq 0$ with, in general, $W_{xy} \neq W_{yx}$, reflecting the directed nature of the dispersal network.
  - Adjacency: $A_{xy} \in \{0,1\}$, fixed to all ones at this stage;
  - Edge weights: $w_{xy} \geq 0$, $\sum_y w_{xy} = 1$, hence $0 \leq w_{xy} \leq 1$;
  - Distance: $d_{xy} \in [0,10000]\,\mathrm{m}$ (euclidean distance between park centroids 2.4);
  - Connectivity: $0 \leq C_y \leq 150$, within a 500 m radius (2.2);
  - Quality: $0 \leq Q_y \leq 3$ (2.1).
- Flux shape parameters δ, γ, ρ.
  These are the only unknown parameters governing the flux term, controlling respectively the relative contribution of distance, connectivity and habitat quality to the movement decision encoded in w_{xy}. As no direct empirical data are available to fix their values, δ, γ and ρ are estimated by the model through the fitting procedure, subject to the positivity constraint:
  - $0 < \delta < 10$
  - $0 < \gamma < 10$
  - $0 < \rho < 10$

It is worth noting that, for the parameters δ, γ, ρ and σ_T, mathematical analysis only established the requirement that they be positive. The corresponding upper bounds were therefore determined empirically through an iterative procedure, by exploring a range of plausible parameter values and evaluating their behaviour through numerical simulations.

Chapter 4

Implementation

The data available for this study consist of the observed (or neural network smoothed 2.5) daily butterfly count time series for each park, which represent the empirical counterpart of the aggregate output predicted by the model in equation 3.4. While this equation specifies how the three terms it comprises jointly determine the predicted abundance, the relative contribution of each term to the resulting dynamics cannot be inferred analytically and must instead be estimated from data.
This is precisely the role of the implementation and optimization procedure described in this chapter: it provides the computational means to translate the theoretical formulation of section 3.4 into a model whose parameters are calibrated against observation, thereby allowing the disentangling of the contribution of each term to the overall flux.

In general terms, optimization consists in identifying the set of parameter values for which the model output best approximates the observed data, according to a chosen loss function; in the present case, this amounts to finding the parameters under which the simulated abundance most closely reproduces the empirical butterfly count data.
As previously discussed, the primary interest of this work lies in characterizing the flux of individuals between parks; accordingly, the fitting procedure also yields estimates of the flight-period and activity terms, which are not ends in themselves but constitute the necessary intermediate quantities for the flux analysis presented in section 3.4.2.

4.1 Python

The entire model simulation, fitting, and diagnostic pipeline was implemented in Python [46]. Data analysis was performed in Python using NumPy [47], SciPy [48], and pandas [49], and figures were generated using Matplotlib [50].
The full source code is openly available in the project's GitHub repository: https://github.com/gioda/farfalle.git.

The implementation follows the three phases (no dispersal, distance-based dispersal, connectivity/quality-based dispersal) introduced in section 3.4, and is organized into a simulation module, a fitting module, and a data-preprocessing module. The three phases share the same temperature-driven activation term and the same numerical integration scheme, and are implemented as methods of a single phase class in order to avoid code duplication and to ensure consistency across phases.

Discretization.

The ODEs were integrated numerically using the fourth-order Runge–Kutta method (RK4 3.4) at a daily time step ($dt = 1$ day), rather than a simpler forward Euler scheme, in order to keep numerical errors negligible over multi-year simulations. The daily temperature forcing was held constant within each integration step.

Fitting and optimization.

Model parameters were estimated by fitting the simulated visible abundance to the neural network predicted daily abundance for each park. To avoid that high-frequency, day-to-day fluctuations in the temperature forcing (and in the NN-predicted target) translated into spurious fluctuations in the simulated dynamics, a 7-day rolling mean was applied to both series prior to fitting. The species-specific maximum activation rate $a_{\max}$ and mobility parameter μ^s were not free-fitted but loaded from a precomputed lookup table, so that the upper bound on a and the value of μ^s were set dynamically per species before optimization.

Parameters were estimated using a normalized MSE loss,

$$\mathcal{L} = \frac{1}{TS} \sum_{t,x} \left(\frac{P_{t,x}^{\text{model}} - P_{t,x}^{\text{NN}}}{\max_t \left(P_{t,x}^{\text{NN}} \right) + \epsilon_x} \right)^2$$

where $P_{t,x}^{\text{model}}$ and $P_{t,x}^{\text{NN}}$ are, respectively, the model prediction and the neural network target, T and S are the number of time steps and spatial points, and $\epsilon_x > 0$ is a small stabilizing constant preventing division by zero.

The loss was minimized with the L-BFGS-B algorithm (`scipy.optimize.minimize` [48]), a quasi-Newton method suited to large-scale problems with box constraints (details in appendix A, A.5), which were imposed on every parameter according to biologically plausible ranges (e.g. see 3.5.1).
Since L-BFGS-B is a local optimizer and the loss surface is not guaranteed to be convex, a multi-start scheme was used: $N_{\text{starts}} = 100$ randomly sampled initial parameter vectors (uniform within bounds, fixed seed) were used in addition to one biologically motivated starting guess, and the run with the lowest loss was retained as the best fit, while the full distribution of per-start losses and estimates was stored to assess parameter identifiability (section 5).

Chapter 5

Results and Discussion

This section presents the results obtained from the implementation and parameter estimation procedures described in chapter 4. In particular, it includes an assessment of the quality of the model fitting and the corresponding flow analyses, with the aim of evaluating both the predictive performance of the proposed framework and its ability to characterize the functional role of urban parks.

Specifically, the representativeness of the model described in chapter 3 with respect to the neural network benchmark (2.5) is further analyzed by evaluating how accurately each model phase reproduces the observed dynamics. For each phase and each site x, it was verified that the model adequately explored the full range of parameter values, both within their initial bounds and in the optimized configuration, ensuring a comprehensive exploration of the parameter space. The goodness of fit was quantified using the normalized mean squared error (nMSE). While the MSE measures the absolute deviation between the ODE trajectory and the neural network predictions in units of squared abundance, the nMSE rescales these residuals by the maximum abundance observed at each site, enabling scale-independent comparisons across parks characterized by highly heterogeneous population sizes (details in appendix A, A.6). This metric corresponds to the objective function minimized during the parameter optimization procedure.
The fitting assessment was complemented by additional isolation-based metrics (see sections 2.3, table 2.7 and figure 2.18) to investigate how the different model phases perform across parks exhibiting different degrees of spatial isolation. Only the most informative indicators are reported.
Finally, a flux-based analysis (section 3.4.2) was carried out, representing the primary objective of this work. Its purpose is to evaluate whether the proposed model can provide qualitative information on the ecological role of urban parks, namely whether they function as sources, sinks, or ecological traps within the urban landscape. Phase 2 is the result of a process of improving the model by adding essential

ecological elements at each step and is the one that best represents the whole, albeit with certain limitations, showing the best agreement with the neural network benchmark and the observed dynamics. Therefore, this flux-based analysis was performed only for the Phase 2 model.

Of the 44 species, the analysis focuses exclusively on the species *Pieris rapae* (fig. 2.4), selected for the preliminary investigation due to its high abundance, generalist ecological behaviour, and strong dispersal capability. In particular, this species was identified in consultation with ecologists as a suitable model organism precisely because its high abundance ensures a large and well-represented dataset across all sampled sites, making the observational data more robust and reliable. These characteristics make it a suitable proxy for investigating metapopulation dynamics and habitat connectivity across the network of 24 parks in Barcelona, where the species is present at all sampled locations (see details in 2.2). Despite these advantages, the choice of this species also introduces certain assumptions in the interpretation of flux-based analyses, which will be discussed in detail in chapter 6. Nevertheless, it was decided to proceed with this species. More specialized species were excluded because the available observational data were limited and less reliable, and their low abundance in sampling counts would have led to a weaker and potentially less representative reconstruction of the ecological network. Further considerations on this issue are presented in section 6.2 of chapter 6.

The results presented in this chapter will be further interpreted in the following chapter 6. For reasons of space and readability, not all analyses could be included here; additional results and supplementary figures are therefore provided in the appendix B for completeness.

5.1 Phase 0

Phase 0 corresponds to the stage in which only the source and output dynamics are represented (details in 3.4.1). In this phase, the model equation

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity}$$

was implemented and fitted against the neural network data (2.5), yielding the parameter estimates summarized in figure 5.1, which displays the frequency distribution of the parameter values obtained across 100 independent optimization runs.

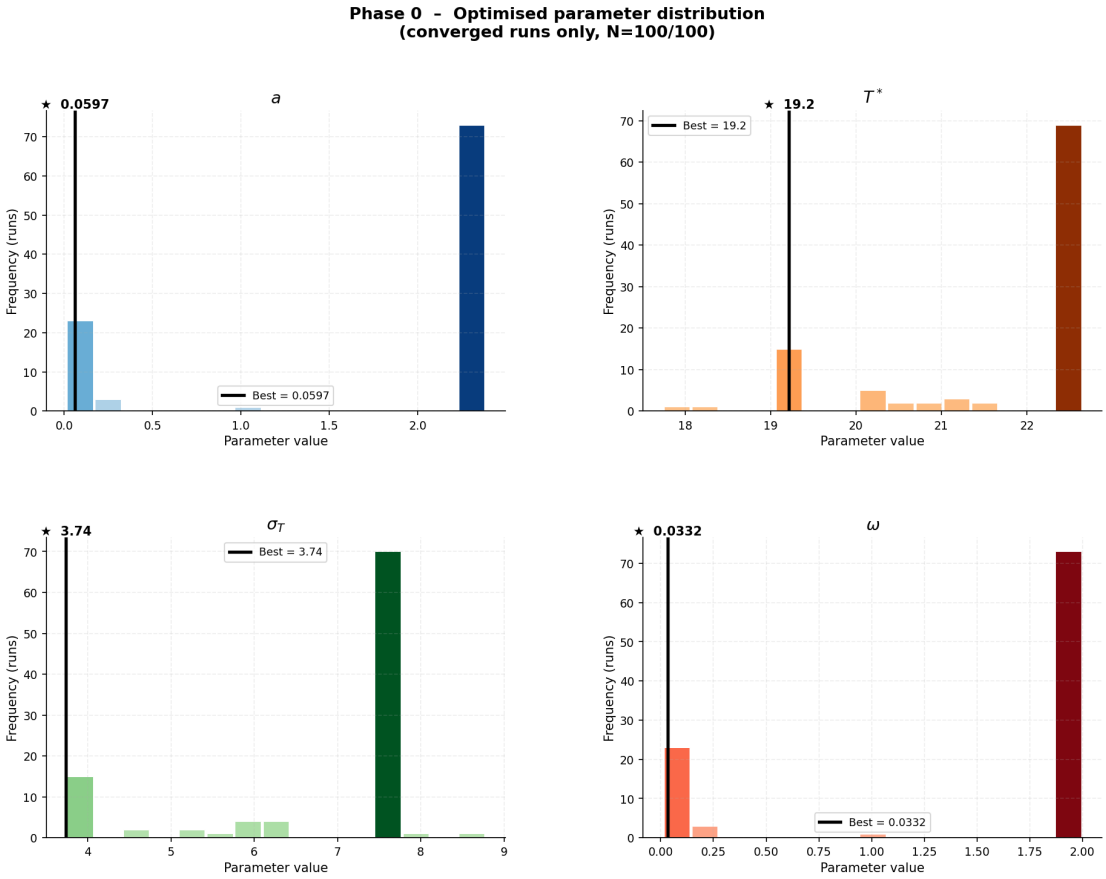


Figure 5.1: Frequency distribution of the Phase 0 parameter estimates obtained from the optimization procedure over 100 independent runs (the marked value corresponds to the solution producing the lowest loss).

The complete set of initial and optimized parameter values for each of the 100 runs is reported in appendix B, B.1, together with a histogram showing the distribution of the initial parameters, analogous to the representation in figure 5.1.
The optimal parameter values obtained from this optimization are reported below:

- a [maximum emergence]: 0.0597 $\Longrightarrow$ very low maximum daily emergence.
- T^* [optimal temperature]: 19.2 $°C$ $\Longrightarrow$ biologically reasonable for Pieris rapae.
- σ_T [thermal breadth]: 3.74 $\Longrightarrow$ the species is active between $\approx 11°\text{C} - 27°\text{C}$.
- ω^s [non-acivity rate]: 0.0332 $\Longrightarrow$ average observable time of ≈ 30 days.

Those marked value identifies the best-performing estimate, corresponding to the minimum loss of 0.0272 (details in figure 5.2).

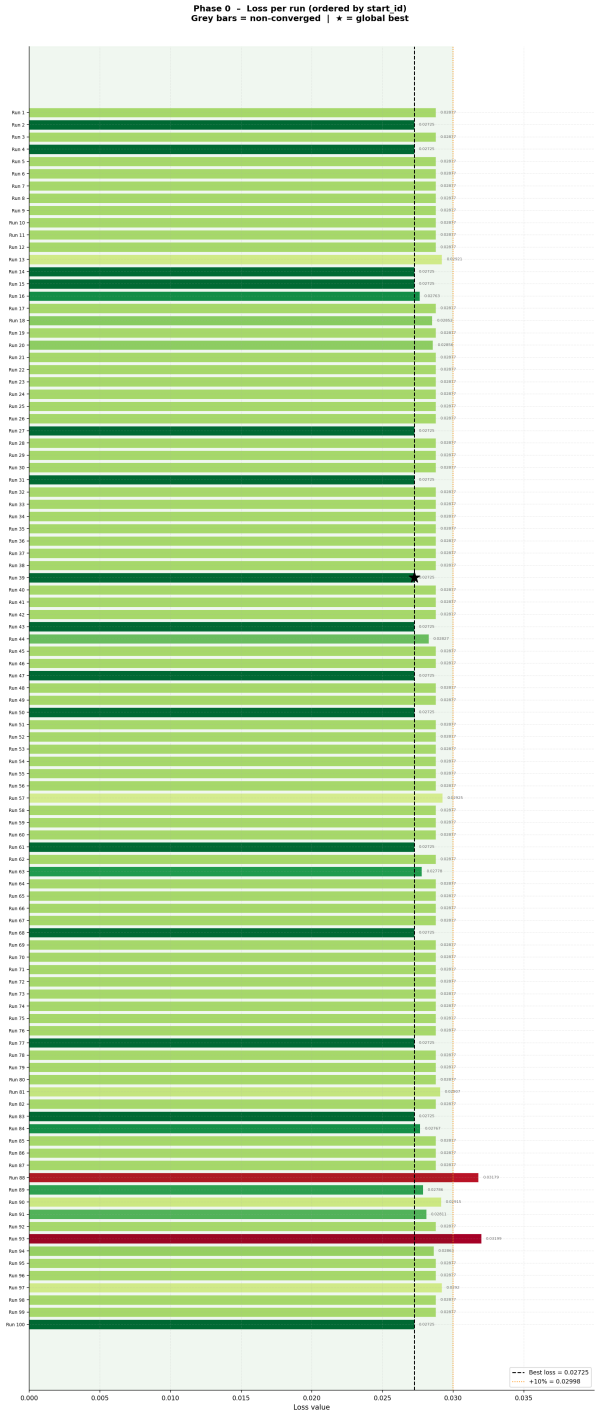


Figure 5.2: Phase 0 loss values obtained from the optimization procedure over the 100 independent runs.

Using these estimated parameters, the model reproduces the neural network output as illustrated in figure 5.3

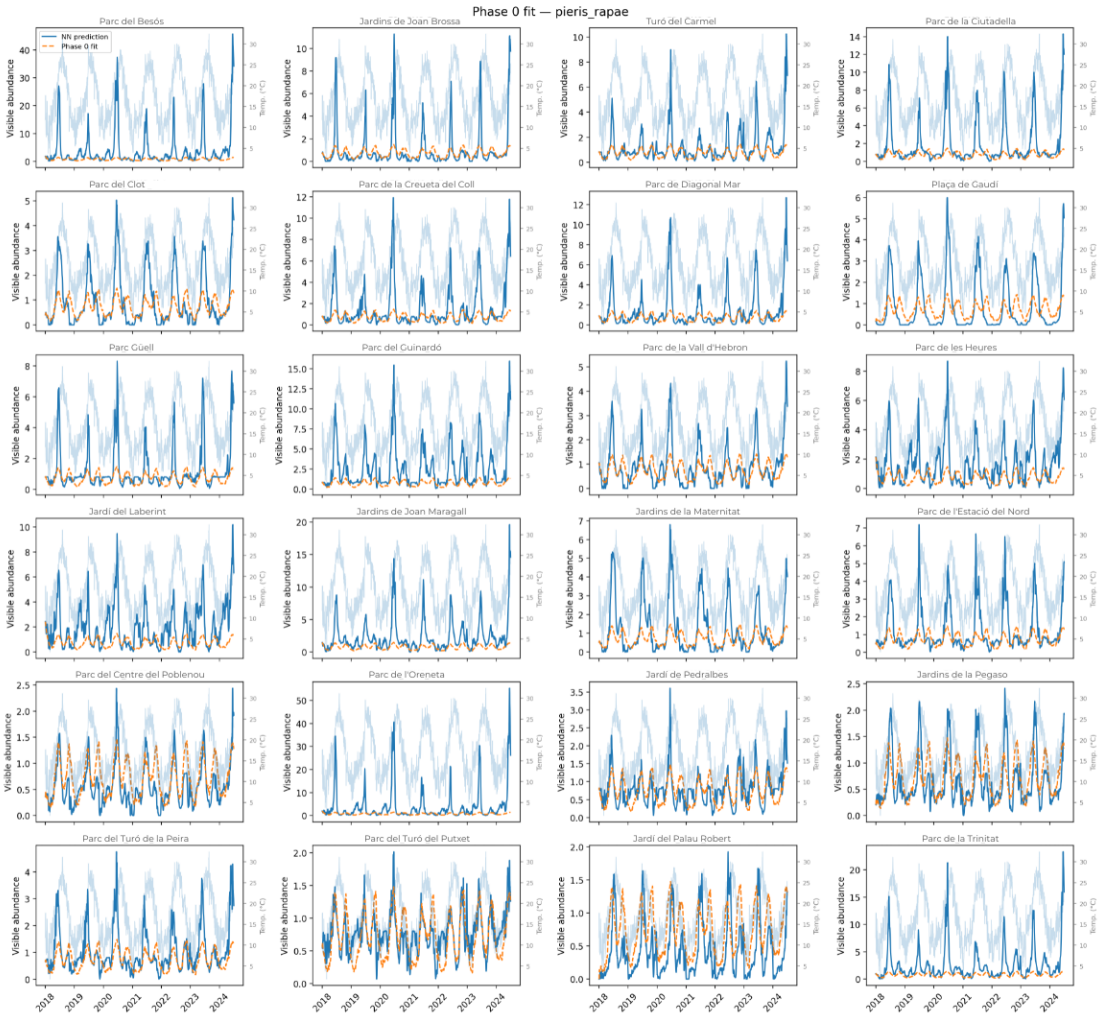


Figure 5.3: Phase 0 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile.

To provide a more detailed assessment of the fitting quality, the nMSE-based analysis introduced earlier (A.6) is reported for each park, allowing a quantitative evaluation of the agreement between the model predictions and the observed dynamics.

The corresponding results are presented below in the figure 5.4.

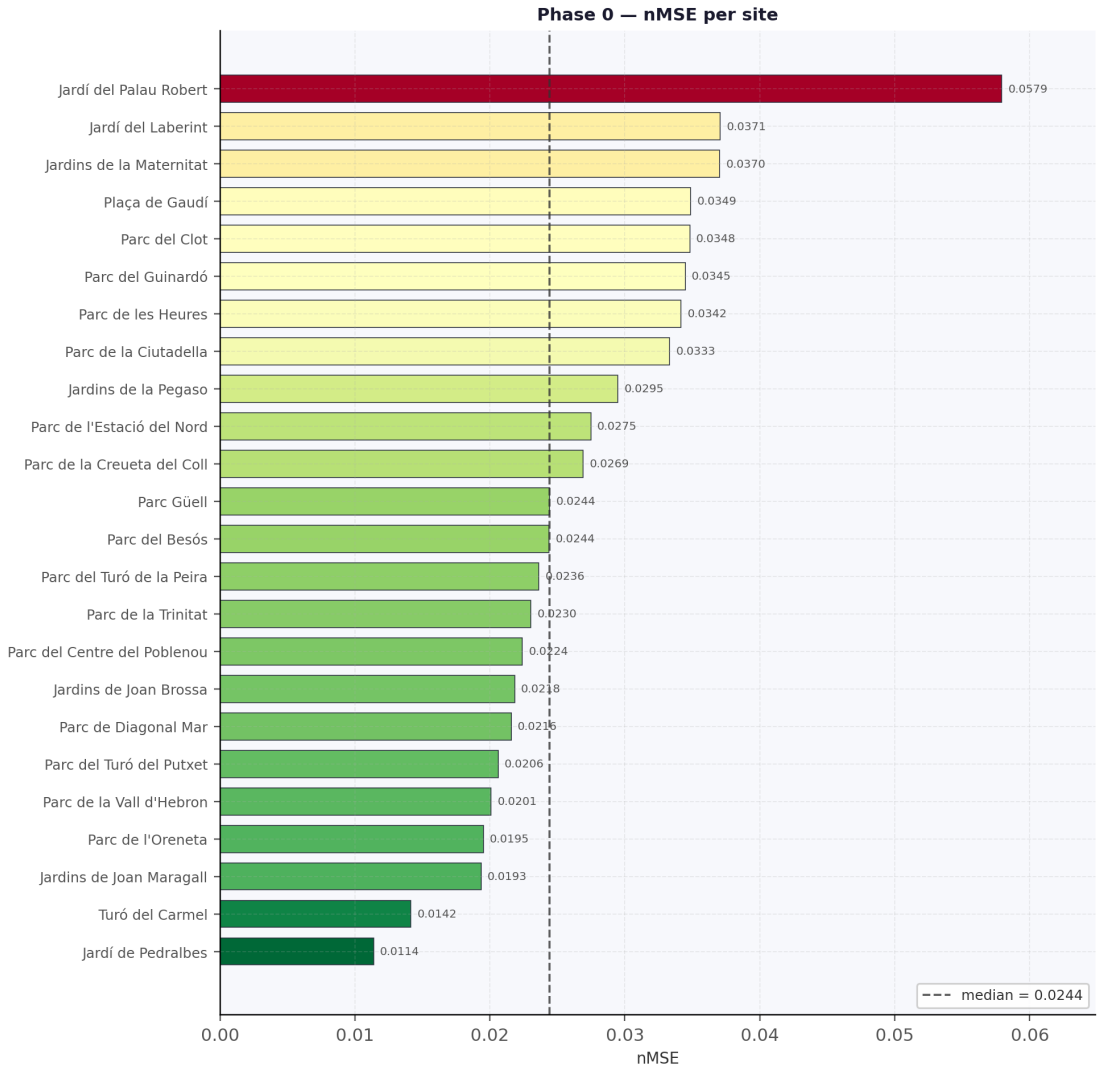


Figure 5.4: Fit quality with mse phase 0.

It can be observed that the park Jardí del Palau Robert yields the lowest quality of fit, while the park Jardí de Pedralbes achieves the best agreement between model predictions and observed data, as quantified by the nMSE-based evaluation introduced in the previous sections.
This variation in performance across parks suggests a non-trivial dependence of the model accuracy on local spatial and ecological characteristics, rather than a uniform behaviour across the entire urban network.

More specifically, Figure 5.5 presents representative examples of Jardí del Laberint and Parc de Diagonal Mar, respectively, enabling a closer inspection of the Phase

0 model behavior over the neural network in two specific parks. The comparison between these two cases further emphasizes how strongly model performance can vary depending on the specific context of each park.

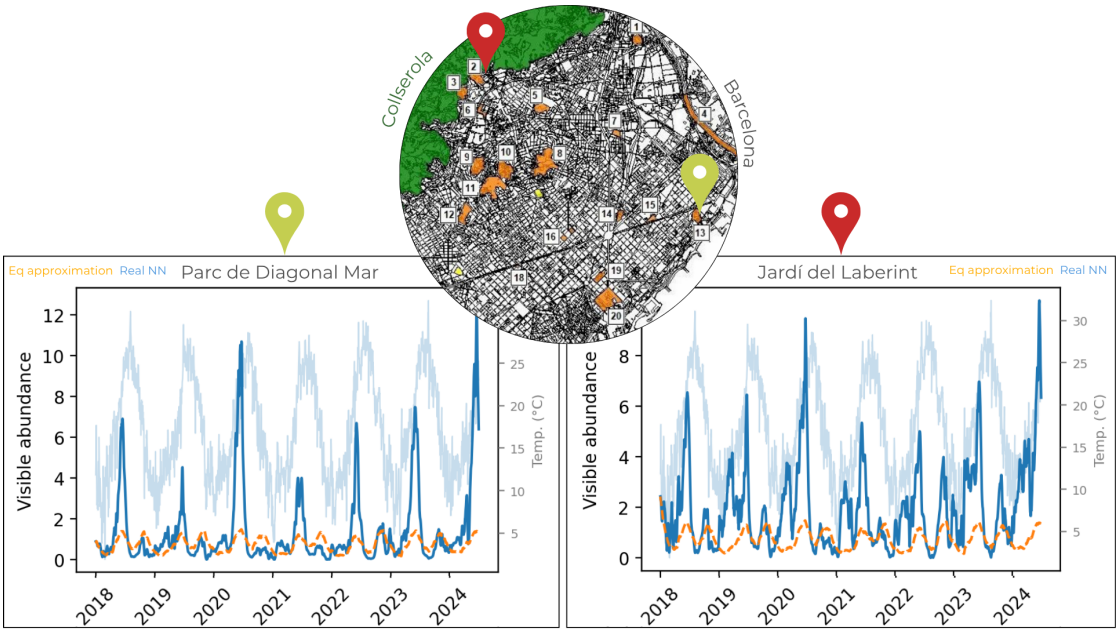


Figure 5.5: Phase 0. On the left, Jardí del Laberint is shown, a coastal park in Barcelona. On the right, Parc de Diagonal Mar is displayed; as indicated by the inset map, it is located at the northern edge of Barcelona, near the boundary with Collserola Park.

These results highlight the dependence of model performance on spatial context and isolation, a topic that will be further discussed in chapter 6.

Finally, to investigate the relationship between the performance of the flow-free model and park isolation, the nMSE was analyzed as a function of the park connectivity measure proposed by ecologists (section 2.2), as shown in figure 5.6.
The aim was to assess whether the model provides a better fit for more isolated parks at this preliminary stage, where inter-park flows are not yet included. This analysis was motivated by the hypothesis that the absence of explicit flow terms may systematically favor more isolated parks, which are expected to be less affected by interactions with neighboring ecosystems.
Further considerations on this point will be discussed in section 5.4 and chapter 6.

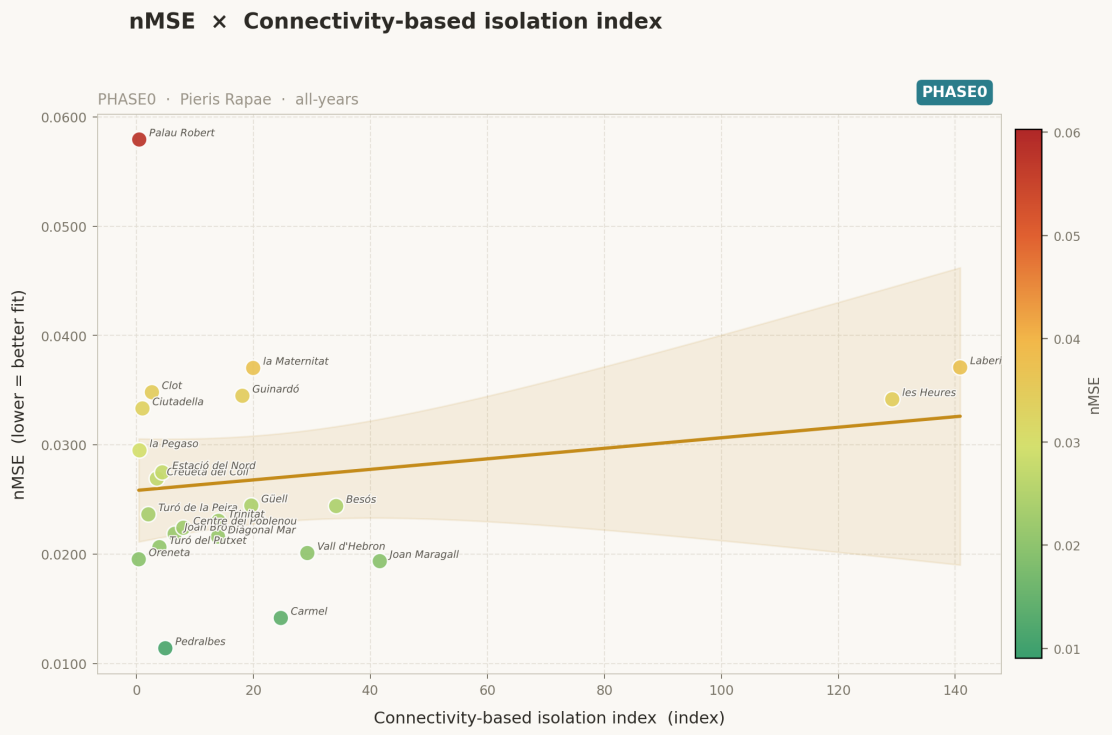


Figure 5.6: Illustration of the relationship between transport and isolation based on connectivity in phase 0. The x-axis shows connectivity values, ranging from more isolated parks (lower values, left) to more connected parks (higher values, right).

5.2 Phase 1

Phase 1 extends the previous formulation by introducing spatial flux interactions between nodes, based solely on euclidean distance (3.4.1). This step was natural given the unsatisfactory results obtained with phase 0. The model equation

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s\left(W_{yx}P_y^s(t) - W_{xy}P_x^s(t)\right)}_{Flux}$$

was implemented and fitted against the neural network data (2.5), with

$$w_{xy} = f(d_{xy}) \Longrightarrow w_{xy} \propto e^{-cost_{xy}} \begin{cases} cost_{xy} = \delta d_{xy} \\ cost_{xx} = 0 \implies d_{xx} = 0 \end{cases}$$

The parameter δ was estimated within the optimization, yielding the distribution over 100 runs in figure 5.7.

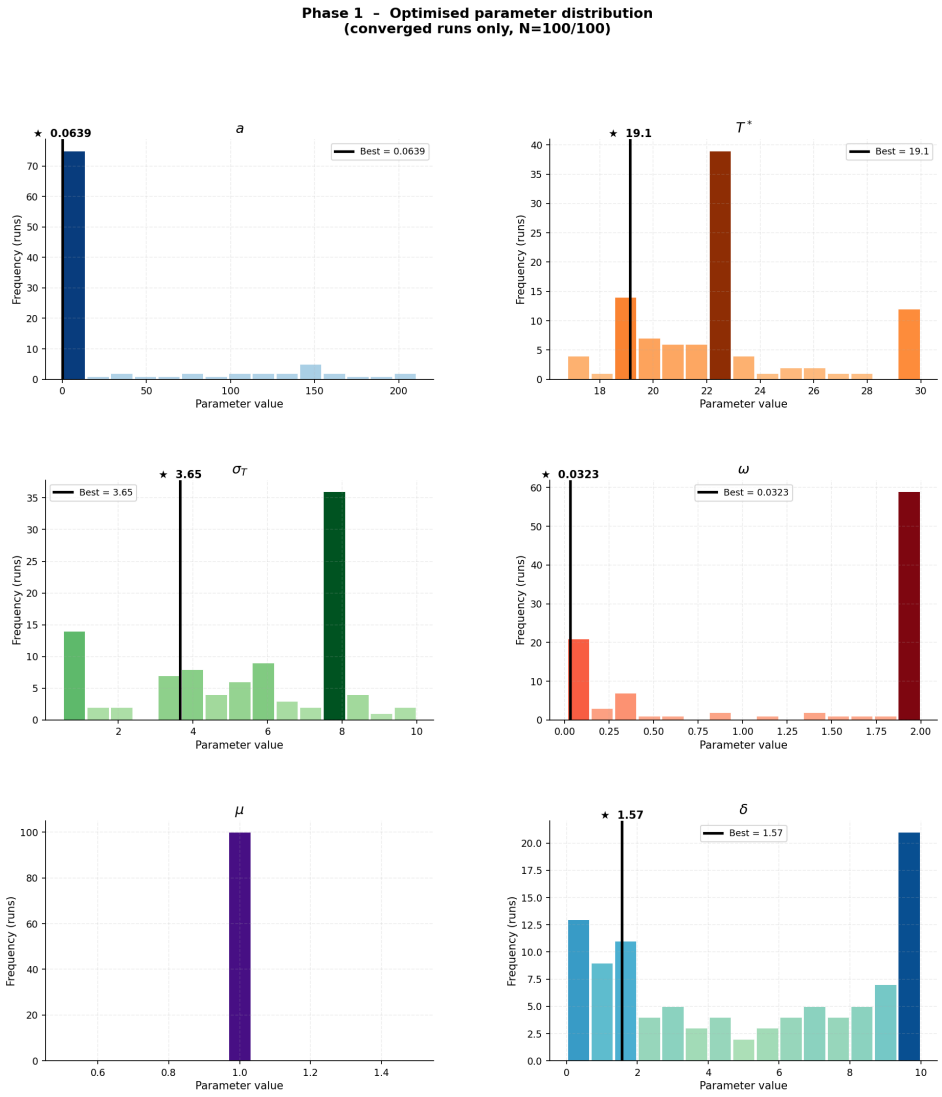


Figure 5.7: Frequency distribution of Phase 1 parameter estimates over 100 optimization runs (the highlighted value corresponds to the minimum-loss solution).

The optimal values, with minimum loss 0.0262 (figure 5.8), are:

- a [maximum emergence]: 0.0639 $\Longrightarrow$ increases slightly.
- T^* [optimal temperature]: 19.1 $°C$ $\Longrightarrow$ same as phase 0.
- σ_T [thermal breadth]: 3.65 $\Longrightarrow$ almost identical to phase 0.
- ω^s [non-acivity rate]: 0.0323 $\Longrightarrow$ average observable time of $\approx$ 31 days.
- δ [distance-scaling parameter]: 1.57

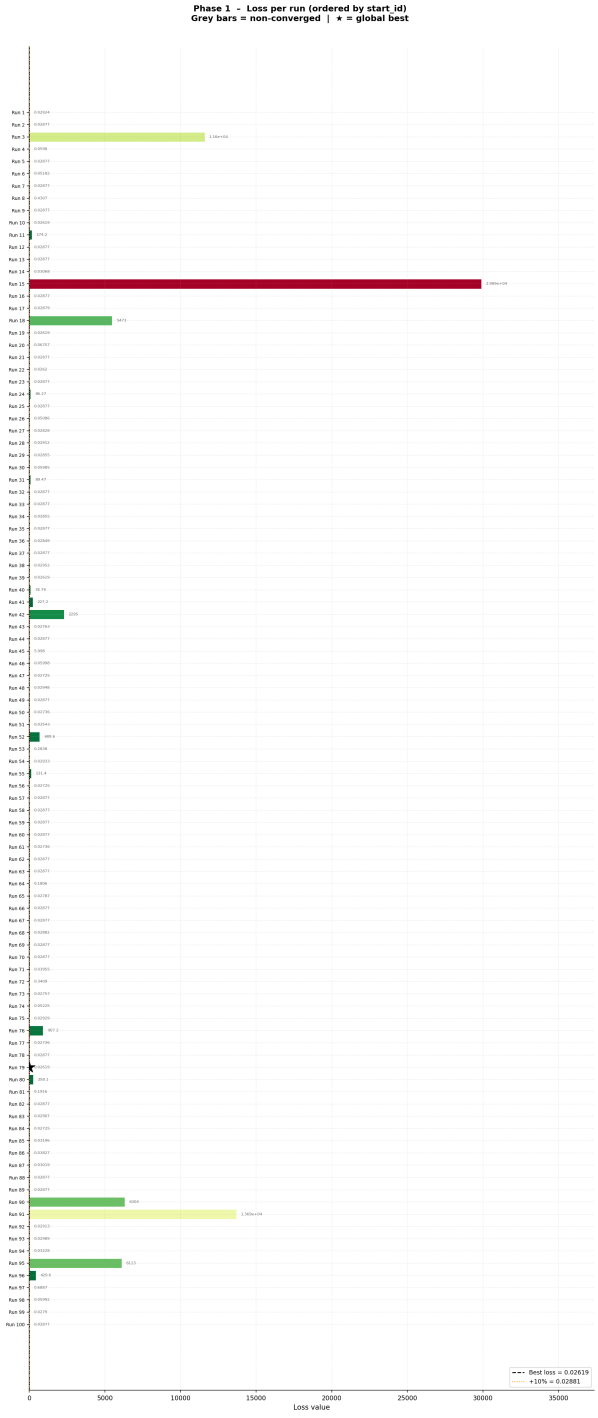


Figure 5.8: Phase 1 loss values obtained from the optimization procedure over the 100 independent runs.

Using the estimated parameters, the model reproduces the neural network output as illustrated in figure 5.9.

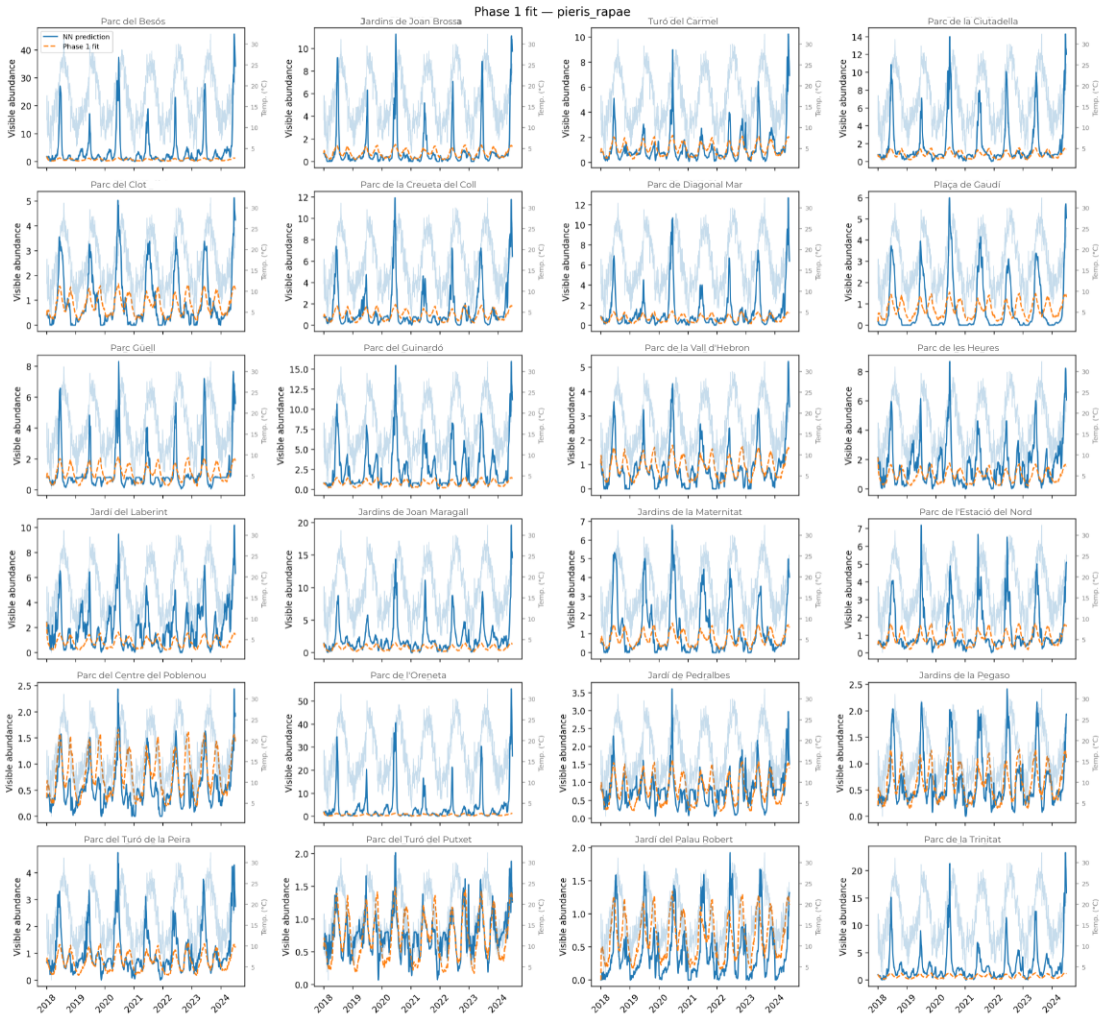


Figure 5.9: Phase 1 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile.

As in the previous phase, the complete set of initial and optimized parameter values for each run is reported in Appendix B, B.2.

Consistently with Phase 0, the same set of analyses was repeated for Phase 1, including the nMSE-based assessment of fit quality, the comparison between best

and worst cases, and the investigation of the relationship between model performance and park isolation and connectivity.
This was done in order to ensure a direct and consistent comparison between the two formulations under the same evaluation framework.

The corresponding nMSE results for Phase 1 are presented in Figure 5.10.

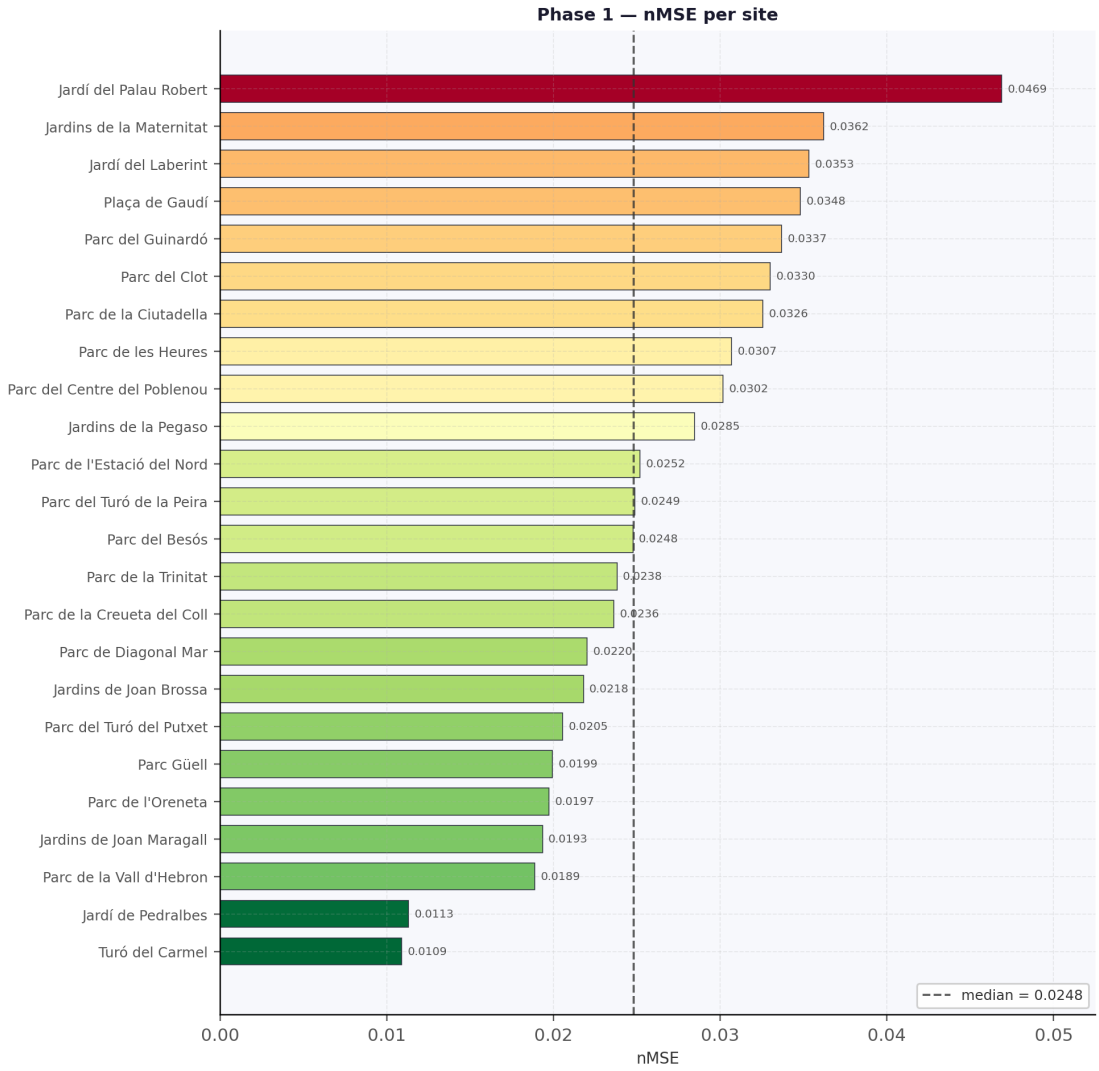


Figure 5.10: Fit quality with mse phase 1.

More specifically, Figure 5.11 presents representative examples from Jardins de Joan Maragall and Jardins de la Maternitat, respectively, providing a closer examination of the model fit under two distinct scenarios.

These two cases were selected as illustrative examples of different spatial configurations and levels of connectivity within the urban green network, allowing for a more detailed qualitative assessment of the model behavior beyond aggregated performance metrics.
A more detailed analysis of these cases will be provided in the following chapter 6.

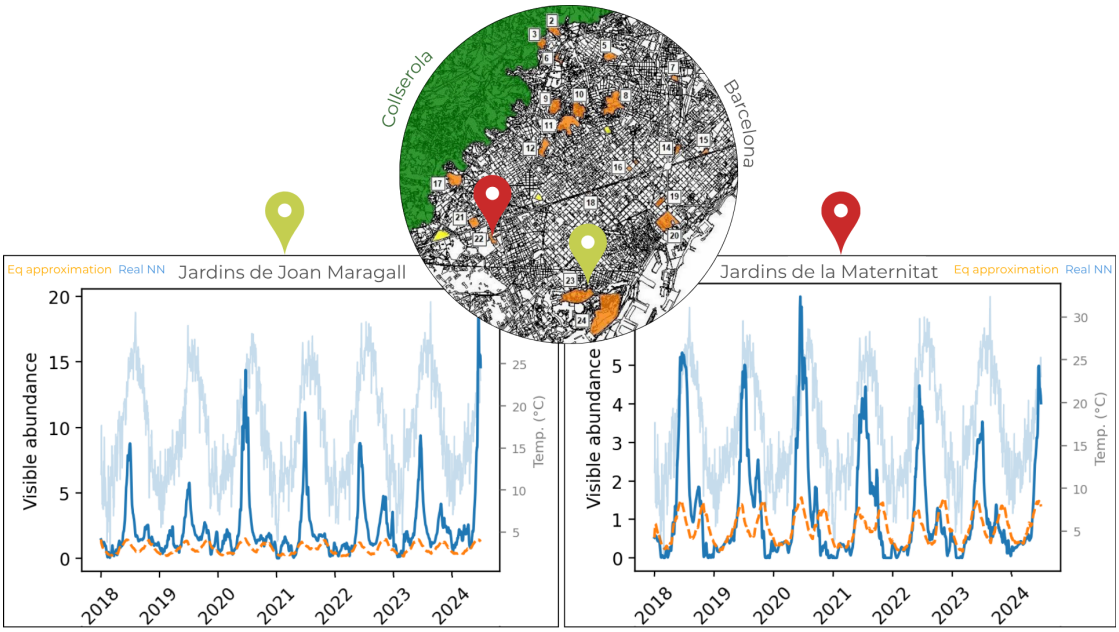


Figure 5.11: Left: Phase 1 representation of Jardins de Joan Maragall, an inland park in Barcelona situated in close proximity to several other green areas. Right: Jardins de la Maternitat, located within Barcelona's urban network in a non-central area of the city and closer to Collserola Park, as indicated by the inset map.

Finally, to assess the dependence of the flux-based model on park isolation and connectivity, the nMSE error was analysed as a function of the connectivity measure proposed by ecologists (section 2.2).
This analysis extends the results obtained in phase 0, with the aim of evaluating whether the introduction of a distance-based flux term alters the previously observed relationship between connectivity and model fit quality.

In particular, it allows one to investigate whether accounting for inter-park interactions modifies the tendency of the model to perform differently across parks with varying degrees of isolation. The corresponding results for the flux formulation are reported in figure 5.12.

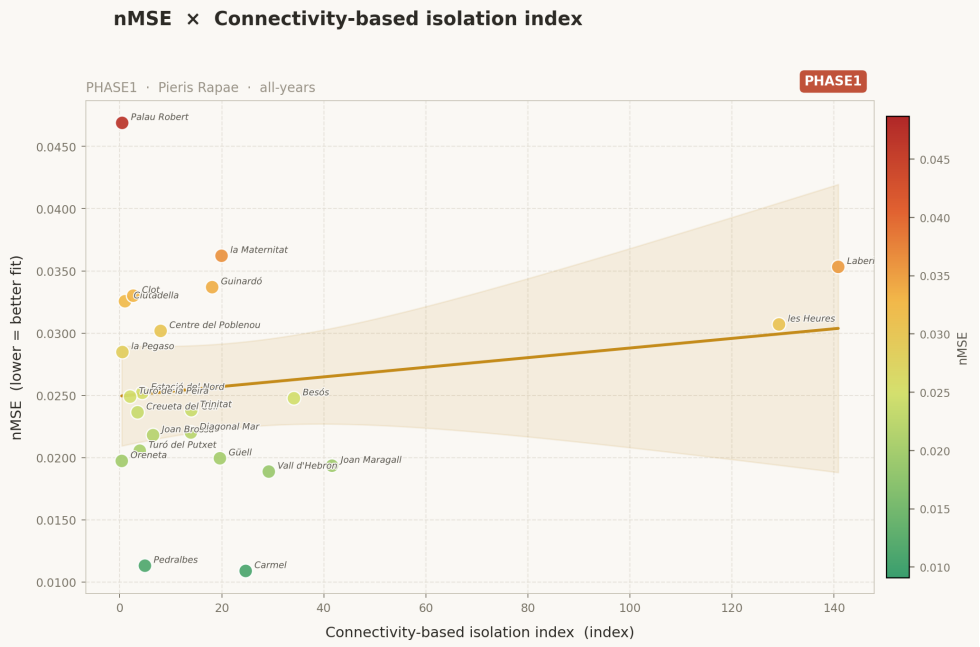


Figure 5.12: Relationship between transport and connectivity-based isolation in phase 1. The x-axis shows connectivity values, from more isolated (left) to more connected (right) parks.

5.3 Phase 2

Although the results improved with phase 1, the previous section showed that distance alone is not sufficient to fully capture the dynamics, motivating the introduction of phase 2, which builds upon the previous formulation by incorporating spatial flux interactions between connected nodes based on connectivity (2.2) and quality (2.1) measures (details in 3.4.1). In this phase, the model equation

$$\frac{dP_x^s}{dt} = \underbrace{\alpha^s(T_x(t))}_{Fly-period} - \underbrace{\omega^s P_x^s(t)}_{Non-Activity} + \underbrace{\sum_y \mu^s\Big(W_{yx}P_y^s(t) - W_{xy}P_x^s(t)\Big)}_{Flux}$$

was implemented and fitted against the neural network data (2.5), with

$$w_{xy} = f(C_y, Q_y) \Longrightarrow w_{xy} \propto e^{-cost_{xy}} \begin{cases} cost_{xy} = -\gamma C_y - \rho Q_y \\ cost_{xx} = -\gamma C_x - \rho Q_x \end{cases}$$

The parameters γ and ρ were estimated within the optimization procedure, as no prior information was available for them, and this yielded the parameter estimates summarized in figure 5.13, which displays the frequency distribution of the parameter values obtained across 100 independent optimization runs.

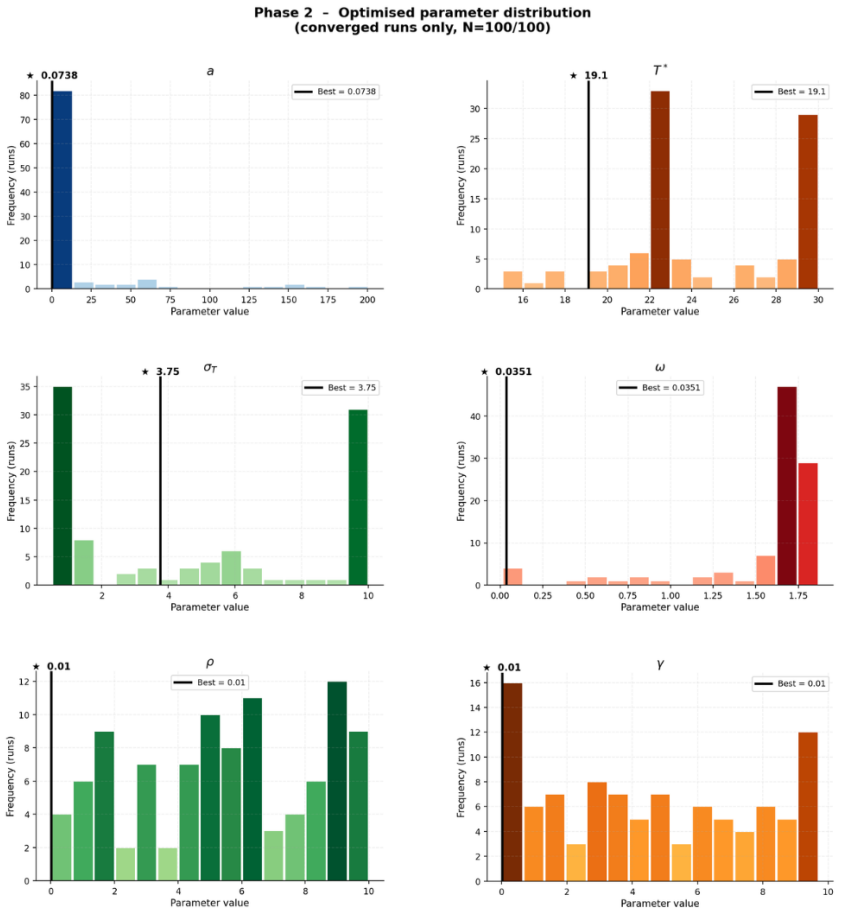


Figure 5.13: Frequency distribution of Phase 2 parameter estimates over 100 independent runs; the highlighted value corresponds to the minimum-loss solution.

The optimal parameter values correspond to the minimum loss of 0.0260 (see figure 5.14) are:

- a [maximum emergence]: 0.0738 $\Longrightarrow$ increases slightly.
- T^* [optimal temperature]: 19.1 $°C$ $\Longrightarrow$ same as phase 1.
- σ_T [thermal breadth]: 3.75 $\Longrightarrow$ almost identical to phase 0 and 1.
- ω^s [non-acivity rate]: 0.0351 $\Longrightarrow$ increased compared to phase 1, leading to a value of average observable of $\approx$ 28 days.
- γ [connectivity-scaling parameter]: 0.01
- ρ [quality-scaling parameter]: 0.01

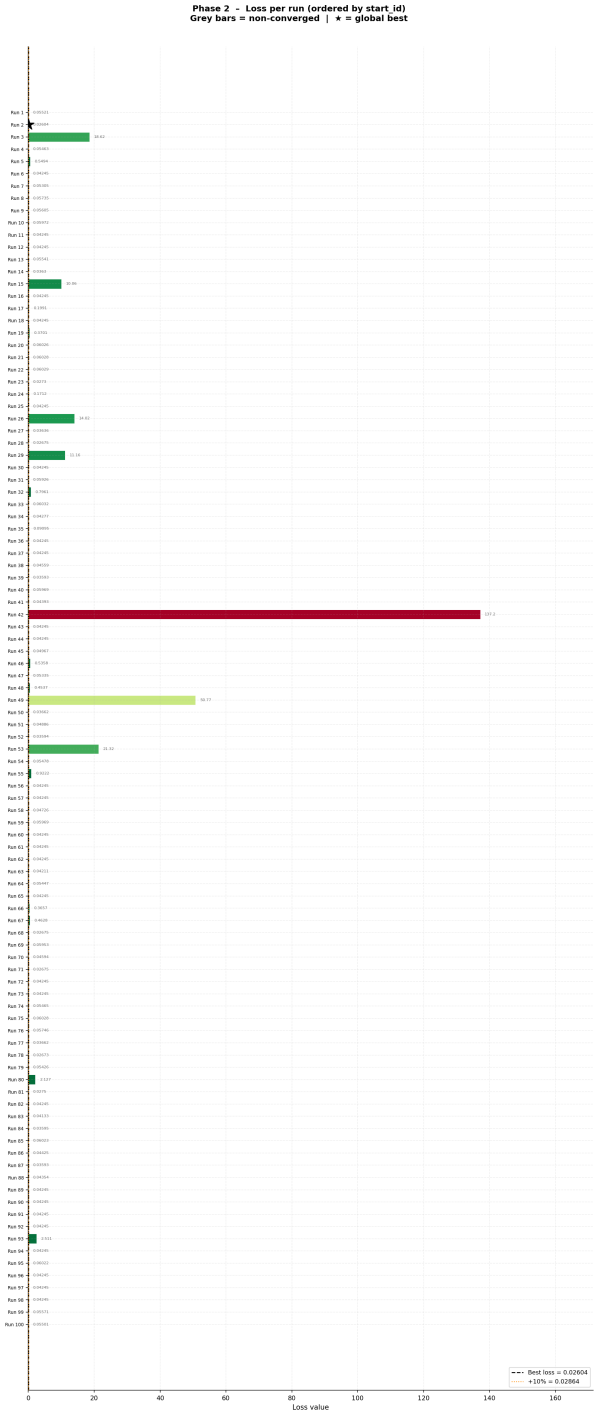


Figure 5.14: Phase 2 loss values obtained from the optimization procedure over the 100 independent runs.

Using these estimated parameters, the model reproduces the neural network output as illustrated in figure 5.15.

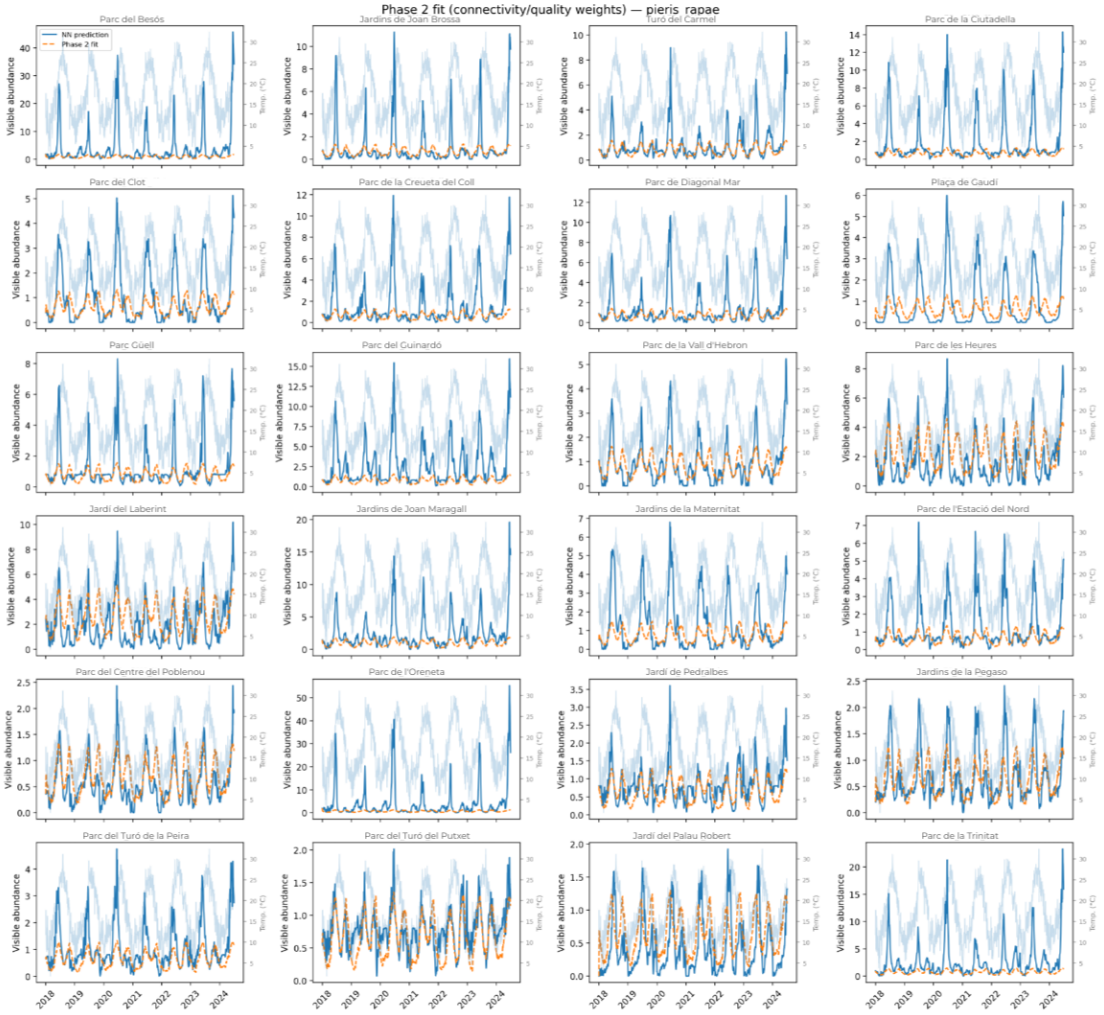


Figure 5.15: Phase 2 model fit obtained using the optimized parameter values. The simulated dynamics (orange) are compared with the neural network predictions (dark blue), while the background (light blue) represents the temperature profile.

The complete set of initial and optimized parameter values for each of the 100 runs is reported in Appendix B, B.3, together with a histogram of the initial parameter distribution, as in figure 5.13.

As in Phase 0 and Phase 1, the same set of analyses was repeated for Phase

2, in order to allow a direct and consistent comparison across all three formulations.

Starting from the nMSE analysis, the corresponding results are presented below in Figure 5.16:

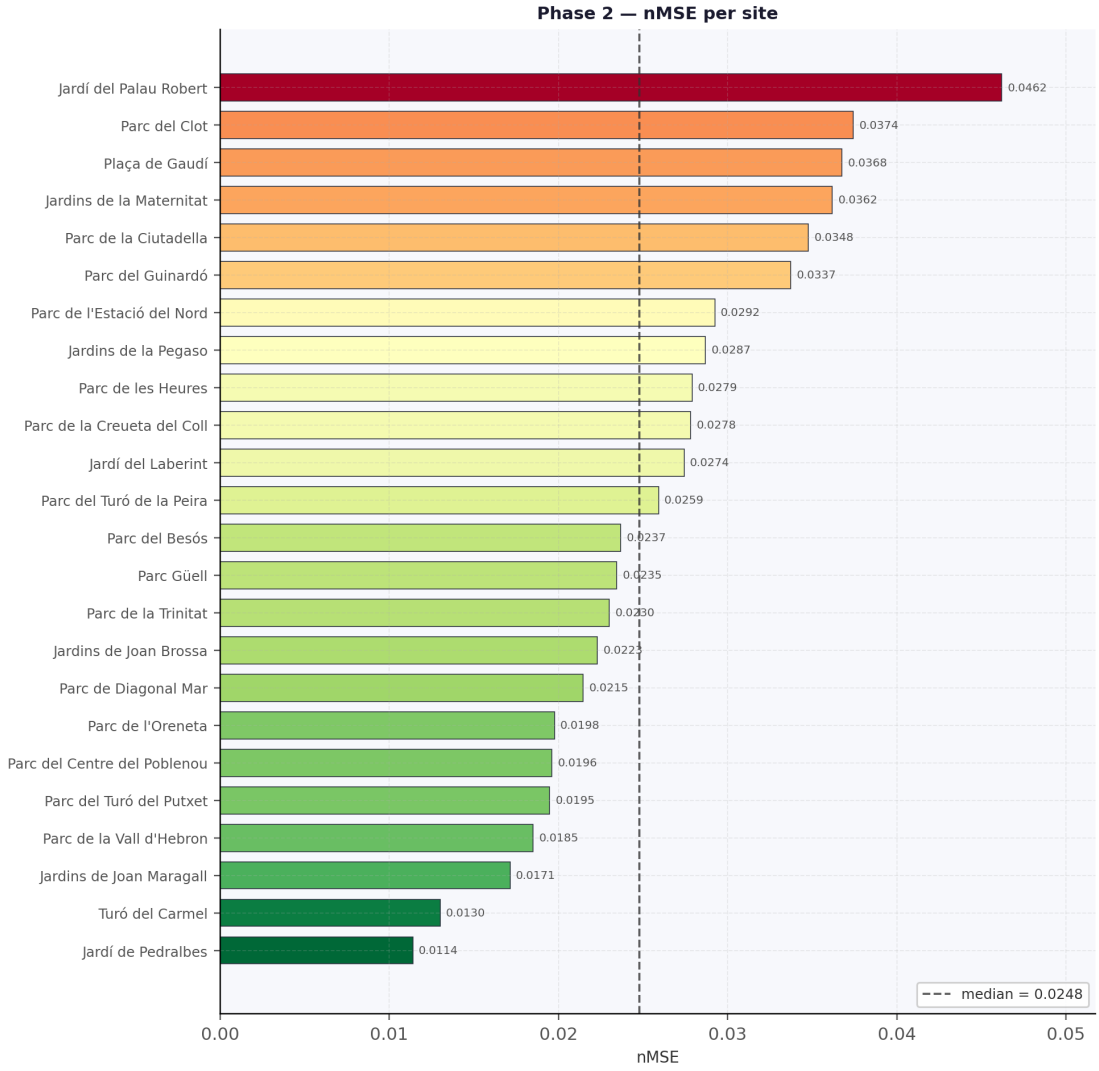


Figure 5.16: Fit quality with mse phase 2.

As in the previous phases, the inclusion of flow terms based on connectivity and quality does not eliminate the heterogeneity in model fitting performance across parks.
The underlying causes of this variability will be discussed in chapter 6.

To further explore this variability, figure 5.17 shows representative examples from Jardins de Joan Maragall and Jardins de la Maternitat, respectively, allowing a closer examination of the model behavior under two distinct cases.
These two parks were selected as illustrative examples of different spatial configurations and connectivity conditions within the urban green network, in order to provide a more detailed qualitative assessment of the model performance beyond aggregate error metrics. In particular, this comparison helps highlight how the model responds to contrasting structural and environmental settings across individual park systems.

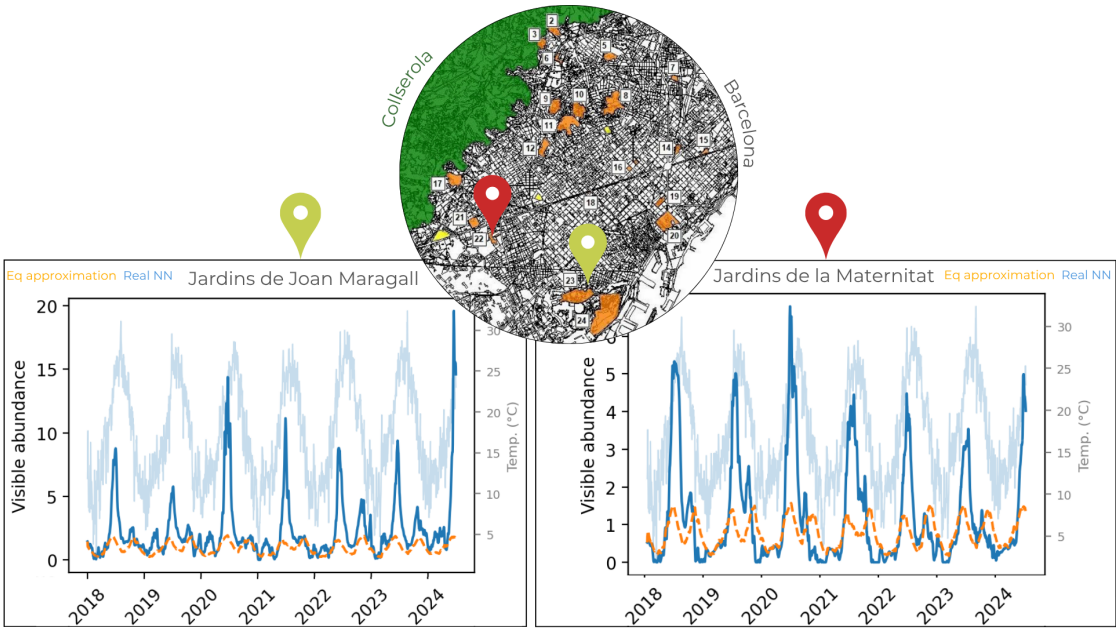


Figure 5.17: Phase 2 fitting. Left: Jardins de Joan Maragall, an inland park surrounded by several neighbouring green areas. Right: Jardins de la Maternitat, located in an off-centre position and in closer proximity to Collserola Park (see inset map).

Finally, as previously done for the other phases, the nMSE error was examined as a function of park isolation.
This analysis aimed to determine whether the joint use of connectivity and quality in the flux term further altered the relationship between fit quality and park connectivity observed in the previous phases.

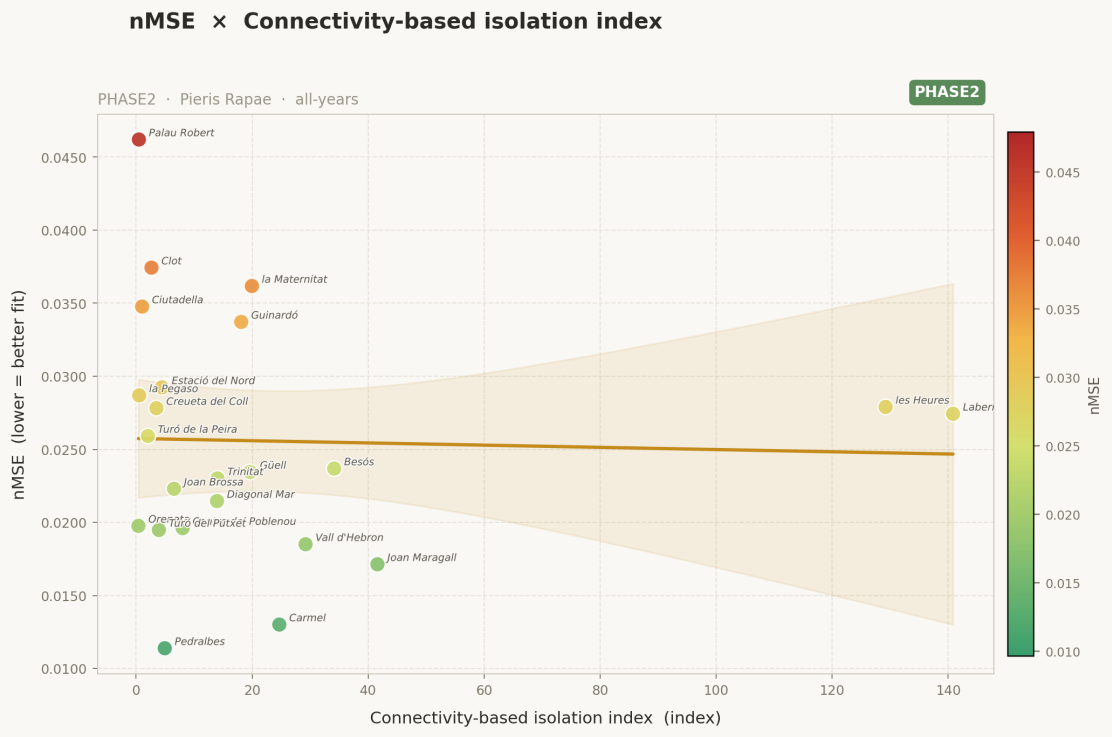


Figure 5.18: Illustration of the relationship between transport and isolation based on connectivity in phase 2. The x-axis shows connectivity values, ranging from more isolated parks (lower values, left) to more connected parks (higher values, right).

5.4 Comparison of Model Phases

Having presented the three phases individually, this section provides a direct comparison of their performance, in order to assess whether the progressive inclusion of spatial flux terms, first based on euclidean distance (Phase 1) and then on connectivity and quality (Phase 2), leads to a systematic improvement in the agreement between simulated and observed dynamics across the 24 parks considered.

Table 5.1 reports the nMSE values obtained for each site in Phase 0, Phase 1 and Phase 2, together with the overall variation Δ (computed as the difference between the last and the first phase) and the corresponding trend.

Comparing Phase 0 to Phase 2, 15 sites show an improvement in fit quality, 9 sites show a worsening, and no site remains unchanged.

Site	**Phase 0**	**Phase 1**	**Phase 2**	**Δ (last-first)**	**Trend**
Parc del Besós	0.02439	0.02476	0.02369	-0.00070	↓ better
Jardins de Joan Brossa	0.02185	0.02180	0.02230	0.00046	↑ worse
Turó del Carmel	0.01415	0.01089	0.01301	-0.00114	↓ better
Parc de la Ciutadella	0.03332	0.03256	0.03477	0.00014	↑ worse
Parc del Clot	0.03481	0.03301	0.03743	0.00263	↑ worse
Parc de la Creueta del Coll	0.02691	0.02364	0.02782	0.00091	↑ worse
Parc de Diagonal Mar	0.02159	0.02201	0.02146	-0.00013	↓ better
Plaça de Gaudí	0.03488	0.03480	0.03675	0.00187	↑ worse
Parc Güell	0.02444	0.01994	0.02345	-0.00098	↓ better
Parc del Guinardó	0.03448	0.03369	0.03372	-0.00076	↓ better
Parc de la Vall d'Hebron	0.02009	0.01887	0.01851	-0.00158	↓ better
Parc de les Heures	0.03416	0.03070	0.02790	-0.00626	↓ better
Jardí del Laberint	0.03708	0.03532	0.02743	-0.00965	↓ better
Jardins de Joan Maragall	0.01935	0.01934	0.01714	-0.00221	↓ better
Jardins de la Maternitat	0.03703	0.03621	0.03618	-0.00085	↓ better
Parc de l'Estació del Nord	0.02747	0.02520	0.02924	0.00177	↑ worse
Parc del Centre del Poblenou	0.02241	0.03018	0.01961	-0.00280	↓ better
Parc de l'Oreneta	0.01952	0.01973	0.01977	0.00024	↑ worse
Jardí de Pedralbes	0.01138	0.01131	0.01139	0.00001	↑ worse
Jardins de la Pegaso	0.02949	0.02848	0.02869	-0.00079	↓ better
Parc del Turó de la Peira	0.02363	0.02489	0.02591	0.00228	↑ worse
Parc del Turó del Putxet	0.02063	0.02055	0.01949	-0.00115	↓ better
Jardí del Palau Robert	0.05793	0.04689	0.04620	-0.01173	↓ better
Parc de la Trinitat	0.02304	0.02381	0.02301	-0.00003	↓ better

Table 5.1: nMSE evolution across phases for the 24 parks considered, comparing Phase 0, Phase 1 and Phase 2.

It is worth noting that Jardí del Palau Robert, which in Phase 0 exhibited by far the highest nMSE among all sites (as already highlighted in section 3.4.1), shows the largest absolute improvement across phases ($\Delta = -0.01173$), suggesting that this park benefits substantially from the inclusion of spatial flux terms.
Similarly, Jardí del Laberint displays a marked improvement ($\Delta = -0.00965$), consistent with the qualitative comparison discussed in figures 5.5, 5.11 and 5.17. Conversely, sites such as Parc del Clot and Plaça de Gaudí show a non-negligible worsening, indicating that the progressive addition of flux terms does not uniformly benefit every park in the network, and that local factors continue to play a role in determining fit quality even in the most complete formulation.

The overall evolution of the fit quality across the three phases is further illustrated in figure 5.19, which allows a visual assessment of the trends summarized in table 5.1.

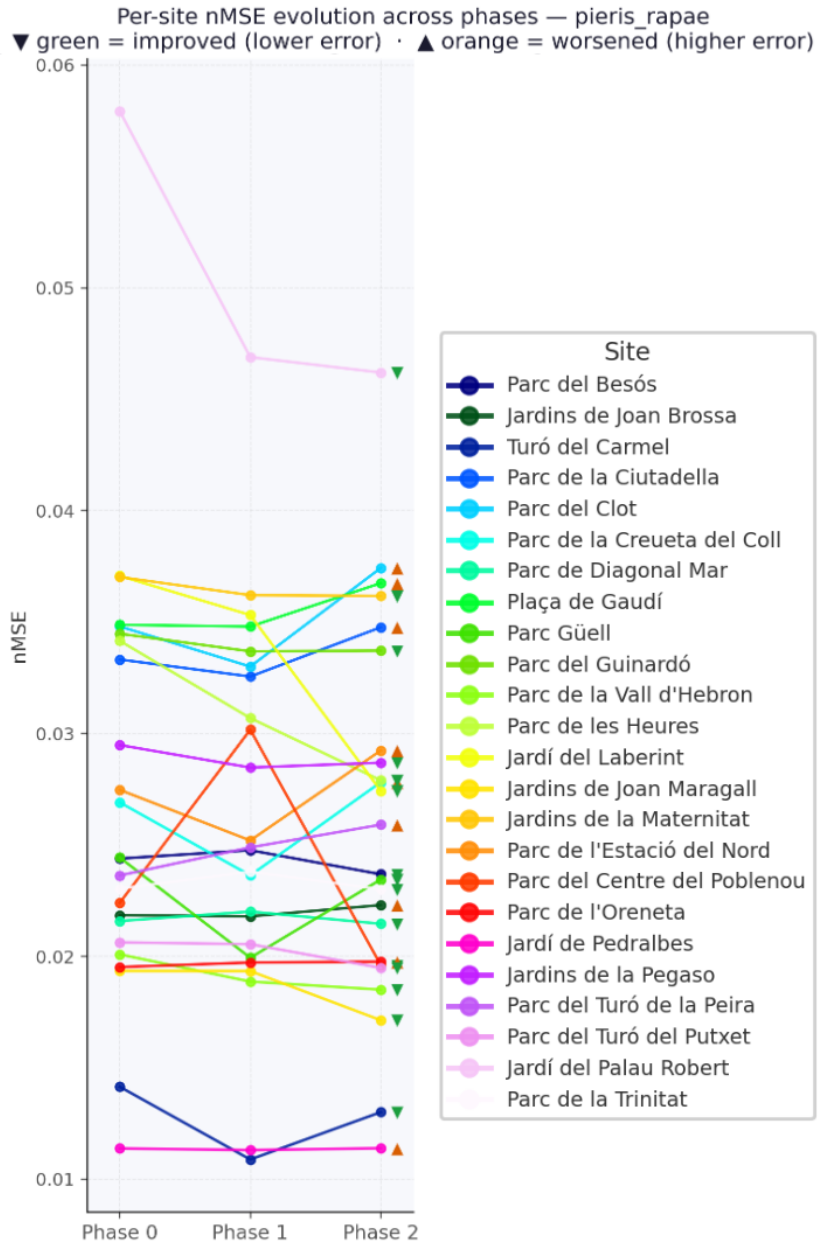


Figure 5.19: Evolution of the nMSE across the three model phases for each of the 24 parks considered.

Beyond the site-specific fit quality, it is also useful to examine how the estimated model parameters evolve across phases, since consistent trends in the parameter values may provide additional support for the robustness of the optimization procedure.

Table 5.2 summarizes the optimal parameter estimates obtained in Phase 0, Phase 1 and Phase 2, together with the overall variation observed between Phase 0 and Phase 2.

Parameter	**Phase 0**	**Phase 1**	**Phase 2**	**Variation (P0 → P2)**
a	0.0597	0.0639	0.0738	+23.6%
T^*	19.2 °C	19.1 °C	19.1 °C	−0.1 °C
σ_T	3.74	3.65	3.75	+0.3%
ω^s	0.0332	0.0323	0.0351	+5.7%
δ	-	1.57	-	-
γ	-	-	0.0100	-
ρ	-	-	0.0100	-
Loss	0.0272	0.0262	0.0260	−4.4%

Table 5.2: Comparison of the optimal parameter estimates across Phase 0, Phase 1 and Phase 2, with overall variation between Phase 0 and Phase 2.

Having summarized how the calibrated parameters and overall loss evolve across the three phases, the following discussion turns to a more detailed analysis of the model anticipated above.

Overall, the biological parameters T^* and σ_T remain remarkably stable across the three phases, varying by less than 1% and by only 0.1 °C respectively, indicating that the thermal response of the population is robustly identified regardless of the specific flux formulation adopted.
The maximum emergence parameter a increases progressively from Phase 0 to Phase 2 (+23.6%), while the non-activity rate ω^s shows a more moderate increase (+5.7%), suggesting that as spatial interactions are progressively incorporated, the model compensates by allowing a higher emergence rate together with a slightly faster decay of non-active individuals.
Finally, the monotonic decrease in the loss value from Phase 0 (0.0272) to Phase 2 (0.0260), corresponding to an overall reduction of 4.4%, confirms that the inclusion of spatial flux terms, and in particular the joint use of connectivity and quality in Phase 2, leads to a modest but consistent improvement in the overall agreement between the model and the neural network data, even though, as shown in table 5.1, this improvement is not uniformly distributed across all sites.

5.5 Flux analysis

Building on the increased representational capacity of phase 2, a flux-based analysis was conducted to address the project's primary objective: classifying habitats as sources, sinks, or ecological traps. As a brief reminder from section 3.4.2:

$$\begin{cases} \text{SOURCE} \Rightarrow \text{net export of individuals and high local fitness} \\ \text{SINK} \Rightarrow \text{net import of individuals but low local fitness} \\ \text{TRAP} \Rightarrow \text{sink dynamics combined with high attractiveness (high } w_{xy}) \end{cases}$$

the overall dynamics of each park are largely determined by the balance between incoming and outgoing flux. The following analysis therefore focuses on the flux term resulting from the model, distinguishing for each park between inflow and outflow contributions and comparing this decomposition with the attractiveness term, closely related to park quality, as illustrated in the figures below.
Figure 5.20 shows a preliminary classification of parks based on net flow.

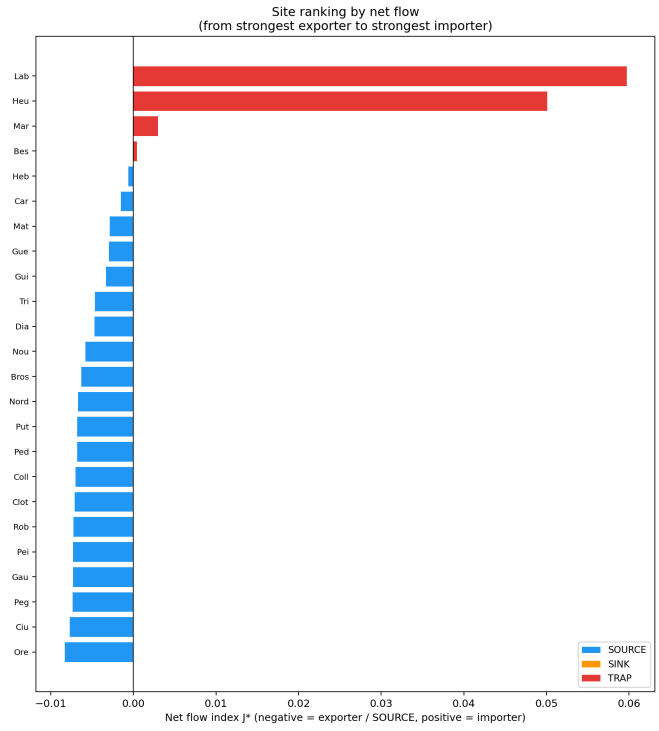


Figure 5.20: Ranking of parks by net flow: sites experiencing net immigration are shown in red, while those experiencing net emigration are shown in blue.

The attractiveness term, closely linked to park quality, was then examined for each site and compared with the net flow ranking presented above (figure 5.21), in order to verify whether the parks identified as strong net immigration sites also correspond to those exhibiting the highest attractiveness.
The results of this comparison are shown in figure 5.22.

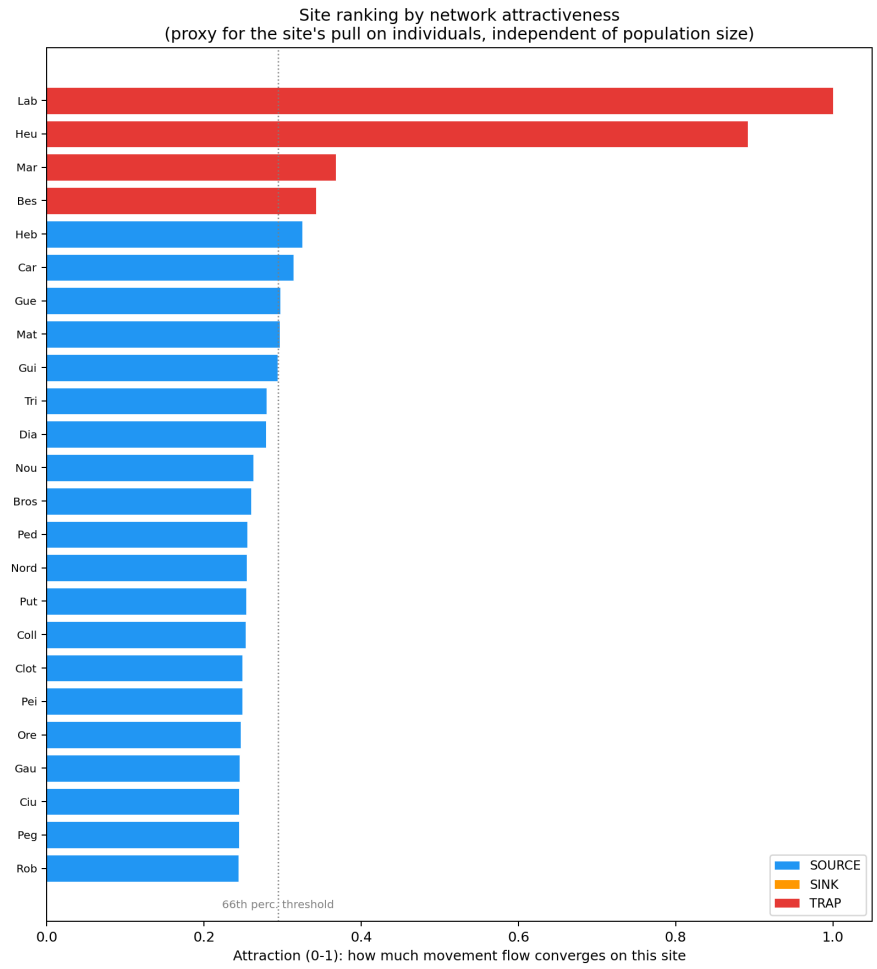


Figure 5.21: Attractiveness ranking of the parks: sites with high immigration, as identified in figure 5.20, are shown in red, while those with high emigration are shown in blue.

It is worth noting from this figure 5.21 that the parks characterized by high immigration coincide with those exhibiting the highest attractiveness, making them well-suited candidates to act as traps within the network.

Finally, figure 5.22 brings together the results of the two previous analyses, relating net flow and attractiveness in order to identify the parks that emerge as the most likely candidates to act as traps within the network.

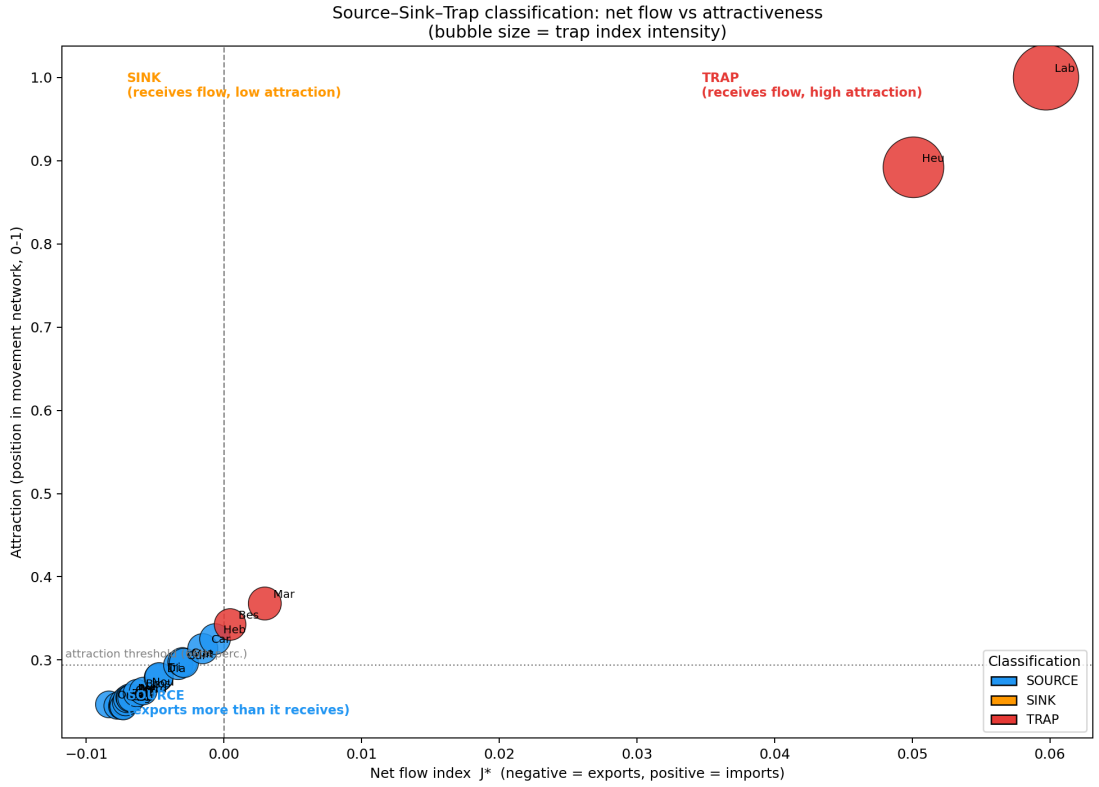


Figure 5.22: Sites ranked by attraction (share of network movement flow converging on each site, from the column sums of W).

From this figure 5.22, the parks Jardí del Laberint, Parc de les Heures, Jardins de Joan Maragall and Parc del Besós emerge as strong candidates for trap behaviour, since they combine a positive net flow with a relatively high attractiveness value. However, net flow alone provides only a preliminary indication of trap dynamics, as it does not explicitly capture the interplay between local population deficit and the attractiveness driving immigration. To investigate the dynamics in a more rigorous and structural way, the trap index was therefore computed for each park, combining these two aspects into a single, more informative metric. The resulting values are mapped in figure 5.23, which overlays two related quantities for each park, ranked from least to most problematic (bottom to top): the trap index itself and the underlying sink severity from which it is derived.

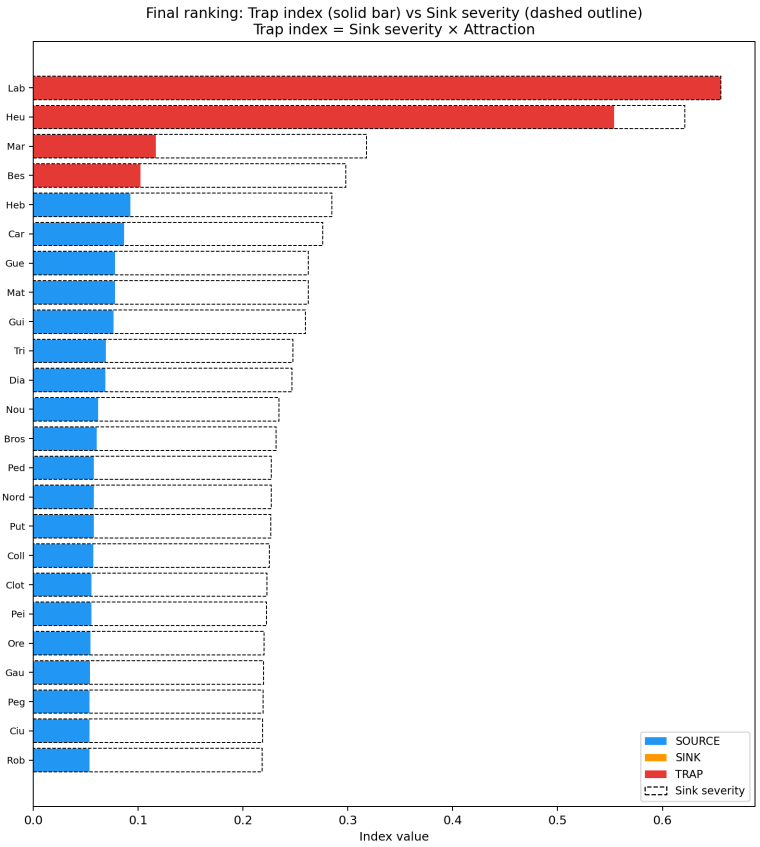


Figure 5.23: Trap index and sink severity by park, ranked from lowest to highest trap index. The solid colored bar shows the trap index, the final metric quantifying how strongly a park behaves as an ecological trap, while the dashed black outline shows the sink severity, the raw local-deficit component before being weighted by the park's attractiveness in the movement network. Colors indicate the final source/sink/trap classification.

Most parks classified as source show a local sink severity comparable to the trap-classified ones, yet their low attraction keeps the final trap index low, indicating they experience some local deficit without effectively drawing in individuals from elsewhere. Jardí del Laberint and Parc de les Heures have attraction close to 1, so their trap index closely matches their sink severity, confirming them as the strongest candidates for genuine ecological traps, with both a high local deficit and a strong pull on immigrants. Jardins de Joan Maragall and Parc del Besós show an intermediate pattern: their sink severity is similar to that of several source parks, but a moderately elevated attraction, is enough to raise their trap index above the source group, resulting in a weaker, borderline trap signal.

Chapter 6

Conclusions

This section discusses and interprets the results obtained in chapter 5 for the species *Pieris rapae*.

Interpreting source, sink, and trap dynamics for a highly mobile, generalist, metapopulation structured species such as *Pieris rapae* requires particular caution. Given its dispersal capacity, individual urban parks may not constitute ecologically meaningful units in isolation, as they are functionally embedded within a broader network of interconnected green spaces; classifying a single site independently of this connectivity risks conflating a scale artifact with a genuine demographic property. Moreover, high mobility can obscure true sink dynamics through a rescue effect, whereby a locally unsustainable population is continuously replenished by immigration from source areas, generating a "pseudo-sink" that appears stable when assessed from abundance data alone. This ambiguity can only be resolved through a flow-based analysis that explicitly quantifies births, deaths, immigration, and emigration, rather than relying on static demographic snapshots. Although the generalist habits of *Pieris rapae* render classic ecological traps less likely than for specialist species governed by rigid habitat-selection cues, the risk is not negligible: ornamental Brassicaceae or other urban-specific cues could still induce oviposition at sites where larval survival is compromised by pesticide use or elevated parasitoid pressure, thereby generating an urban-specific trap effect. These considerations underscore that, for mobile, metapopulation-structured species, the source-sink-trap framework is ecologically informative only when integrated with a flow-based, network-aware analysis, rather than inferred from closed, patch-level population estimates.

It is precisely for this reason that the model developed in this thesis constitutes a valuable analytical tool. However, examining the geographic distribution of the parks classified as traps in section 5.5, namely Jardí del Laberint, Parc de les Heures, Jardins de Joan Maragall, and Parc del Besós, reveals that all four are

located at the periphery of Barcelona. This suggests that these sites are likely subject to substantial external dynamics that are not yet accounted for in the present model. The results obtained should therefore be interpreted with this, bearing in mind that this is an ongoing project and that the findings presented here represent only the first outputs of a model that remains partially representative of reality.

Beyond the classification of individual parks as sources, sinks, or traps, a complementary perspective on model performance emerges from examining how the quality of the fit itself evolves across the three formulations developed in this thesis, and how this evolution relates to the spatial structure of the urban park network.

Already in phase 0, where parks are treated as closed, non-interacting systems, a clear pattern emerges: the quality of the fit is not uniform across the network but depends systematically on the local spatial and ecological context of each site (figure 5.4). As shown in figure 5.6, fitting error increases with park connectivity, indicating that the closed-system assumption is most appropriate for relatively isolated parks, whereas its adequacy deteriorates as inter-park interactions become more significant.
The introduction of a euclidean distance based flux term in phase 1 improves the fit across the network overall (figure 5.10), yet does not eliminate this heterogeneity: as illustrated in figure 5.11 through the contrasting cases of Jardins de Joan Maragall, an inland park surrounded by several neighbouring green areas that is well reproduced despite an underestimation of peak abundances and Jardins de la Maternitat, located off-centre and closer to Collserola Park, whose dynamics remain poorly captured, the model's performance continues to depend on factors beyond distance alone. Indeed, figure 5.12 confirms that the increasing trend between connectivity and fitting error persists under the flux formulation, a result that is unsurprising given that connectivity, as constructed here, reflects levels of urban integration that a purely euclidean distance based model cannot capture; more connected parks therefore remain systematically less well represented even after the inclusion of spatial flux.

Throughout both formulations, the parameters T^* and σ_T remain essentially unchanged, a stability that reinforces their interpretation as intrinsic, species-specific properties that should not be affected by the details of the system's spatial dynamics. By contrast, the parameters a and ω^s both increase from phase 0 to phase 1: the growth in a points to a stronger role of immigration processes in shaping the phase dynamics, while the increase in ω^s reduces the number of days over which the species remains observable, bringing the model into closer agreement with the empirical average lifespan of *Pieris rapae*, estimated at approximately 20 days [51].

The move to phase 2, in which movement is governed by a full cost-based formulation incorporating both landscape connectivity C_y and habitat quality Q_y, represents a further step in this progression. Notably, the two parameters controlling connectivity and quality converge to comparable values, suggesting that both factors contribute with similar importance to butterfly flow dynamics; this outcome should nonetheless be interpreted with some caution, given the inherent difficulty of estimating a limited set of six parameters capable of adequately representing the dynamics of 24 distinct parks, particularly when Barcelona is still treated as a closed system that neglects potentially significant source terms originating outside the city. Despite this limitation, the overall nMSE decreases for all parks in phase 2 (figure 5.16), confirming that the fuller formulation captures the movement and dispersal behaviour of *Pieris rapae* more accurately than the previous two. This improvement is again illustrated in figure 5.17 through the same two representative parks, Jardins de Joan Maragall and Jardins de la Maternitat, whose contrasting fit quality mirrors the pattern already observed in phase 1. More importantly, the relationship between fitting error and isolation changes qualitatively in this final formulation: whereas figure 5.18 shows an essentially flat trend, in marked contrast to the increasing trend observed in phases 0 and 1, indicating that the explicit inclusion of connectivity and habitat quality as drivers of the flux term homogenizes the model's representation of parks across the full range of isolation values, regardless of whether they are peripheral or well embedded within the urban network.

This convergence across phases does not, however, imply a uniformly good fit for every site. When the spatial distribution of residual errors is examined across all three formulations (figures 5.4, 5.10, and 5.16), a consistent pattern emerges whereby the parks least well represented by the model tend to be those most centrally embedded within the urban fabric, suggesting that the simplified treatment of urbanization and habitat quality adopted here does not fully capture the physical characteristics of such densely surrounded sites; a more detailed and updated characterization of these quality metrics, particularly for small, centrally located parks, would likely help address this limitation. Jardí del Palau Robert stands out as the least well-represented site throughout, a result already visible in phase 0 (figure 5.4) and persisting, albeit with markedly improved representativeness, across later phases (table 5.1, figure 5.19). Although apparently isolated, its proximity to the heavily trafficked Avinguda Diagonal likely introduces a pronounced and asymmetric urbanization effect, weaker toward the coast than on the avenue-facing side, which the current formulation does not resolve. At the opposite end of the spectrum, Turó del Carmel is consistently among the best-represented parks, owing to its position in a hillier, more interconnected area of Barcelona where the model successfully captures butterfly flow between neighbouring sites. Not all sites

improve monotonically with model complexity, however: parks such as Jardins de Joan Brossa, located far from the main urban core, show a deterioration in fit that is likely attributable to external influences not captured by the current model, and, together with other similarly behaving sites, would benefit from closer discussion with ecologists, since the underlying modelling assumptions do not appear fully justified in these cases.

Taken together, these results indicate that phase 0, which excludes movement and dispersal altogether, and phase 1, which incorporates movement based solely on euclidean distance, are both insufficient to capture the spatial heterogeneity observed across the network, implying that geographic distance alone does not adequately describe dispersal processes in an urban landscape. The substantial improvement achieved in phase 2 through the inclusion of habitat quality Q_y and landscape connectivity C_y demonstrates that a more complete, cost-based representation of movement is necessary to obtain a realistic description of butterfly dynamics across heterogeneous urban parks.
While the model is not without limitations, particularly with regard to peripheral parks subject to external dynamics and to small, centrally located sites whose urbanization and quality characteristics are only approximately captured, it nonetheless provides a valuable and promising foundation for the ecological analysis of *Pieris rapae* population dynamics in urban settings.

6.1 Final consideration

Taken together, these results indicate that phase 0, which excludes movement and dispersal altogether, and phase 1, which incorporates movement based solely on euclidean distance, are both insufficient to capture the spatial heterogeneity observed across the network, implying that geographic distance alone does not adequately describe dispersal processes in an urban landscape. The substantial improvement achieved in phase 2 through the inclusion of habitat quality Q_y and landscape connectivity C_y demonstrates that a more complete, cost-based representation of movement is necessary to obtain a realistic description of butterfly dynamics across heterogeneous urban parks.
While the model is not without limitations, particularly with regard to peripheral parks subject to external dynamics and to small, centrally located sites whose urbanization and quality characteristics are only approximately captured, it nonetheless provides a valuable and promising foundation for the ecological analysis of *Pieris rapae* population dynamics in urban settings.

More broadly, this progression across phases reflects the central contribution of the

present work: the development of a graph-based system of differential equations capable of modelling butterfly population dynamics with the primary objective of estimating dispersal fluxes and using them to infer habitat suitability. Despite the limited availability of ecological data and the simplifying assumptions required by the current formulation, most notably the treatment of Barcelona as a closed system and the approximate characterization of urbanization and habitat quality discussed above, the proposed framework provides a coherent and biologically meaningful description of the underlying processes governing butterfly populations. The consistent improvement in fit quality observed, together with the biologically plausible and stable estimates obtained for species-specific parameters such as T^* and σ_T, demonstrates that this modelling approach constitutes a viable alternative for investigating dispersal dynamics and their ecological implications, even at this early stage of development. As further ecological processes, such as the external source terms and finer-scale urbanization effects identified above, are incorporated alongside additional data, the framework is expected to become an increasingly robust tool for the quantitative analysis of butterfly populations, and to support biodiversity monitoring, habitat assessment, and conservation planning.

Beyond its methodological contribution, the framework developed in this thesis is motivated by a broader environmental concern. Urban green spaces are increasingly recognized as critical refugia for pollinator biodiversity, yet they are also among the habitats most directly and rapidly reshaped by human decision-making, through land-use conversion, fragmentation, pesticide application and the ongoing densification of cities. Quantifying source, sink, and trap dynamics at the scale of individual parks is therefore not merely a modelling exercise: it provides an evidence base that can inform how urban planners and policymakers manage green infrastructure, for instance by identifying which sites most urgently require ecological restoration or improved connectivity, and which interventions risk inadvertently converting a valuable habitat into a demographic trap. This concern is further sharpened by climate change, which is expected to alter flight periods, generation number, and range distributions of butterfly species. A dispersal-explicit, park-level model of the kind proposed here therefore offers a tool not only for understanding present-day metapopulation dynamics, but for anticipating how these dynamics may shift as climatic conditions and urban land use continue to change, and for guiding conservation and habitat-management decisions accordingly.

6.2 Future development

As this project remains under active development, the present thesis should be regarded as a preliminary investigation of the proposed framework, necessarily constrained by the time available for its completion. Several aspects that may play a significant role in shaping butterfly population dynamics have therefore not yet been incorporated into the current formulation, among which the treatment of voltinism and the closed-system assumption for Barcelona stand out as the most important limitations.

With regard to voltinism, as discussed in section 2.2, butterfly species can be classified as univoltine or multivoltine according to the number of generations they produce each year. The current formulation does not distinguish between these life-history strategies and is therefore unable to reproduce explicitly the occurrence of multiple generations within a single season. A natural extension of the model would be to incorporate voltinism directly into the mathematical formulation, primarily through the fly-period function $\alpha^s(T_x(t))$ in equation 3.4, which is currently represented by a single bell-shaped curve. Allowing this function to exhibit multiple peaks in future developments would provide a substantially more realistic representation of multivoltine species.

A second important limitation concerns the assumption that Barcelona behaves as a closed system. Throughout this work, dispersal has been modelled exclusively among the 24 urban parks included in the study area; in reality, however, the butterfly populations within the city are also influenced by surrounding natural areas (figure 6.1), particularly large green spaces such as Collserola, which likely act as important source habitats contributing individuals to the urban population.

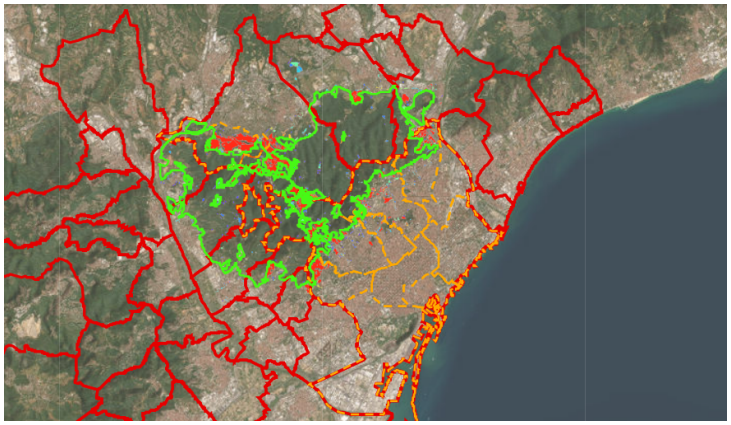


Figure 6.1: Map of the Barcelona coastal territory (orange boundary) and the surrounding natural areas (red and green for Collserola Park), which represent a major source of individuals for the system dynamics beyond the city limits [4].

Neglecting these external inputs may partially explain the discrepancies observed between simulated and empirical population dynamics throughout chapter 5, since an important component of the dispersal process is thereby omitted from the model. A natural direction for future research is thus the explicit incorporation of external source areas into the dispersal dynamics, for instance through an additional source term representing immigration from habitats outside the city. Since spatial data are already available for the Collserola Natural Park (section 2.17), this area offers a natural starting point for extending the model, before generalizing the same framework to other surrounding natural areas in order to obtain a more realistic description of the exchanges between the urban parks and the broader regional landscape.

Beyond these structural extensions, it would also be valuable to broaden the analysis to all species considered in this study, as only preliminary results, focused on *Pieris rapae*, have been obtained here due to time constraints. Such a comprehensive assessment should examine all parks identified as potentially problematic across species, rather than focusing on a single species in isolation, in order to determine whether certain parks consistently act as ecological traps regardless of the species involved, or whether different species perceive the same parks in markedly different ways. Cluster-based analyses would offer further insight in this respect: grouping species according to voltinism, and separately according to broader ecological traits such as degree of generalism, could help establish whether species sharing similar life-history strategies exhibit comparable responses to park characteristics. Extending the source-sink-trap analysis presented here for *Pieris rapae* to the remaining species considered in this study would ultimately allow the construction of a complete, species-integrated map of the parks in terms of their source, sink, and trap classification.

The implementation of these improvements is expected to substantially increase both the explanatory and predictive capabilities of the proposed framework, enabling a more realistic representation of butterfly population dynamics. By incorporating additional ecological processes and relaxing some of the simplifying assumptions adopted in the current formulation, the model will be better equipped to capture the complexity of species interactions with their environment and the mechanisms underlying dispersal across heterogeneous landscapes. As the framework continues to be refined and validated against empirical observations, it has the potential to become a robust quantitative tool for investigating metapopulation dynamics, habitat connectivity, and the effects of environmental change, ultimately supporting more reliable ecological analyses and providing valuable guidance for biodiversity monitoring, conservation planning, and the management of urban green infrastructure.

Appendix A

Mathematical Background

This appendix collects the mathematical definitions, theorems, and numerical schemes referenced throughout the thesis, in support of the model's formulation and analysis.

A.1 Graph theory

Definition 1 (Weighted directed graph)**.** *A weighted directed graph is a triple* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$*, where:*

- $\mathcal{V} = \{1, \dots, n\}$ *is a finite set of nodes (or vertices);*
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ *is the set of edges, where an ordered pair* $(x, y) \in \mathcal{E}$ *denotes the existence of a directed connection from node* x *to node* y*;*
- $\mathcal{W} : \mathcal{E} \to \mathbb{R}_{\geq 0}$ *assigns to each edge a non-negative weight* W_{xy}*, quantifying the strength, cost, or capacity associated with that connection.*

Unlike an undirected graph, a directed graph does not require $W_{xy} = W_{yx}$*; the two directions are, in general, independent [52, 53].*

Definition 2 (Adjacency matrix and weighted adjacency matrix)**.** *Given a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ *with* $|\mathcal{V}| = n$*, the adjacency matrix* $\mathbf{A} \in \{0,1\}^{n \times n}$ *is defined entry-wise by*

$$A_{xy} = \begin{cases} 1 & \text{if } (x, y) \in \mathcal{E}, \\ 0 & \text{otherwise,} \end{cases}$$

and encodes the structural connectivity of the graph independently of edge intensity. The weighted adjacency matrix $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$ *combines structure and intensity via the Hadamard (entrywise) product*

$$\mathbf{W} = \mathbf{A} \odot \mathbf{w}, \qquad W_{xy} = A_{xy}\, w_{xy},$$

where $w_{xy} \geq 0$ denotes the raw edge weight. In general, $\mathbf{W}$ *is not symmetric, i.e.* $W_{xy} \neq W_{yx}$*, reflecting the directed nature of* $\mathcal{G}$ *[52].*

Definition 3 (Row-stochastic matrix)**.** *A non-negative matrix* $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$ *is said to be row-stochastic if every row sums to one, i.e.*

$$\sum_{y=1}^{n} W_{xy} = 1 \qquad \forall\, x \in \{1, \ldots, n\}.$$

Row-stochastic matrices arise naturally as transition matrices of discrete-time Markov chains, where the entry W_{xy} *is interpreted as the conditional probability of transitioning from state* x *to state* y *in a single time step [54].*

Definition 4 (Markov chain on a graph)**.** *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ *be a weighted directed graph with row-stochastic weight matrix* $\mathbf{W}$*. The stochastic process* $(X_t)_{t \in \mathbb{N}}$*, taking values in* $\mathcal{V}$*, is a Markov chain on* $\mathcal{G}$ *if it satisfies the Markov property*

$$\mathbb{P}(X_{t+1} = y \mid X_t = x, X_{t-1}, \ldots, X_0) = \mathbb{P}(X_{t+1} = y \mid X_t = x) = W_{xy},$$

i.e. the future state depends on the past only through the present state. A probability distribution π *over* $\mathcal{V}$ *is called a stationary distribution of the chain if*

$$\pi \mathbf{W} = \pi, \qquad \sum_{x} \pi_x = 1, \qquad \pi_x \geq 0 \ \forall x$$

[54].

Definition 5 (Graph Laplacian)**.** *Let* $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$ *be a weighted (in general non-symmetric) adjacency matrix, and let* $\mathbf{D} = \operatorname{diag}(D_{11}, \ldots, D_{nn})$ *be the associated out-degree matrix, with*

$$D_{xx} = \sum_{y=1}^{n} W_{xy}.$$

The (weighted, directed) graph Laplacian is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}^T.$$

When $\mathbf{W}$ *is row-stochastic,* $\mathbf{D} = \mathbf{I}$*, and the Laplacian reduces to* $\mathbf{L} = \mathbf{I} - \mathbf{W}^T$*. The operator* $-\mathbf{L}$ *acting on a vector* $\mathbf{P} \in \mathbb{R}^n$ *measures, at each node, the deviation of the local value from the weighted average of its in-neighbours, and is the discrete analogue of the diffusion (Laplace) operator [52, 53].*

A.2 Softmax function

Definition 6 (Softmax function)**.** *Given a vector of real-valued scores (or negative costs) $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$, the softmax function $\sigma : \mathbb{R}^n \to \Delta^{n-1}$ maps z to a point on the probability simplex*

$$\Delta^{n-1} = \left\{ p \in \mathbb{R}^n_{\geq 0} : \sum_{i=1}^{n} p_i = 1 \right\},$$

via

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \qquad i = 1, \ldots, n$$

[55, 56].

Proposition 1 (Properties of the softmax function)**.** *The softmax function σ satisfies the following properties:*

1. ***Normalization:*** *$\sigma(z)_i \geq 0$ for all i, and $\sum_{i=1}^{n} \sigma(z)_i = 1$, so that $\sigma(z)$ defines a valid probability distribution over $\{1, \ldots, n\}$.*

2. ***Translation invariance:*** *for any $c \in \mathbb{R}$, $\sigma(z + c\mathbf{1}) = \sigma(z)$, where $\mathbf{1} = (1, \ldots, 1)$; only relative differences between components of z affect the resulting distribution.*

3. ***Order preservation:*** *σ is strictly increasing in each argument, i.e. $z_i > z_j \implies \sigma(z)_i > \sigma(z)_j$.*

4. ***Sensitivity to relative differences:*** *since $\sigma(z)_i / \sigma(z)_j = e^{z_i - z_j}$, the relative probability assigned to two outcomes depends exponentially on the difference between their scores, amplifying even small differences in z.*

5. ***Limiting behaviour:*** *as the gap $z_i - \max_{j \neq i} z_j \to +\infty$, $\sigma(z)_i \to 1$ and $\sigma(z)_j \to 0$ for $j \neq i$ (the distribution concentrates on the arg max); conversely, as all components of z become equal, $\sigma(z) \to (1/n, \ldots, 1/n)$ (the uniform distribution) [55].*

In the context of Chapter 3, the softmax is applied with $z_{xy} = -cost_{xy}$, so that destinations y associated with lower movement cost are assigned proportionally higher transition probability w_{xy}, while the normalization guarantees $\sum_y w_{xy} = 1$ for every origin x.

A.3 Existence and uniqueness of ODE solutions

Theorem 1 (Cauchy–Lipschitz / Picard–Lindelöf)**.** *Let $D \subseteq \mathbb{R} \times \mathbb{R}^n$ be open, and let $f : D \to \mathbb{R}^n$ be a continuous function that is locally Lipschitz continuous in its second argument, i.e. for every $(t_0, y_0) \in D$ there exist a neighbourhood $U \subseteq D$ of (t_0, y_0) and a constant $K > 0$ such that*

$$\|f(t, y_1) - f(t, y_2)\| \leq K\|y_1 - y_2\| \qquad \forall\, (t, y_1), (t, y_2) \in U.$$

Then, for every initial condition $(t_0, y_0) \in D$, there exists $\varepsilon > 0$ such that the initial value problem

$$\frac{dy}{dt} = f(t, y), \qquad y(t_0) = y_0,$$

admits a unique solution $y : (t_0 - \varepsilon, t_0 + \varepsilon) \to \mathbb{R}^n$. Moreover, this solution depends continuously on the initial condition y_0 [57, 58].

In particular, every linear system of the form

$$\frac{dy}{dt} = \mathbf{B}(t)y + b(t),$$

with $\mathbf{B}(t)$ and $b(t)$ continuous on an interval $\mathcal{T}$, satisfies the hypotheses of Theorem 1 globally on $\mathcal{T}$, since $f(t, y) = \mathbf{B}(t)y + b(t)$ is (globally) Lipschitz in y with Lipschitz constant $K = \|\mathbf{B}(t)\|$. This guarantees existence, uniqueness, and continuous dependence on initial data for the linearised population dynamics analysed in Section 3.4.

A.4 Perron–Frobenius theorem

Theorem 2 (Perron-Frobenius, non-negative matrix form)**.** *Let $\mathbf{W} \in \mathbb{R}^{n\times n}_{\geq 0}$ be a non-negative, irreducible matrix. Then:*

1. *$\mathbf{W}$ has a real eigenvalue $\lambda_{\max} > 0$, called the Perron root, satisfying $\lambda_{\max} \geq |\lambda|$ for every other eigenvalue λ of $\mathbf{W}$ (i.e. $\lambda_{\max}$ is the spectral radius of $\mathbf{W}$);*
2. *$\lambda_{\max}$ is a simple eigenvalue (algebraic multiplicity one);*
3. *there exists an eigenvector v associated with $\lambda_{\max}$ whose entries are all strictly positive, and this is, up to scalar multiples, the unique eigenvector of $\mathbf{W}$ with this property [59].*

Corollary A.4.1 (Row-stochastic matrices)**.** *If $\mathbf{W} \in \mathbb{R}^{n\times n}_{\geq 0}$ is row-stochastic, i.e. $\sum_y W_{xy} = 1$ for all x, then $\mathbf{W}\mathbf{1} = \mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)^T$; hence $\lambda = 1$ is an eigenvalue of $\mathbf{W}$ with eigenvector $\mathbf{1}$. Since every eigenvalue λ of a row-stochastic*

matrix satisfies $|\lambda| \leq 1$ *(by the Gershgorin circle theorem, each row defines a disc of radius* 1 *centred at* W_{xx}*, contained in* $\{z : |z| \leq 1\}$*), it follows that* $\lambda_{\max} = 1$ *is the dominant eigenvalue, with associated right eigenvector* $v_{\max} \propto \mathbf{1}$*. If, in addition,* $\mathbf{W}$ *is irreducible, Theorem 2 guarantees that* $\lambda = 1$ *is simple and that the corresponding left eigenvector* π *(satisfying* $\pi\mathbf{W} = \pi$*) can be normalized to a strictly positive probability vector, the unique stationary distribution of the associated Markov chain [54, 60].*

A.5 The L-BFGS-B Optimization Algorithm

This section briefly reviews the mathematical structure of the L-BFGS-B algorithm, used throughout this work (via `scipy.optimize.minimize`) to fit the model parameters to the neural-network predicted abundance data (Section 4).

Quasi-Newton methods and BFGS

Newton's method for unconstrained minimization requires the Hessian H_k of the objective $\mathcal{L}(\boldsymbol{\theta})$ at each iteration, which is prohibitive to compute here, since every evaluation of $\mathcal{L}$ requires a full RK4 integration of the metapopulation ODE system (Section 3.4). Quasi-Newton methods avoid this cost by building an approximation $B_k^{-1} \approx H_k^{-1}$ from gradient information alone. Given the parameter and gradient differences

$$s_k = \boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k, \qquad y_k = \nabla\mathcal{L}(\boldsymbol{\theta}_{k+1}) - \nabla\mathcal{L}(\boldsymbol{\theta}_k),$$

the BFGS update refines B_k^{-1} using s_k, y_k, and their inner product $y_k^\top s_k$, guaranteeing a positive-definite curvature estimate and superlinear convergence near a minimum, without ever forming the true Hessian [61].

The limited-memory variant (L-BFGS)

Storing the dense B_k^{-1} matrix is costly for problems with many parameters. L-BFGS retains only the most recent m pairs (s_k, y_k) (typically $m \sim 5$–20) and reconstructs the action of B_k^{-1} on the gradient implicitly via the two-loop recursion, reducing the cost per iteration from $O(n^2)$ to $O(mn)$ – a minor consideration here given the small parameter vector (four to six parameters per phase), but standard practice nonetheless [61].

Box constraints (the "B" in L-BFGS-B)

L-BFGS-B extends L-BFGS to handle simple bound constraints $\ell_i \leq \theta_i \leq u_i$, $i = 1, \ldots, n$, without requiring a general-purpose constrained solver. At each iteration, a generalized Cauchy point is computed by projecting the steepest-descent

direction onto the feasible box, identifying which variables are "active" at their bounds; a quasi-Newton step is then computed on the remaining free variables and projected back onto the box if necessary. This makes L-BFGS-B well suited to the present problem, where parameters are constrained to biologically plausible ranges (e.g. $T^\star \in [15{,}30]\,^\circ\mathrm{C}$, $\omega, \sigma_T > 0$, $\delta, \gamma, \rho \geq 0$) but no other constraints (e.g. linear or nonlinear relations between parameters) are required [62].

Practical implementation

In this work, the algorithm was used through the `scipy.optimize.minimize` library (`method='L-BFGS-B'`), with gradients of $\mathcal{L}(\boldsymbol{\theta})$ approximated numerically by finite differences, since no closed-form gradient of the RK4-integrated model output was available. Convergence was assessed via SciPy's default tolerance settings on the projected gradient norm and the relative decrease of the objective between iterations. As discussed in Section 4, because L-BFGS-B is a local optimizer, a multi-start strategy was layered on top of it to mitigate the risk of convergence to poor local minima [61, 62].

A.6 Mathematical formulation of the goodness-of-fit metrics

This appendix details the goodness-of-fit metrics used to evaluate the spatial performance of the metapopulation model against the neural network (NN) target dynamics. Let $P_x^s(t)$ denote the ODE-simulated abundance of *Pieris rapae* at site x on day t, and let $\hat{P}_x^s(t)$ denote the corresponding NN-predicted abundance, used as a proxy for the observed population dynamics. Let T be the total number of days considered (in this study, $T = 2373$, corresponding to approximately 6.5 years of daily data).

A.6.1 Mean Squared Error (MSE)

The site-level Mean Squared Error is defined as

$$\mathrm{MSE}_x = \frac{1}{T} \sum_{t=1}^{T} \left(P_x^s(t) - \hat{P}_x^s(t) \right)^2 . \tag{A.1}$$

This quantity measures the average squared deviation between the ODE trajectory and the NN target, expressed in units of abundance squared. Because it is not normalized, MSE_x is scale-dependent: sites with larger absolute population sizes will tend to exhibit larger MSE values even if the relative fit quality is comparable, which limits its usefulness for cross-site comparison [63].

A.6.2 Normalized Mean Squared Error (nMSE)

To allow scale-independent comparison across sites with heterogeneous population sizes, the residuals are rescaled by the site-specific maximum of the NN target trajectory:

$$\mathrm{nMSE}_x = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{P_x^s(t) - \hat{P}_x^s(t)}{\max_t \hat{P}_x^s(t)}\right)^2 . \tag{A.2}$$

Dividing each residual by $\max_t \hat{P}_x^s(t)$ before squaring constrains the contribution of each time step to a comparable range across sites, regardless of their absolute abundance scale. This is the loss function minimized during parameter fitting, since it prevents high-abundance sites from dominating the optimization at the expense of low-abundance ones.

Appendix B

Additional results

This appendix is devoted to additional result figures that were not discussed in the main text but were found to be of fundamental importance for the overall analysis.

B.1 Phase 0

The figure B.1 summarize the initial parameter values sampled across the 100 independent optimization runs.

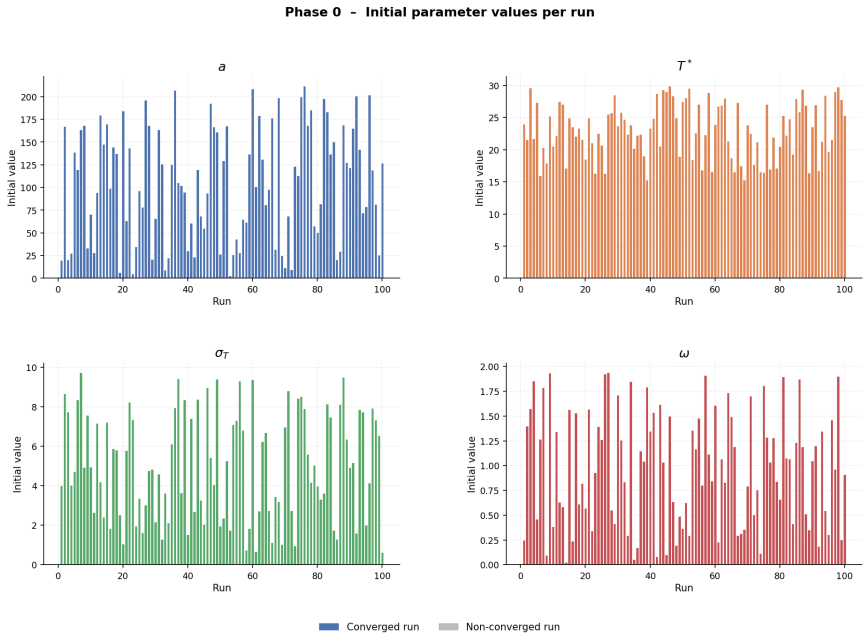


Figure B.1: Phase 0, initial values of the biological parameters a, T^*, σ_T and ω^s across the 100 optimisation runs.

While figure B.2 shows the distribution of these sampled values.

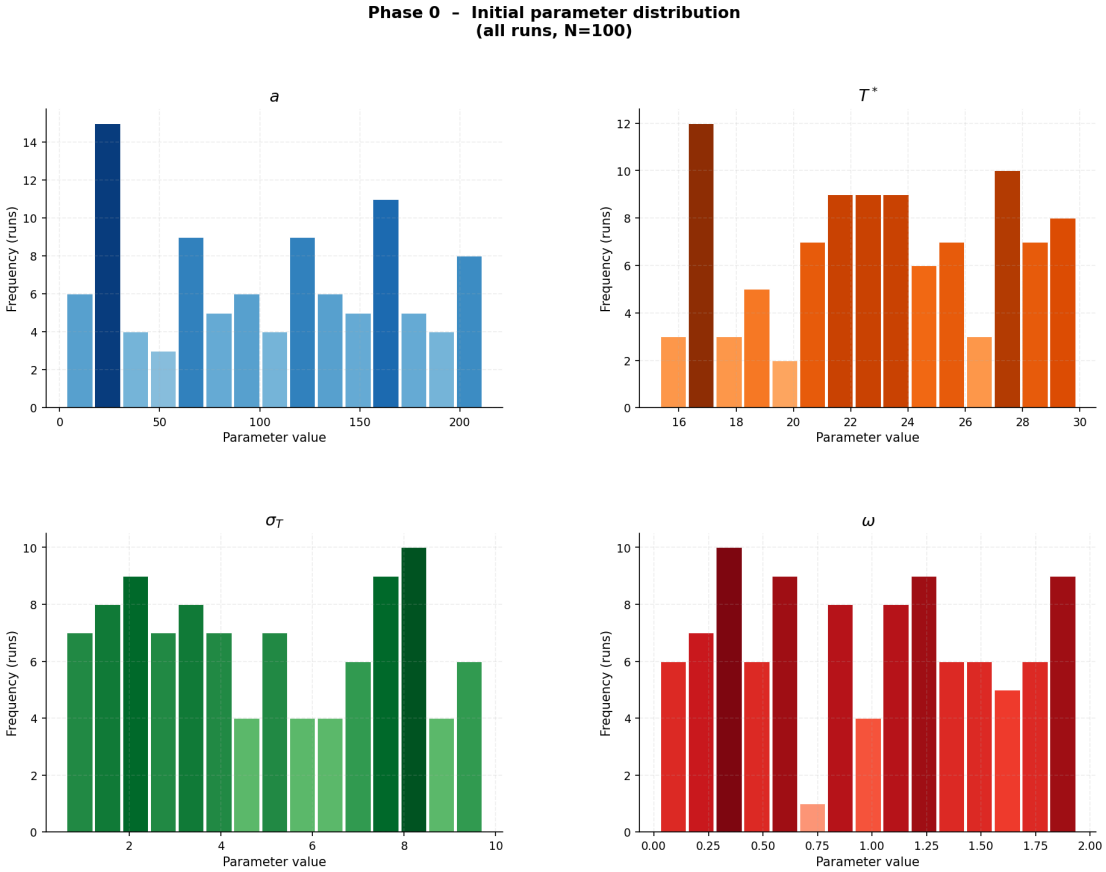


Figure B.2: Phase 0, distribution of the initial parameter values sampled across the 100 optimisation runs.

Once optimisation has converged, figure B.3 presents the resulting parameter estimates for all 100 runs, with bar colour encoding the normalised loss achieved in each run and a star marking the run that yielded the global best fit.

Taken together, these three figures allow the robustness and consistency of the optimisation procedure to be assessed already at this simplest, closed-system stage of the model.

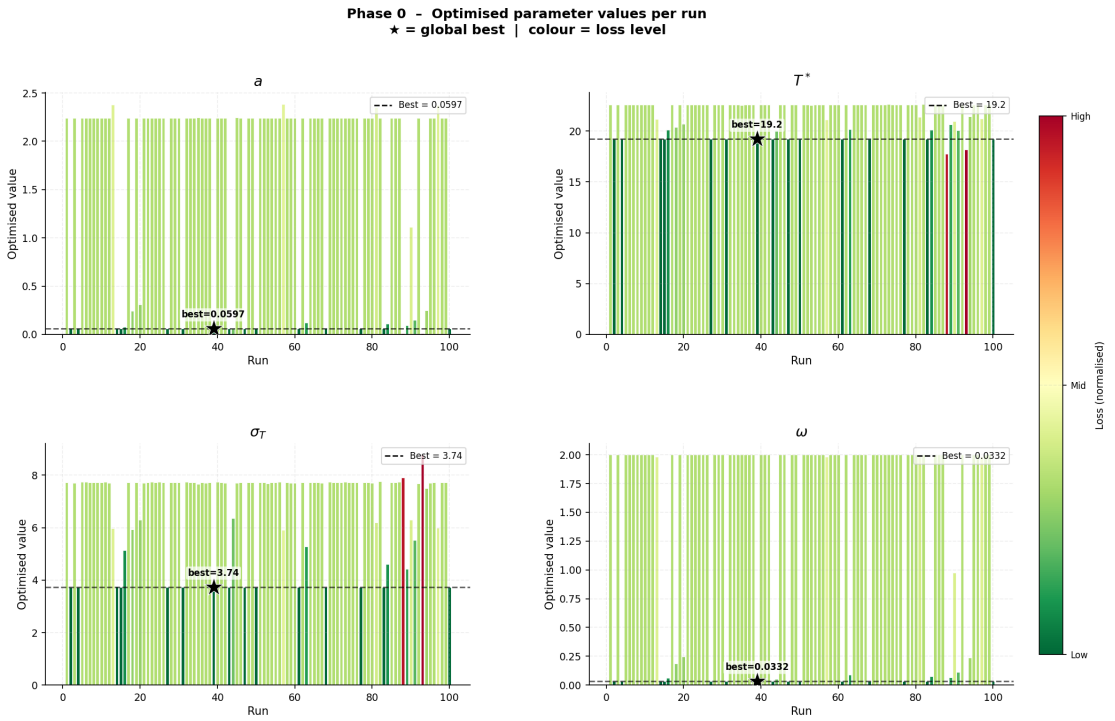


Figure B.3: Phase 0, optimised values of the biological parameters a, T^*, σ_T and ω^s across the 100 optimisation runs. Bar colour encodes the normalised loss of the corresponding run, and the star marks the run achieving the global best fit.

B.2 Phase 1

This section presents the corresponding optimisation results for phase 1, in which the biological parameters a, T^*, σ_T, ω^s and δ were estimated. It is worth noting that μ was initially included among the parameters to be estimated, since no empirical data were available at the start of the simulations; once reliable data for this parameter became available, however, μ was subsequently kept fixed at a value representative of the species as a whole ($\mu = 1$ for *Pieris rapae*).

As in phase 0, the same three-stage view of the optimisation process is provided here: figure B.4 reports the initial parameter values assigned across the 100 optimisation runs.

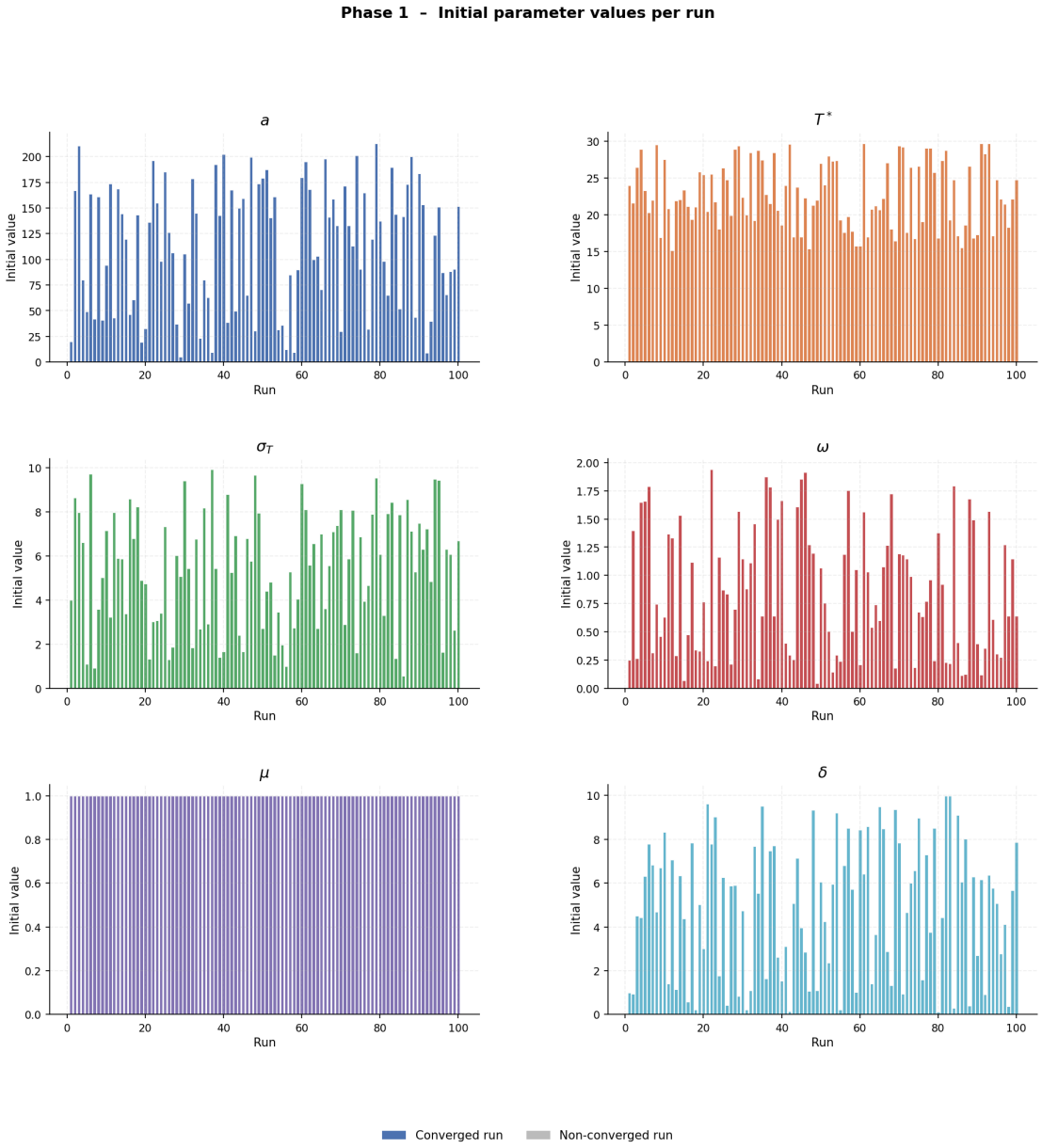


Figure B.4: Phase 1, initial values of the biological parameters a, T^*, σ_T, ω^s, μ and δ across the 100 optimisation runs.

Figure B.5 shows the distribution of these sampled initial values.

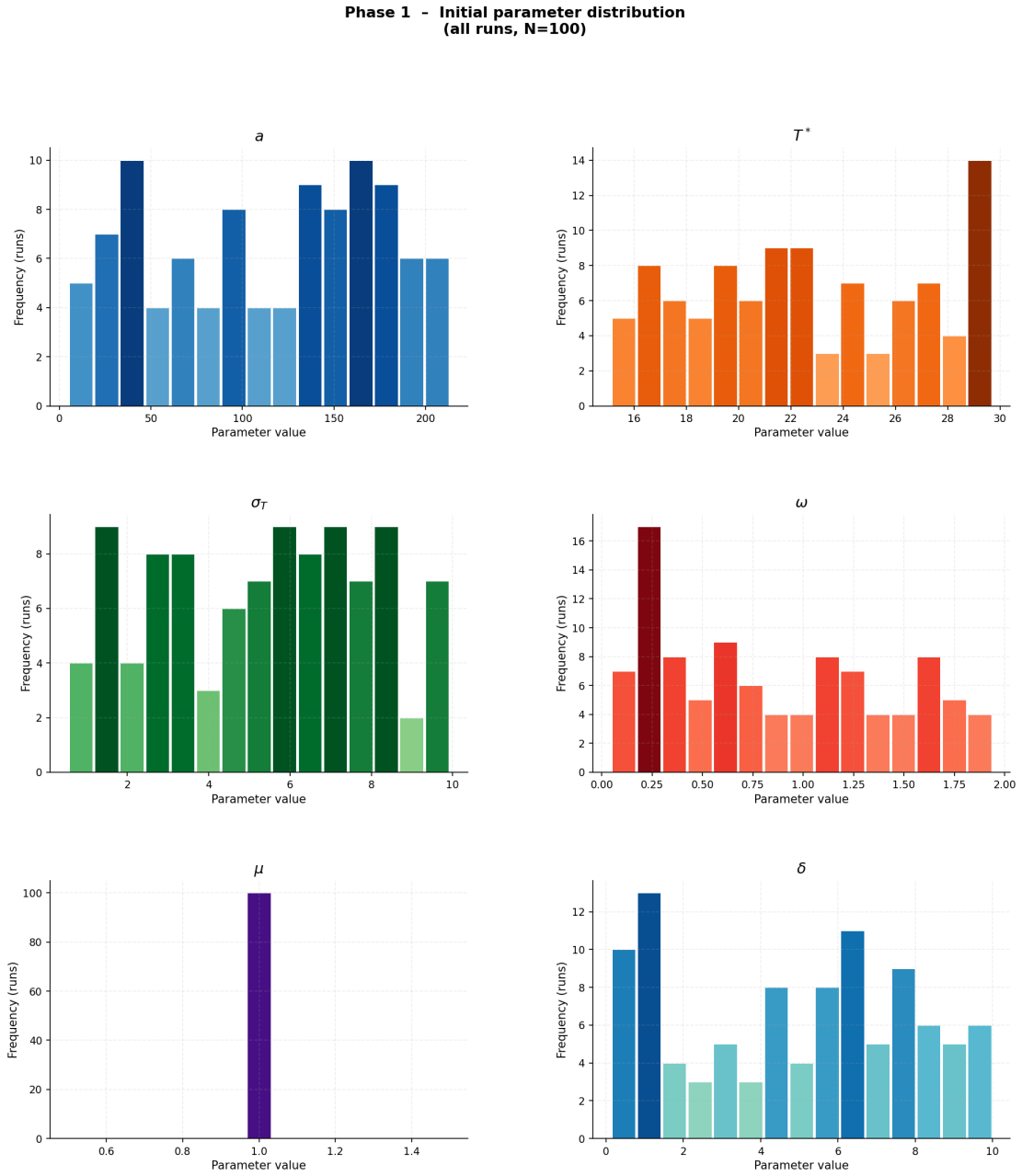


Figure B.5: Phase 1, distribution of the initial parameter values sampled across the 100 optimisation runs.

Instead, figure B.6 presents the optimised solutions obtained from each run. Together, these figures allow the robustness of the optimisation process to be assessed under the additional complexity introduced by the dispersal parameter δ.

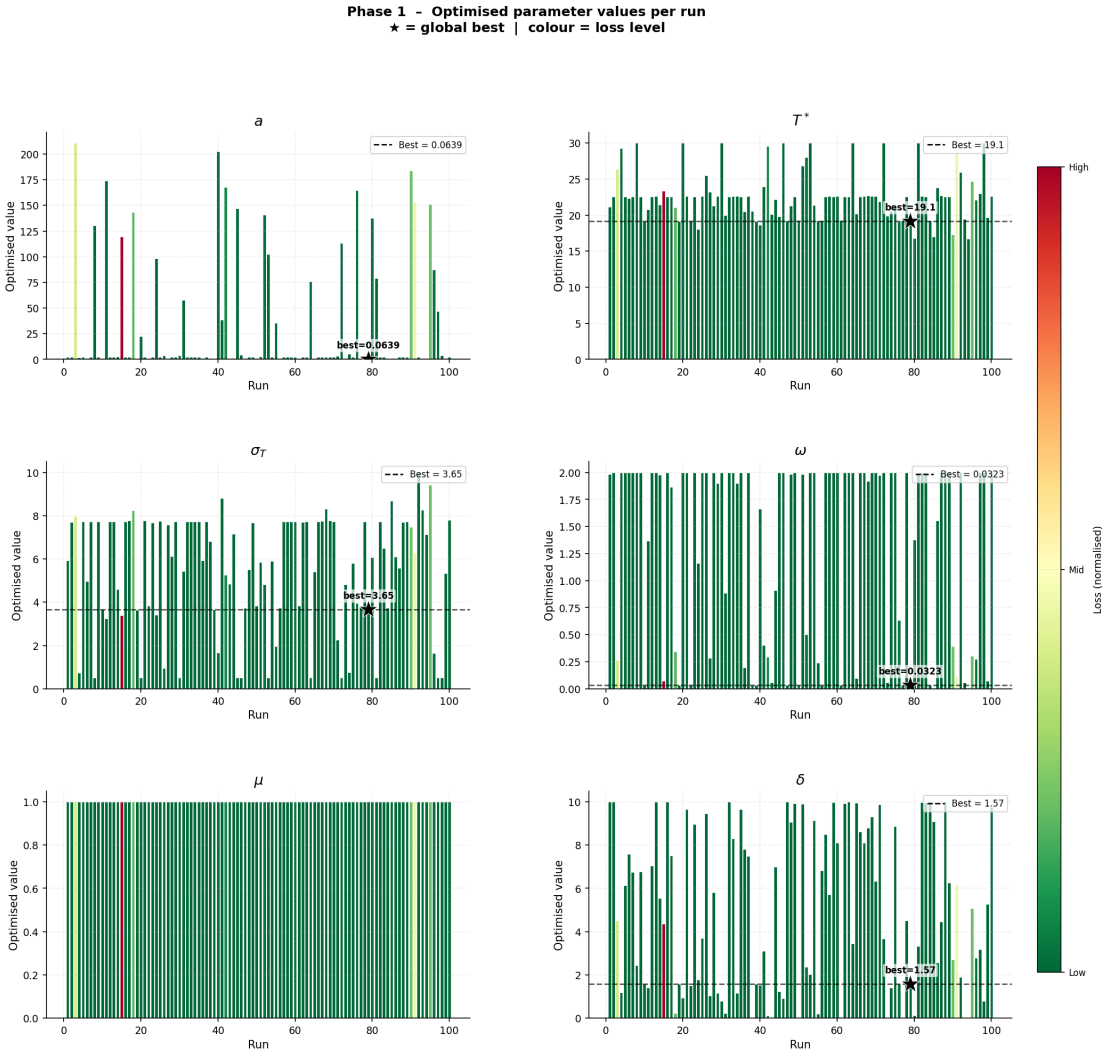


Figure B.6: Phase 1, optimised values of the biological parameters a, T^*, σ_T, ω^s, μ and δ across the 100 optimisation runs.

B.3 Phase 2

This final section summarises the optimisation results for phase 2, corresponding to the most comprehensive parameter estimation scenario considered in this work. As for the previous two phases, figure B.7 shows the initial parameter values assigned to each of the 100 optimisation runs.

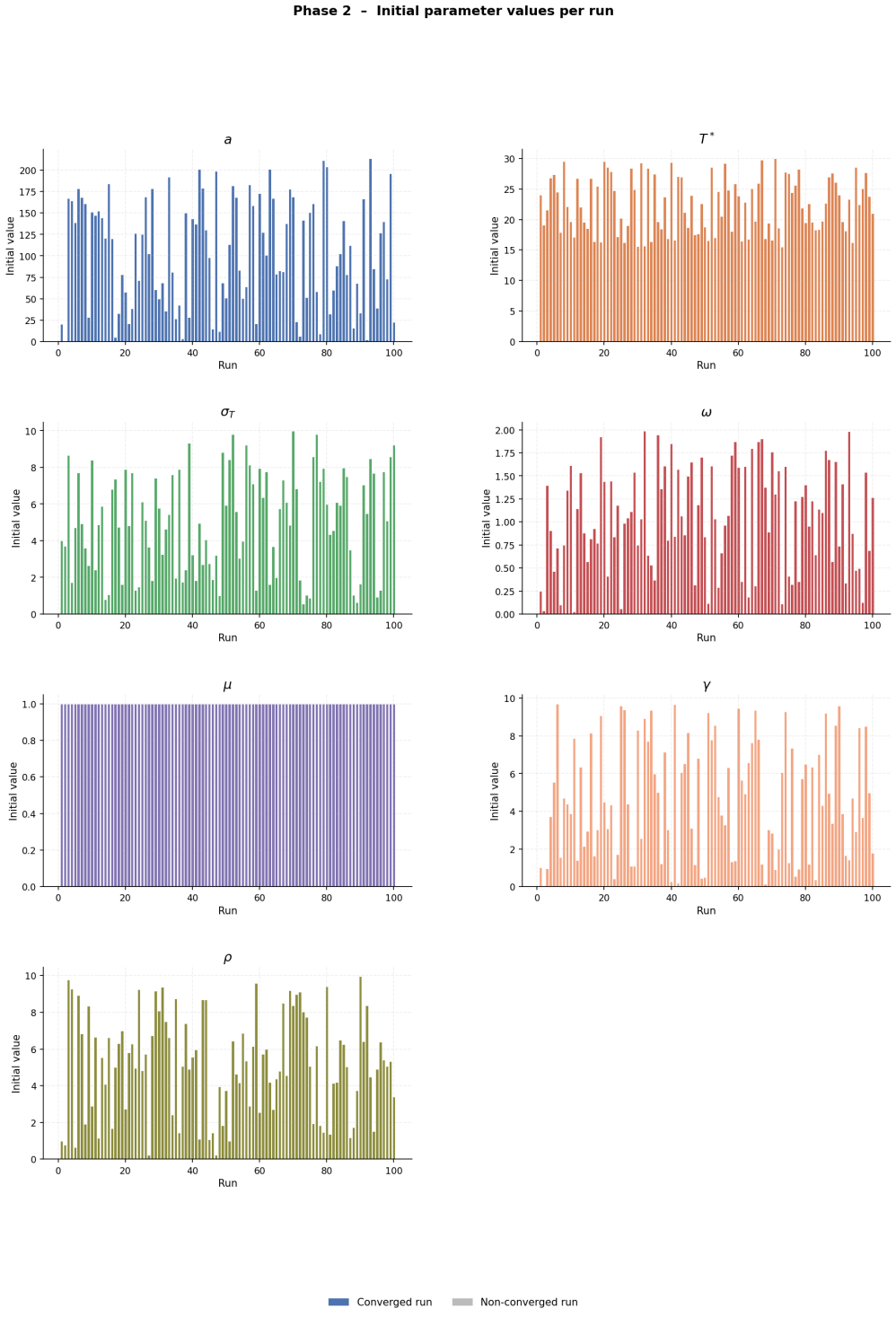


Figure B.7: Phase 2, initial values of the biological parameters a, T^*, σ_T, ω^s, μ, γ and ρ across the 100 optimisation runs.

Figure B.8 presents the distribution of the sampled initial values for all estimated parameters.

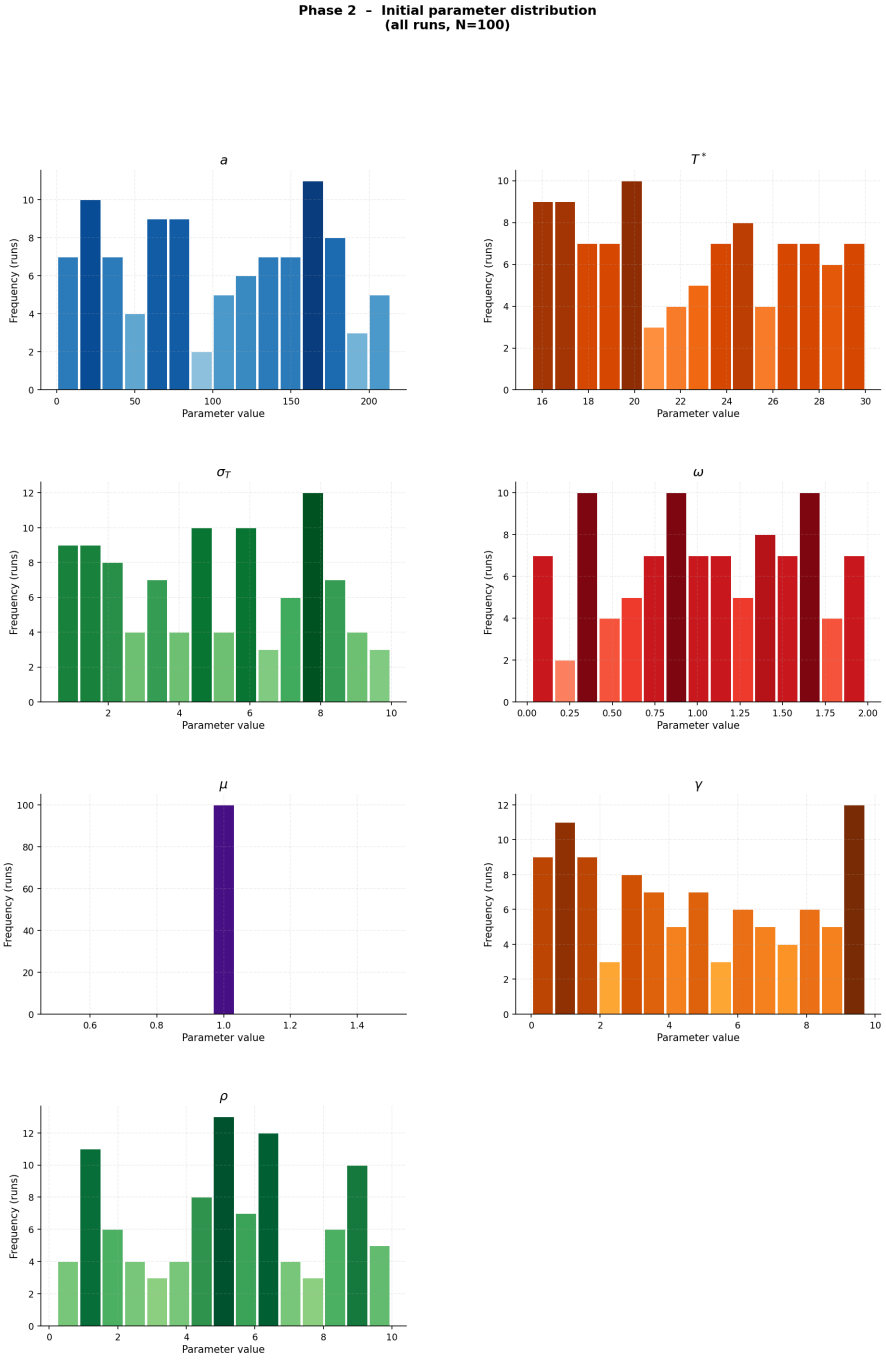


Figure B.8: Phase 2, distribution of the initial parameter values sampled across the 100 optimisation runs.

Finally figure B.9 reports the corresponding optimised parameter estimates.

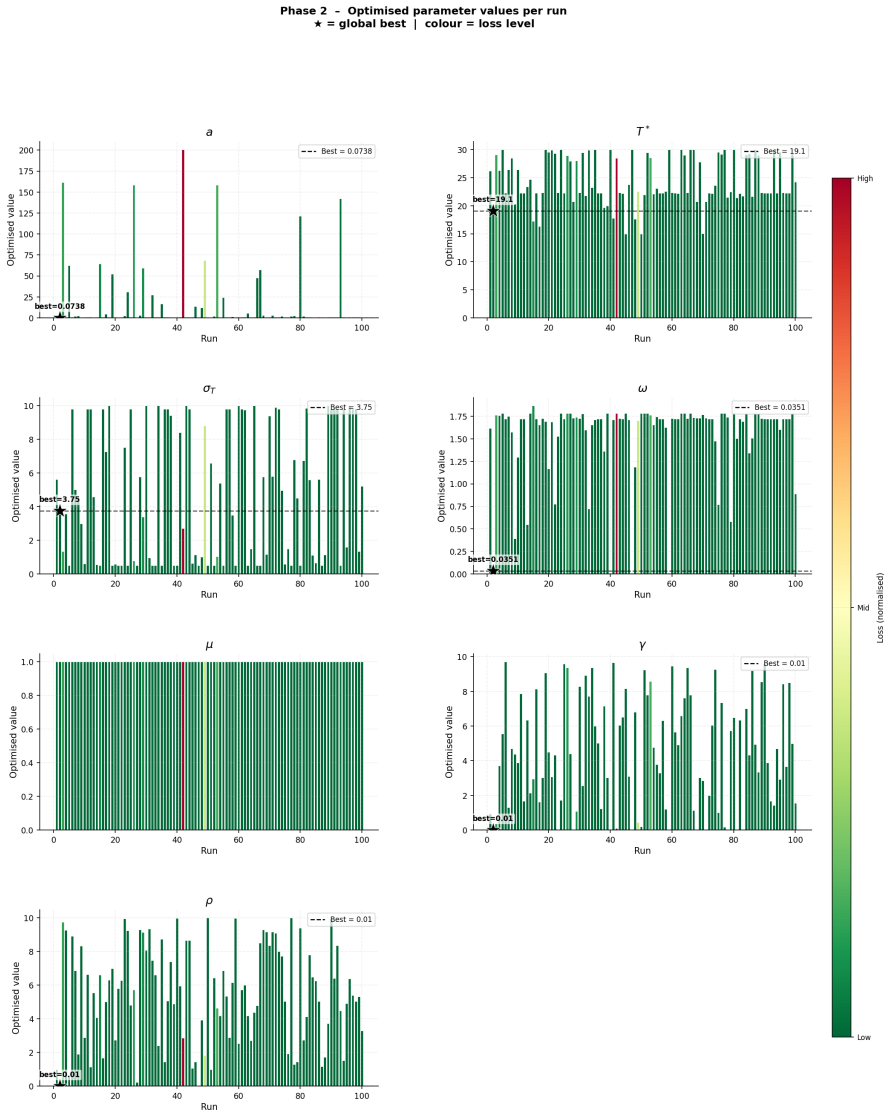


Figure B.9: Phase 2, optimised values of the biological parameters a, T^*, σ_T, ω^s, μ, γ and ρ across the 100 optimisation runs. Bar colour encodes the normalised loss of the corresponding run, and the star marks the run achieving the global best fit.

These results provide additional evidence of the stability and reproducibility of the optimisation framework even when the full set of biological and dispersal-related parameters, a, T^*, σ_T, ω^s, μ, γ and ρ, is estimated simultaneously.

Bibliography

[1] Clàudia Pla-Narbona, Constantí Stefanescu, Joan Pino, Francisco J. Cabrero-Sañudo, Enrique García-Barros, Miguel L. Munguira, and Yolanda Melero. Butterfly biodiversity in the city is driven by the interaction of the urban landscape and species traits: a call for contextualised management. *Landscape Ecology*, 37(1):81–92, 2022.

[2] Martino Pesaresi and Panagiotis Politis. GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030), 2023. PID: http://data.europa.eu/89h/9f06f36f-4b11-47ec-abb0-4f8b7b1d72ea.

[3] S. Novell-Masot, P. Colom, Y. Melero, and G. Dalmasso. A neural network framework for modelling seasonal count time series from standardised ecological monitoring schemes. Manuscript in preparation, 2026.

[4] Consorci del Parc Natural de la Serra de Collserola. Parc natural de la serra de collserola.

[5] Guangzhao Chen, Xia Li, Xiaoping Liu, Yimin Chen, Xun Liang, Jiye Leng, Xiaocong Xu, Weilin Liao, Yue'an Qiu, Qianlian Wu, and Kangning Huang. Global projections of future urban land expansion under shared socioeconomic pathways. *Nature Communications*, 11(1):537, Jan 2020.

[6] Karen C. Seto, Burak Güneralp, and Lucy R. Hutyra. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40):16083–16088, 2012.

[7] Christina Cobbold. Spatial ecology: Lecture 1, metapopulations.

[8] Ilkka Hanski and Otso Ovaskainen. The metapopulation capacity of a fragmented landscape. *Nature*, 404:755–758, 2000.

[9] Yolanda Melero, Constantí Stefanescu, Stephen C.F. Palmer, Justin M.J. Travis, and Joan Pino. The role of the urban landscape on species with contrasting dispersal ability: Insights from greening plans for barcelona. *Landscape and Urban Planning*, 195:103707, 2020.

[10] Daniel Sol, Yolanda Melero, Lisieux Fuzessy, and César González-Lagos. Urbanization. In Bob B. M. Wong and Ulrika Candolin, editors, *Behavioural Responses to a Changing World*. Oxford University Press, 12 2024.

[11] Otso Ovaskainen and Ilkka Hanski. How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theoretical Population Biology*, 64(4):481–495, 2003.

[12] Jelena H. Pantel, Thomas Lamy, Maxime Dubart, Jean-Pierre Pointier, Philippe Jarne, and Patrice David. Metapopulation dynamics of multiple species in a heterogeneous landscape. *Ecological Monographs*, 92(3):e1515, 2022.

[13] Robin Hale and Stephen Swearer. Ecological traps: Current evidence and future directions. *Proceedings of the Royal Society B: Biological Sciences*, 283:20152647, 02 2016.

[14] Bruce A. Robertson and Richard L. Hutto. A framework for understanding ecological traps and an evaluation of existing evidence. *Ecology*, 87(5):1075–1085, 2006.

[15] James Battin. When good animals love bad habitats: Ecological traps and the conservation of animal populations. *Conservation Biology*, 18(6):1482–1491, 2004.

[16] Joern Fischer and David B. Lindenmayer. Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, 16(3):265–280, 2007.

[17] Philip J. Harrison, Ilkka Hanski, and Otso Ovaskainen. Bayesian state-space modeling of metapopulation dynamics in the glanville fritillary butterfly. *Ecological Monographs*, 81(4):581–598, 2011.

[18] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl_3):7280–7287, 2002.

[19] Julia Bicker, René Schmieding, Michael Meyer-Hermann, and Martin J. Kühn. Hybrid metapopulation agent-based epidemiological models for efficient insight on the individual scale: A contribution to green computing. *Infectious Disease Modelling*, 10(2):571–590, 2025.

[20] Dirk Helbing, Illés Farkas, and Tamás Vicsek. Simulating dynamic features of escape panic. *Nature*, 407:487–490, 09 2000.

[21] Fernando Huancas, Anibal Coronel, Esperanza Lozada, and Jorge Torres. Mathematical modeling of population dynamics of pollinators: A survey. *Biology*, 14(9), 2025.

[22] CREAF – Centre for Ecological Research and Forestry Applications. urban Butterfly Monitoring Scheme (uBMS), 2018.

[23] Butterfly Conservation Europe and UK Centre for Ecology & Hydrology. European Butterfly Monitoring Scheme (eBMS), 2014.

[24] Constanti Stefanescu, Jofre Carnicer, and Josep Peñuelas. Determinants of species richness in generalist and specialist mediterranean butterflies: the negative synergistic forces of climate and habitat change. *Ecography*, 34(3):353–363, 2011.

[25] Yolanda Melero, Constanti Stefanescu, and Joan Pino. General declines in mediterranean butterflies over the last two decades are modulated by species traits. *Biological Conservation*, 201:336–342, 09 2016.

[26] Marco Moretti, André T. C. Dias, Francesco de Bello, Florian Altermatt, Steven L. Chown, Francisco M. Azcárate, James R. Bell, Bertrand Fournier, Mickaël Hedde, Joaquín Hortal, Sébastien Ibanez, Erik Öckinger, José Paulo Sousa, Jacintha Ellers, and Matty P. Berg. Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. *Functional Ecology*, 31(3):558–567, 2017.

[27] Thomas Merckx and Hans Van Dyck. Urbanization-driven homogenization is more pronounced and happens at wider spatial scales in nocturnal and mobile flying insects. *Global Ecology and Biogeography*, 28(10):1440–1455, 2019.

[28] Romain Julliard, Joanne Clavel, Vincent Devictor, Frédéric Jiguet, and Denis Couvet. Spatial segregation of specialists and generalists in bird communities. *Ecology Letters*, 9(11):1237–1244, 2006.

[29] Constanti Stefanescu, Jofre Carnicer, and Josep Peñuelas. Determinants of species richness in generalist and specialist mediterranean butterflies: the negative synergistic forces of climate and habitat change. *Ecography*, 34(3):353–363, 2011.

[30] Albert Fernández-Chacón, Constantí Stefanescu, Meritxell Genovart, James D. Nichols, James E. Hines, Ferran Páramo, Marco Turco, and Daniel

Oro. Determinants of extinction-colonization dynamics in mediterranean butterflies: the role of landscape, climate and local habitat features. *Journal of Animal Ecology*, 83(1):276–285, 2014.

[31] Oliver Schweiger, Alexander Harpke, Martin Wiemers, and Josef Settele. Climber: Climatic niche characteristics of the butterflies in europe. *ZooKeys*, 367:65–84, 01 2014.

[32] Mikko Kuussaari, Matias Saarinen, Eeva-Liisa Korpela, Juha Pöyry, and Terho Hyvönen. Higher mobility of butterflies than moths connected to habitat suitability and body size in a release experiment. *Ecology and Evolution*, 4(19):3800–3811, 2014.

[33] Andrew J. Suggitt, Constantí Stefanescu, Ferran Páramo, Tom Oliver, Barbara J. Anderson, Jane K. Hill, David B. Roy, Tom Brereton, and Chris D. Thomas. Habitat associations of species show consistent but weak responses to climate. *Biology Letters*, 8(4):590–593, 04 2012.

[34] Anne Eskildsen, Luísa G. Carvalheiro, W. Daniel Kissling, Jacobus C. Biesmeijer, Oliver Schweiger, and Toke T. Høye. Ecological specialization matters: long-term trends in butterfly species richness and assemblage composition depend on multiple functional traits. *Diversity and Distributions*, 21(7):792–802, 2015.

[35] Joe Middleton-Welling, Rachel A. Wade, Roger L. H. Dennis, Leonardo Dapporto, and Tim G. Shreeve. Optimising trait and source selection for explaining occurrence and abundance changes: A case study using british butterflies. *Functional Ecology*, 32(6):1609–1619, 2018.

[36] R. Vila, C. Stefanescu, and J. M. Sesma. *Guia de les papallones diürnes de Catalunya*. Lynx Edicions, Barcelona, 2018.

[37] M. O. Hill and A. J. E. Smith. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon*, 25(2/3):249–255, 1976.

[38] EPSG Geodetic Parameter Dataset. EPSG:3035 – ETRS89-extended / LAEA Europe. https://epsg.io/3035, 2024.

[39] S. D. Graves and A. M. Shapiro. Exotics as host plants of the california butterfly fauna. 110:413–433, 2003.

[40] Ilkka Hanski. Metapopulation dynamics. *Nature*, 396(6706):41–49, nov 1998.

[41] Marie-Hélène Lizée, Jean-François Mauffrey, Thierry Tatoni, and Magali Deschamps-Cottin. Monitoring urban environments on the basis of biological traits. *Ecological Indicators*, 11:353–361, 03 2011.

[42] Marie-Hélène Lizée, Thierry Tatoni, and Magali Deschamps-Cottin. Nested patterns in urban butterfly species assemblages: Respective roles of plot management, park layout and landscape features. *Urban Ecosystems*, pages 1–20, 03 2016.

[43] European Commission, Joint Research Centre. Global human settlement layer (ghsl), 2026. https://human-settlement.emergency.copernicus.eu/.

[44] M. Pesaresi, M. Schiavina, P. Politis, S. Freire, K. Krasnodębska, J. H. Uhl, A. Carioli, C. Corbane, L. Dijkstra, P. Florio, H. K. Friedrich, J. Gao, S. Leyk, L. Lu, L. Maffenini, I. Mari-Rivero, M. Melchiorri, V. Syrris, J. Van Den Hoek, and T. Kemper. Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data. *International Journal of Digital Earth*, 17(1), 2024.

[45] Yolanda Melero, Constantí Stefanescu, and Joan Pino. General declines in mediterranean butterflies over the last two decades are modulated by species traits. *Biological Conservation*, 201:336–342, 2016.

[46] Python Software Foundation. Python language reference, 2026.

[47] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with numpy. *Nature*, 585:357–362, 2020.

[48] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272, 2020.

[49] Wes McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010.

[50] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[51] Ai SHIBATA, Kazuhiko WATANABE, Masanobu YOSHIO, and Minoru ISHII. Daily activity of the small white butterfly, pieris rapae crucivora boisduval (lepidoptera : Pieridae), in an artificial light cycle. *Japanese Journal of Entomology (New Series)*, 5(3):59–69, 2002.

[52] Ravindra B. Bapat. *Graphs and Matrices*. Universitext. Springer, London, 2010.

[53] Fan R. K. Chung. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI, 1997.

[54] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.

[55] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.

[56] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, 2016.

[57] Morris W. Hirsch, Stephen Smale, and Robert L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press, Waltham, MA, 3rd edition, 2012.

[58] Earl A. Coddington and Norman Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, New York, 1955.

[59] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA, 2000.

[60] Hal Caswell. *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer Associates, Sunderland, MA, 2nd edition, 2001.

[61] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2nd edition, 2006.

[62] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[63] Norman R. Draper and Harry Smith. *Applied Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, New York, 3rd edition, 1998.

Acknowledgements

Ringraziare è sempre difficile, soprattutto se si è in un caso come il mio, ovvero circondata da persone meravigliose.

In primis grazie al professore Giovanni, per questa opportunità e per il suo modo di studiare con fervore qualsiasi cosa. Alla professoressa Yolanda per aver portato le farfalle nella mia vita, insieme a un clima accademico idilliaco. Al professor Preziosi, grazie per essere l'umanità del Politecnico.
Sempre e comunque in primis, alla mia famiglia, a Piera, Massimo e Marco, un grazie colmo di affetto. Mi ripeterò ma vi ringrazio per avermi resa ciò che sono e per avermi supportata per tutti questi anni; questa laurea è anche soprattutto vostra.
A un altro membro della mia famiglia, a Carola a cui devo tutto e per la quale un grazie non sarà mai abbastanza.
Ai miei amici, non farò un elenco di nomi perché sarebbe davvero lungo, purtroppo sono una persona fortunata circondata da tante persone stupende di cui sono eternamente grata. Spero che nonostante non veniate citati esplicitamente sentiate tutta la gratitudine che provo nei vostri confronti e spero di riuscire a ringraziarvi non solo in una tesi, ma ogni giorno standovi accanto.
A Barcellona, a chi l'ha resa speciale. A chi, a Chieri e dintorni, è sempre restato negli anni e a chi nonostante tutto rimane sempre come se ci fossimo appena conosciuti. A chi ha reso il Politecnico un posto felice e a chi fuori dal Politecnico è diventato per me un posto felice. A chi nonostante la distanza, è sempre più vicino e sa far sparire chilometri di spazio con amore puro. A chi sa migliorarmi ogni giorno, a chi mi mostra la bellezza della vita ogni giorno e a chi ogni giorno continua a farmi meravigliare di ogni cosa, sostenendomi.
Spero vi siate ritrovati tutti in qualcosa perchè io vi ritrovo continuamente in tutta la mia vita e in questa tesi soprattutto. Senza di voi non avrei mai avuto il coraggio di godermi a pieno questa laurea e questa esistenza.

Grazie, infinitamente.