POLITECNICO DI TORINO

Master's Degree in ENVIRONMENTAL AND LAND ENGINEERING, CLIMATE CHANGE

Master's Degree Thesis

Climate Change and Nipah virus: exploring the links between climate variability, extremes, and zoonotic spillover in Bangladesh

Supervisor
Prof. Jost-Diedrich GRAF VON HARDENBERG
Co-Supervisors
Prof. Marino GATTO
Prof. Isabella CATTADORI

Candidate
Agnese CIMAROLI

OCTOBER 2024

Abstract

Zoonotic infectious diseases, transmitted from wild or domestic animals into humans, represent a major threat for public health as they can cause serious disease outbreaks and potentially escalate into pandemics, as recently observed for SARS-2-Covid-19. Climate change and anthropogenic ecosystem alterations are amplifying this risk by reshaping interactions between wildlife, pathogens, and human populations. Consequently, “jump-zones”-regions where zoonotic spillovers are increasingly likely-are fast expanding, particularly at tropics.

Among bat-borne zoonoses, Nipah virus (NiV) is recognized by the World Health Organization as one of the top 10 priority diseases, causing near-annual outbreaks in Bangladesh, a region highly vulnerable to climate change impacts. Despite the critical role bats play in pathogen transmission, their responses to climatic factors remains largely underexplored, highlighting the urgent need for focused research, also in view of the current climate change.

This thesis investigates the relationship between relevant climatic factors, including extreme weather events, and Nipah virus spillover occurrences in Bangladesh, where fruit bats (Pteropus species) serve as natural reservoirs of the virus. Through detailed spatio-temporal analysis, coupled with a regression analysis using Generalized Linear Models (GLMs), this study aims to identify key climate variables influencing the frequency and spatial distribution of spillover events occurred over a 18-year period (2001-2018). Furthermore, addressing gaps identified in the literature, this study extends the temporal focus beyond the typical winter months to examine year-round climate dynamics, using ERA5 reanalysis data, which provide higher spatial resolution and more reliable climate information for tropical regions than traditional sources.

Key findings reveal that specific climate conditions—colder winters, warmer and drier monsoons, and reduced post-monsoon rainfall—are strongly associated with a surge in NiV spillover events. These climatic stressors likely disrupt bat foraging and increase thermoregulatory stress, which are known to drive recrudescence and shedding of virus. Notably, the observed correlation between monsoon conditions and the frequency of spillover events is a novel finding for the Nipah virus in Bangladesh, although similar pattern have been observed with Hendra virus, another paramyxovirus, in Australia. Additionally, the spatial analysis identified two distinct geographic clusters within the Nipah Belt, each defined by specific climatic and environmental stressors. This underscores the spatial heterogeneity of spillover risk drivers and emphasizes the importance of targeted, region-specific interventions to mitigate spillover risks.

Given the accelerated pace of climate change impacts, and its complex interaction with bat behaviour and viral dynamics, a deeper understanding of the climatic triggers associated with spillover events is crucial. Such knowledge could enhance predictive models, supporting the design of targeted interventions to prevent or mitigate NiV outbreaks, particularly in high-risk regions.

Acknowledgements

Firstly, I would like to thank my supervisor, Jost-Diedrich Graf Von Hardenberg, for supporting my interest in this topic, generously sharing his expertise, and dedicating valuable time to guide me through this work.

I am also grateful to Professor Marino Gatto and Prof. Isabella Cattadori for their advice, support, and for continuously sharing their insights and experience in developing this research. Despite the physical distance, with Prof. Cattadori in the U.S. and Professor Gatto at the Politecnico di Milano, they always made time to guide me, and I truly appreciate their effort.

Finally, I really appreciate the opportunity to explore an idea that was important to me, seeing it takes shape and grow in a multidisciplinary context, where different perspectives enriched the project. Working on something I was passionate about has been deeply rewarding, and I look forward to contributing to this field in the future, to the best of my abilities.

Table of Contents

Introduction **1**

Chapter 1. Anthropogenic changes drive the emergence and spread of zoonosis: Nipah virus **3**

1.1 Wild mammals and zoonotic infections . . . 3
1.2 Bats: Unique mammalian virus host . . . 4
1.3 South and Southeast Asia hot-spots of infections . . . 5
1.4 Nipah Virus . . . 5
  1.4.1 Etiology . . . 6
  1.4.2 Uniqueness of Nipah virus infectious events . . . 6
  1.4.3 Why a major concern . . . 6
  1.4.4 Natural reservoir of NiV . . . 7
1.5 Anthropogenic pressure changes bat ecology and increases spillover risk . . 11
  1.5.1 Redistribution of resources in space and time . . . 12
  1.5.2 Altered connectivity . . . 13
  1.5.3 Enabling conditions for virus spillover from bats to recipient hosts 15
1.6 Climate change and zoonosis: Nipah virus . . . 19
  1.6.1 Climate Change in South Asia . . . 20
  1.6.2 Focus: climate's role in driving zoonotic disease emergence and transmission . . . 22
  1.6.3 Drivers of bat-borne zoonosis in Bangladesh: Focus on the Nipah virus . . . 24
1.7 Limits identified in NiV research and their influence on thesis choices . . 28

Chapter 2. The Bangladesh case **30**

2.1 Description of the study region . . . 30
  2.1.1 Climate of Bangladesh . . . 30
  2.1.2 Climate change exposure . . . 32
  2.1.3 Demographics and economy . . . 33
  2.1.4 Land-use and Nipah belt . . . 35
2.2 Spatio-temporal reconstruction of Nipah virus spillover events in Bangladesh (2001-2018) . . . 36
2.3 Preliminary data analysis . . . 37

Chapter 3. Climate analysis **41**

3.1 Analysis of climate anomalies . . . 41
  3.1.1 Climate data . . . 42
  3.1.2 Data pre-processing steps . . . 43
  3.1.3 Methods . . . 44

3.1.4 Results 48
3.1.5 Discussion 49
3.2 Analysis of weather extremes 52
3.2.1 Methods 54
3.2.2 Results 56
3.2.3 Discussion 58

Chapter 4. Spatial analysis of environmental stressors 63
4.1 Spatial distribution of environmental stressors 63
4.1.1 Data and data pre-processing 64
4.2 Spatial clustering of districts 69
4.2.1 Methods 69
4.2.2 Results and Discussion 72

Chapter 5. Regression analysis of climate variables, extremes and spillover events within spatial clusters 77
5.1 Regression analysis: Generalized Linear Models 77
5.1.1 Data and data pre-processing 77
5.1.2 Methods 80
5.1.3 Results - Case A: single spatial cluster 85
5.1.4 Discussion - Case A 87
5.1.5 Results - Case B: two spatial clusters 93
5.1.6 Discussion - Case B 95
5.2 Climate drivers of spillover events: current findings and future projections 99

Conclusions 102

Appendix A. 106
A.1 Climate variability: ENSO 106
A.2 Bangladesh Case 107
A.2.1 Detailed insights into embolism and Xylem refilling phenomenon . 107
A.2.2 Reconstruction of Nipah virus spillover events 108
A.2.3 Data analysis Tools and Methods for annual spillover distribution 117
A.3 Climate analysis 119
A.3.1 Trend analysis 119
A.3.2 Extreme indexes 121
A.4 Spatial analysis 126
A.4.1 Detailed methodology of K-means clustering process 126
A.5 Regression analysis: Generalized Linear Models (GLMs) 129
A.5.1 Steps in the univariate GLM process 131
A.5.2 GLMs evaluation 137
A.5.3 GLMs results - Case A: single spatial cluster 140
A.5.4 GLMs results - Case B: two spatial clusters 145

Bibliography 150

List of Tables

2.1 Annual count of NiV cases for year range 2001-2018. Sources: WHO, IECDR. 36

2.2 Annual count of spillover events considered in this study. Sources: [33], [59], [94]. 37

2.3 Annual number of reconstructed spillover events for each year from 2000/2001 to 2017/2018. The year is defined from May to April, rather than January to December. The numbers highlighted in green represent the annual count variations resulting from the shift from the January-December year to the May-April year. 39

2.4 Descriptive statistics of annual spillover events (year from May to April). 39

3.1 Main technical details of the ERA5 reanalysis datasets used in this study, produced by the Copernicus Climate Change Service (C3S) [96]. 42

3.2 Results of the analysis of differences in standardized anomalies between years with high and low spillover events. "Positive" or "Negative" indicates that high spillover years are characterized by statistically significant positive or negative anomalies compared to low spillover years. The range of statistically significant values is shown in square brackets. The first percentage indicates the proportion of districts across Bangladesh reporting statistically significant values. The second percentage, in parentheses, indicates the proportion of districts affected by spillover events over the 18 years analyzed. An hyphen (-) indicates that none or fewer than 19% of districts reported statistically significant values. This 19% threshold was specifically chosen because all other percentages fall below 5%, thereby establishing a clear and significant boundary between relatively rare and more frequent occurrences of statistically significant reports. 52

3.3 The selected indices of extreme events: name, symbol and definition. . . . 54

3.4 Results of the analysis of differences in extreme events indexes (i.e. count day variables) between years with high and low spillover events. "Higher" or "Lower" indicates that high spillover years are characterized by statistically significant higher or lower number of days of the corresponding count variable compared to low spillover years. The range of statistically significant values (ΔN_{days}) is shown in square brackets. The first percentage indicates the proportion of districts across Bangladesh reporting statistically significant values. The second percentage, in parentheses, indicates the proportion of districts affected by spillover events over the 18 years analyzed. An hyphen (-) indicates that none or fewer than 19% of districts reported statistically significant values. This 19% threshold was specifically chosen because all other percentages fall below 6%, thereby establishing a clear and significant boundary between relatively rare and more frequent occurrences of statistically significant reports. 56

5.1 Summary of climate variables and indices analyzed. Variables shown in black and red were statistically significant in the climate analysis (see Chapter 3). Variables in red were ultimately not included in the regression analysis (see considerations in paragraph 5.1.1). Variables in blue were added to replicate the methodology from the reference article [33], allowing for a direct comparison of results. 78

5.2 Summary of univariate GLM analyses. Coefficients highlighted in red indicate p-values greater than 0.05, suggesting non-significance. Asterisks denote significance levels: * for p-values less than 0.05, ** for p-values less than 0.01, and *** for p-values less than 0.001. Coefficients highlighted in blue represent variables excluded due to multicollinearity issues. Periods and variables included in further analysis are shown in black. 86

5.3 Summary of the final full set of models, including those with interaction terms, obtained by testing selected candidate negative binomial GLMs with specific interaction terms. For each model, the table provides: model expressions, $\Delta AICc$ values (in ascending order and relative to the reference model), Pseudo-R^2, the optimal dispersion parameter obtained by minimizing the negative log-likelihood function, and regression coefficients (β). Coefficients in red indicate non-significant results (p-value $>$ 0.05), while black coefficients with asterisks denote significant levels: * (p-value $<$ 0.05), ** (p-value $<$ 0.01), and *** (p-value $<$ 0.001). 88

5.4 Comparison between the best-fitting models derived from the current analysis and the reference article [33]. 89

5.5 Summary of univariate GLM analyses for different clusters over the reference period 2000-2018. Coefficients highlighted in red indicate p-values greater than 0.05, suggesting non-significance. Asterisks denote significance levels: * for p-values less than 0.05, ** for p-values less than 0.01, and *** for p-values less than 0.001. Coefficients highlighted in blue represent variables excluded due to multicollinearity issues. Periods and variables included in further analysis are shown in black. 94

5.6 Summary of the final full set of models, including those with interaction terms, obtained by testing selected candidate negative binomial GLMs with specific interaction terms for both clusters. For each model, the table provides: model expressions, $\Delta AICc$ values (in ascending order and relative to the reference model), Pseudo-R^2, the optimal dispersion parameter obtained by minimizing the negative log-likelihood function, and regression coefficients (β). Coefficients in red indicate non-significant results (p-value $>$ 0.10), while black coefficients with asterisks denote significant levels: * (p-value $<$ 0.05), ** (p-value $<$ 0.01), and *** (p-value $<$ 0.001). 95

5.7 Comparison between the best-fitting models derived for Cluster 1 and Cluster 2. . . . 96

A.1 Annual count of NiV cases and corresponding spillover events from 2001 to 2007. 109

A.2 Nipah virus spillover events in Bangladesh from 2001 to 2011, detailing district, date, and sources. Events classified as 'real' are shown in black, 'reconstructed' based on inferred information in blue, and 'dubious' in red. 115

A.3 Nipah virus spillover events in Bangladesh from 2012 to 2018, detailing district, date, and sources. Events classified as 'real' are shown in black, 'reconstructed' based on inferred information in blue, and 'dubious' in red. 116

A.4 Description of data, data sources and associated details. 126

A.5 Outcome of the exhaustive feature selection process: the top five models with $\Delta AICc < 2$, ranked by ascending AICc values, along with their associated statistics. 142

A.6 Results of model improvement efforts incorporating interaction factors for each candidate model selected based on the exhaustive model selection process, with a selection criterion of $\Delta AICc < 2$. Models identified as improvements over the original baseline model (highlighted in bold) are indicated in blue. 143

A.7 Outcome of the model selection process for Cluster 1 and 2: top models ranked by ascending AICc values, with associated statistics reported. In Cluster 1 only one model presents a $\Delta AICc < 2$ (highlighted in blue). 147

A.8 Results of model improvement efforts for Cluster 1 and 2, incorporating interaction factors for each candidate model selected based on the exhaustive model selection process, with a selection criterion of $\Delta AICc < 2$. Models identified as improvements over the original baseline model (highlighted in bold) are indicated in blue. 148

List of Figures

1.1 A bat from the Family Pteropodidae, Order Chiroptera, and genus Pteropus, commonly known as the flying fox. Source: [38]. 8

1.2 Geographic distribution of Henipavirus outbreaks and fruit bats of the *Pteropodidae* family [34]. Countries at risk (orange) and countries with reported outbreaks (red) are shown [34]. Violet spots mark bat collection sites positive for Henipavirus (HNVs), and blue stars mark HNVs outbreak locations (1997-2008). Green (solid) and violet (dotted) contours indicate the home ranges of *Pteropus* genus and *Pteropodidae* family of fruit bats, respectively. Data source: Global Alert and Response Department, World Health Organization, WHO; map production: Public Health Information and Geographic Information Systems (GIS). Reproduced and adapted according to WHO permission ID: 390902 [34]. 9

1.3 Routes of transmission for Nipah Virus (NiV). This figure has been revised and adapted from [16]. 17

1.4 A bat feeding on date palm sap. The sap from date palm trees can serve as a vehicle for the Nipah virus, facilitating its transmission from bats to humans. Source: [38]. 18

1.5 Temporal overlap of ecological drivers of spillover events in Bangladesh (monthly total across 18 years), highlighting the interplay of bat's nutritional stress (fruit availability [56]), physiological stress (gestation/lactation [48]), thermoregulatory stress (seasonal changes), and the alignment of bat visits with human consumption of raw date palm sap [56]. 27

2.1 (a) Digital Elevation Model (DEM) of Bangladesh elaborated in QGIS software. Data source: SRTM DEM, from NASA web site, 1 arc sec (approximately 30 m) resolution. (b) Red areas represent the population density concentration across Bangladesh in 2001. Source data: WorldPop, 1 km resolution. 31

2.2 (a) Monthly temperature (average, maximum, minimum, and range) and (b) precipitation averaged across Bangladesh during the reference period from 2000 to 2018. 32
2.3 Spatial distribution of district-level seasonal averages for (a) mean temperature and (b) precipitation across Bangladesh during the reference period (2000-2018). 33
2.4 Population trends and growth rate in Bangladesh (2000-2018). Data source: World Bank. 34
2.5 Land cover and land-use in Bangladesh for 2020. Data source: ESA WorldCover. 34
2.6 Comparison of total spillover events by month and by district. 37
2.7 Comparison of total spillover events by month and by division. 38
2.8 Yearly distribution of spillover events (N.tot = 101), with years above the mean (dashed line) marked in red and below the mean in blue. 39
2.9 (a) Histogram of annual (May to April) spillover events with the fitted Negative Binomial distribution (red line). Total spillover counts across affected districts (2001-2018) are compared to the fitted Negative Binomial model. Key statistics, including mean, variance, and dispersion parameter α, are reported. (b) Quantile-Quantile plot (QQ-plot) as a graphical method for determining if the dataset follows a negative binomial probability distribution. 40

3.1 Probability distribution of synthetic differences in standardized anomalies for average winter temperature in Faridpur District. The extreme tails are defined by the percentiles $P_{2.5}, P_5, P_{95}, P_{97.5}$. The observed value, marked with a green dashed line, exceeds the lower $P_{2.5}$ threshold, indicating it falls within the extreme tails of the distribution. This placement confirms the anomaly is statistically significant at the 5% level, demonstrating a significant deviation from the expected climate conditions. 47
3.2 Maps of observed differences in standardised average temperature anomalies between years with high and low spillover events during: (a) winter, and (b) monsoon seasons. **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both. 51
3.3 Maps of observed differences in standardised precipitation anomalies between years with high and low spillover events during: (a) monsoon and (b) post-monsoon seasons. **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both. 53
3.4 Anomalies in average temperatures during the winter and monsoon seasons across various districts of Bangladesh from year 0 to year 17 under analysis. The top panel plots the annual number of spillover events, distinguishing periods before and after the establishment of Nipah virus surveillance. Red dots signify years with high spillover events, while blue dots represent years with low spillover events. The heat maps below show the temperature deviations from the norm for each district during the winter and monsoon seasons, respectively. This visualization aims to explore potential correlations between temperature anomalies and the frequency of spillover events. 60

3.5 July anomalies in average temperatures and precipitation across various districts of Bangladesh from year 0 to year 17 under analysis. The top panel plots the annual number of spillover events, distinguishing periods before and after the establishment of Nipah virus surveillance. Red dots signify years with high spillover events, while blue dots represent years with low spillover events. The heat maps below show the temperature deviations from the norm for each district during the winter and monsoon seasons, respectively. This visualization aims to explore potential correlations between temperature anomalies and the frequency of spillover events. 61
3.6 Maps of observed differences in day counts between years with high and low spillover events for indexes: (a) dry days (b) cold days (c) dry days followed by cold nights). **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both. 62
3.7 Spatial averages of: (a) Monthly totals for all statistically significant count variables, aggregated across districts and years: DD_y, CD_y, DCD_y; (b) Observed monthly differences in the number of dry days, cold days, and dry days followed by cold nights between years characterized by high and low spillover events. 62

4.1 Land Cover and Land Use thematic map of Bangladesh pre-processed in QGIS (Resolution: 10 m). Raster source: ESRI Land Cover - ArcGIS Living Atlas of the World (Sentinel-2 L2A). 65
4.2 Hazard map of flood prone areas categorised by severity (Moderate and Severe) and type (tidal surge, river and flash flooding). 66
4.3 Pre-processed gridded data show cattle density per square kilometer (left). Average cattle density per district, color-coded to indicate different density levels (right). Bold black lines on district boundaries denote areas affected by NiV spillover events, identifying regions within the Nipah belt. 67
4.4 Spatial distribution of fruit crops in Bangladesh, with darker green shades indicating districts with higher percentages of land area dedicated to fruit cultivation relevant for bat diets, showcasing regions where these resources are most abundant. 68
4.5 Spatial distribution of sugar crops in Bangladesh including Date Palms, with darker orange shades indicating districts with higher percentages of land area dedicated to sugar cultivation relevant for bat diets, showcasing regions where these resources are most abundant. 68
4.6 Elbow method's plot for determining the optimal number of clusters (K). The plot shows the sum of squared distances (inertia) against the number of clusters, with the "elbow" indicating the optimal K. 71
4.7 (**Left**) Silhouette plot for $K = 5$ showing the s coefficient values for each cluster, with the average silhouette score $\bar{s}$ indicated by the vertical red dashed line; (**Right**) Visualization of the clustered data after PCA, with each point representing a district colored according to its cluster assignment. 71
4.8 **(Left)** K-means clustering of Bangladesh districts based on land use, land cover and monthly climatic features, showing the optimal configuration with $K = 5$ clusters. Different colors represent the clusters across Bangladesh. Within the Nipah belt, clusters 1 and 2 are specifically highlighted with black and red dots, respectively. **(Right)** Map of Silhouette coefficients illustrating cohesion and separation strength of each district's cluster, represented with a red-to-blue color gradient. The red boundaries indicate the NiV-affected districts. 73

4.9 **Nipah belt as a unique cluster** Comparison of established ecological drivers for NiV spillover events between the Nipah belt (shown in red) and region outside the belt (shown in orange). District-level values and percentages are averaged within the two clusters: the Nipah belt and the remaining districts of Bangladesh. 74
4.10 Spatial distribution of district-level percentages for land devoted to (a) fruit and (c) sugar crops, as well as (b) flood-prone areas. Black-bordered districts indicate those belonging to the Nipah belt, which were affected by spillover events during the 2000-2018 period. 75
4.11 **Two clusters within the Niaph belt** Comparison of established ecological drivers for NiV spillover events between the two clusters within the Nipah belt: cluster 1 (shown in purple) and cluster 2 (shown in green). District-level values and percentages are averaged within the two clusters. 75
4.12 Spatial distribution of district-level (c) cattle density and percentages for land devoted to (a) agricultural crops and (b) human settlements. Black-bordered districts indicate those of Cluster 2, while blue-bordered districts indicate those of Cluster 1. 76
4.13 Contour plot of average annual temperature and precipitation. Red-bordered districts represent those in Cluster 2, while blue-bordered districts indicate those in Cluster 1. 76

5.1 Variation in the yearly number of Nipah spillover events explained by: (a) mean winter temperatures, and (b) mean July temperature. Observed values are represented as dots, predicted values by the negative binomial regression model are shown as dashed lines, and the 95% confidence intervals of the predicted values are indicated by the shaded blue areas. 87
5.2 Top GLM without interaction terms and temporal trends: (a) The model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo-*R^2 and coefficient estimates along with their statistical significance. Each data point corresponds to a specific year, labeled from 0 to 17, corresponding to the annual periods from 2000/2001 to 2017/2018 (where the year spans from May to April). The model's predicted values are depicted by the smooth surface. (b) Annual time series of the considered explanatory variables (i.e. mean July and winter temperatures), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh. 90
5.3 Diagnostic analysis of the improved M_1^* negative binomial GLM: (Left) enhanced influence plot with scatter point sizes proportional to Cook's distance, and (Right) index plot of Cook's distance, both with indicated thresholds for identifying outliers and influential points, respectively. 91
5.4 Top GLM with interaction terms and temporal trends: (a) The model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo-*R^2 and coefficient estimates along with their statistical significance. Each data point corresponds to a specific year, labeled from 0 to 17, corresponding to the annual periods from 2000/2001 to 2017/2018 (where the year spans from May to April). The model's predicted values are depicted by the smooth surface. (b) Annual time series of the interaction term (i.e. $T_{avg,july} \cdot T_{avg,winter}$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh. 92
5.5 Comparison between observed annual spillover counts (green dots) and predictions from the top three models. 92
5.6 Histogram of annual (May to April) spillover events with fitted Negative Binomial distribution (red line). Key statistics, including mean, variance, and dispersion parameter α are reported for: (a) Cluster 1, and (b) Cluster 2. 93

5.7 Best negative binomial GLM and temporal trends for Cluster 1: (a) The best model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo*-R^2 and coefficient estimates along with their statistical significance. (b) Annual time series of the explanatory variable (i.e. $T_{avg,winter}$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh. 97

5.8 **Cluster 1:** Diagnostic analysis of the best-fit negative binomial GLM: enhanced influence plot (left) with scatter point sizes proportional to Cook's distance, and index plot of Cook's distance (right), both with indicated thresholds for identifying outliers and influential points, respectively. 97

5.9 Best reference negative binomial GLM and temporal time series for Cluster 2: (a) The best model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo*-R^2 and coefficient estimates along with their statistical significance. (b) Annual time series of the explanatory variables included in the $M^{(0)}_{ref,C2}$ model (i.e. $T_{avg,july} \cdot DD_y$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh. 98

5.10 **Cluster 2:** Diagnostic analysis of the best-fit negative binomial GLM: enhanced influence plot (left) with scatter point sizes proportional to Cook's distance, and index plot of Cook's distance (right), both with indicated thresholds for identifying outliers and influential points, respectively. 99

5.11 **Cluster 2:** Comparison between observed annual spillover counts (green dots) and predictions from the top six models. 99

A.1 Global impact of El Niño: temperature and precipitation patterns that are typical of April to September (left) and October to March (right)of El Niño conditions. Map by NOAA Climate.gov and FAO. 106

A.2 (a) Figure 1 taken from [93] reporting the location of spillover events in years 2001-2007; (b) Figure 1 taken from [93] imported, georefrenced and re-projected in QGIS; (c) Resulting information. 110

A.3 *Spaghetti plot of (a) monthly mean areal temperature [°C/day] and (b) precipitation across 2000-2018 period and districts with seasonal highlights. Each colored line represents a district, with the black dashed line indicating the ensemble mean across all districts.* 119

A.4 Spatial distribution of district-level densities of Livestock, specifically: (a) Goat, (b) sheep and (c) pigs. Black-bordered districts indicate those belonging to the Nipah belt, which were affected by spillover events during the 2000-2018 period. 127

A.5 *Case A - Unique cluster:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variables (i.e. mean temperature, minimum temperature, percentage of cold days, percentage of dry days followed by cold nights), annual indexes (i.e. annual dry days, annual cold days, annual dry days followed by cold nights) and July variables (i.e. mean temperature, minimum temperature and precipitation). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***). 141

A.6 Diagnostic plots assessing the validity and verifying the assumptions of the top GLM model without interaction terms. 144

A.7 Diagnostic analysis of the best-fit negative binomial GLM: (Left) enhanced influence plot with scatter point sizes proportional to Cook's distance, and (Right) index plot of Cook's distance, both with indicated thresholds for identifying outliers and influential points, respectively. 144

A.8 *Case B - Cluster 1:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variables (i.e. mean temperature, minimum temperature, percentage of cold days, percentage of dry days followed by cold nights), annual index (i.e. annual dry days followed by cold nights) and July variable (i.e. mean temperature). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***). 145
A.9 *Case B - Cluster 2:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variable (i.e. mean temperature), annual indexes (i.e. annual dry days, annual cold days, annual dry days followed by cold nights) and July variables (i.e. mean temperature, minimum temperature and precipitation). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***). 146
A.10 **Cluster 1:** Diagnostic plots to evaluate the best reference model's validity and check GLM assumptions. 147
A.11 **Cluster 2:** Diagnostic plots to evaluate the best reference model's validity and check GLM assumptions. 149

Acronyms

AIC Akaike's Information Criterion. 82

AICc Akaike's Information Criterion Corrected for small sample sizes. 77, 82, 87, 91

AR6 Sixth IPCC Assessment Report. 19, 20, 22

CMIP6 Coupled Model Intercomparison Project phase 6. 100, 101

DJF December, January and February, months of the Winter season. 78

DPS Date Palm Sap. 16, 18, 19, 26, 27, 35, 49, 50, 58, 63, 94, 100, 103

EBOV Ebola virus. 3

ECMWF European Centre for Medium-Range Weather Forecasts. 42

EID Emerging Infectious Disease. 1, 2, 23, 48

ENSO El Niño Southern Oscillation. 23

GCM Global Climate Model. 100

GHG Greenhouse gases. 19

GIS Geographic Information System. viii, 9, 43

GLM Generalised Linear Model. vii, xi–xiii, 51, 77, 80–86, 88, 91–95, 97–99, 129, 131, 133, 135, 138, 139, 144, 147, 149

HeV Hendra virus. 6, 23

HNV Henipavirus. viii, 3, 9, 12–16

HPS Hantavirus Pulmonary Syndrome. 23

icddr,b International Centre for Diarrheal Disease Research, Bangladesh. 36

IECDR Institute of Epidemiology, Disease Control and Research. vi, 36

IPCC Intergovernmental Panel on Climate Change. 19, 20, 22

IUCN International Union for Conservation of Nature. 11

JJAS June, July, August and September, months of the Monsoon period. 78

L-BFGS-B Limited-memory Broyden–Fletcher–Goldfarb–Shanno Bound algorithm. 81, 82

MAM March, April and May, months of the Pre-monsoon period. 78

MLE Maximum Likelihood Estimation. 80, 81, 85, 93, 130–132

NiV Nipah virus. vi, viii, x–xii, 1, 2, 6–8, 10–12, 14–19, 23, 24, 28, 29, 36, 37, 41, 49, 52, 53, 63–65, 67, 69, 70, 72–76, 89, 90, 92, 96–98, 101–105

NWP Numerical Weather Prediction. 43

ON October and November, months of the Post-monsoon period. 78

PCA Principal Component Analysis. x, 71, 72

SSP Shared Socioeconomic Pathways. 100

WHO World Health Organization. vi, viii, 9, 36

WMO World Meteorological Organization. 45

Glossary

Death burden is the total number of human deaths [1].. 24

Emerging infectious diseases are those that (1) have not occurred in humans before; (2) have occurred previously but affected only small numbers of people in isolated places, such as AIDS and Ebola hemorrhagic fever; or (3) have been present throughout human history but have only recently been recognized as distinct diseases due to a infectious agent, as in the case of Lyme disease and gastric ulcers [2].. 22

Fomite is any inanimate object that, when contaminated with or exposed to infectious agents, can transfer disease to a new host.. 23

Prevalence is the proportion of infected and infectious individuals in a population, reflecting the current infection history [3].. 12

Re-emerging infectious diseases are diseases that once were major health problems globally or in a particular country, and then declined dramatically, but are again becoming health problems for a significant proportion of the population (i.e. malaria and tuberculosis) [2].. 22

Seroprevalence measures the proportion of animals that have been exposed to the virus and seroconverted (produced antibodies to the pathogen), thus reflecting the cumulative infection history [3].. 9

Transmissibility is the diseases capacity to spread within human populations [1].. 24

Viral shedding is the expulsion and release (shed) of infectious virions (i.e., infectious virus particles) from the host (body) into the environment, where the virus may infect another recipient host [4], [5], [6]. Bats can shed the virus they carry trough body secretions (saliva, blood) and excretions (urine and feces/guano) and can contaminate foods and surfaces with subsequent spread to another host [4], [5], [6].. 14

Virulence is defined as the degree to which a pathogenic organism can cause disease [1].. 24

Zoonosis a disease that is caused by an infectious pathogen or parasite that originates in (or is maintained in the wild by) one or more non-human hosts but can be transmitted to and cause disease in humans [7].. 3

Introduction

Climate and anthropogenic land-use change are the primary drivers of the emergence and re-emergence of infectious diseases (EID) [8], 60.3% of which are zoonosis, with 71.8% originating from wildlife, posing an escalating threat to global health security [9]. Zoonosis, transmitted to humans from an infected vertebrate host [10], have the potential to spread rapidly across borders, overwhelm unprepared healthcare systems, and cause severe economic losses—as dramatically demonstrated by the COVID-19 pandemic.

Climate change, coupled with human-induced ecosystem changes (i.e. deforestation, biodiversity loss, urbanization, road construction, agricultural expansion), heightens disease transmission risks by directly and indirectly shaping the dynamics and interactions between hosts, vectors and pathogens [10], [8]. This is driving the fast expansion of geographic areas flagged as "*jump-zones*", regions where EIDs are increasingly likely to spill over into human populations [11].

Bats, which harbor some of the most virulent zoonotic pathogens capable of infecting humans [1], are particularly sensitive to climate fluctuations and serves as key indicators of ecosystem health, being their biology, physiology and behaviour closely tied to external environmental conditions [12]. These connections, combined with their unique dispersal ability, evolutionary traits and opportunistic behaviour, makes bats a unique reservoir host among wild mammals, likely to drive cross-species viral transmission events in the near future [13]. Indeed, climate change is expected to shift the habitat ranges of wildlife hosts and led to novel aggregation in areas of high population density and biodiversity, positioning South and Southeast Asia as bat-borne spillover hotspots [13]. However, despite their significance, bat responses to climatic factors remain largely underexplored [14], underscoring the urgent need for focused research—one of the key motivations behind this thesis.

The **Nipah virus** (NiV) is recognized by the World Health Organization as one of the top 10 priority diseases with the potential to cause a **pandemic** [15], [16]. Unlike other bat-borne pathogens, which tend to be sporadic and localized, NiV has emerged in multiple, distant locations over a 26-year period (1998-2024), leading to significant loss of life[17]. With a mortality rate of 40-70% and no available cure or vaccine [18], understanding the precise drivers of NiV spillover events is crucial for developing targeted mitigation strategies.

An in-depth literature review has revealed gaps and opportunities for improvement in understanding the relationship between climate and NiV spillover events. These findings have guided the methodological choices in this thesis, influencing the selection of climate data source (specifically, ERA5 reanalysis data for its higher spatial resolution and reliability in tropical regions), the exploration of relevant climate variables, and the extension of analysis periods (i.e., beyond the typical winter months). These decisions have contributed meaningfully to the progress made in this study.

This thesis investigates the relationship between Nipah virus spillover events and climatic conditions in Bangladesh from 2000 to 2018, with the aim to identifying climatic variables and weather extremes that may have influenced the frequency and spatial distribution of these events. The structure of the thesis is as follows: after an overview of the state of the art in *Chapter 1*, which explores the role of bats as reservoirs of zoonotic viruses and the impact of climate and anthropogenic changes on the emergence and spread of EID, *Chapter 2* provides a spatio-temporal reconstruction and analysis of the spillover events occurred in Bangladesh in the reference period. In *Chapter 3* a detailed temporal analysis, exploring anomalies, trends, and extremes in temperature and precipitation, was conducted to identify distinctive and statistically significant climate patterns associated with years of high versus low spillover occurrence. In *Chapter 4*, these climatic insights, along with other environmental and anthropogenic stressors, are analyzed for their spatial distribution across Bangladesh. A clustering procedure groups districts by common stressors to explore regional variations in transmission drivers. Finally, *Chapter 5* presents a regression analysis using Generalized Linear Models (GLMs) to test the hypothesized correlations found throughout the study. Notable findings include climate patterns previously observed only in relation to the Hendra virus in Australia, providing new insights into how climatic conditions may drive NiV spillovers in Bangladesh.

Chapter 1

Anthropogenic changes drive the emergence and spread of zoonosis: Nipah virus

1.1 Wild mammals and zoonotic infections

Most of the infectious diseases of humans have originated in wild and domestic animals [19]. Among the 10,000 species of viruses capable of infecting humans, the vast majority circulate silently, or cause subclinical illness, in wild mammals [13]. Therefore, compared to any other class of vertebrates, mammals are responsible, directly or indirectly, for the majority of the emerging infectious diseases that affect humans, making this group and their pathogens highly relevant to public health [13].

The rapid changes in climate and land use have impacted the abundance and distribution of many mammal species and, with this, their ability to shift to different environmental settings and adapt to novel habitat conditions. During these shifts, animals can move long distances and establish in new areas, often bringing their parasites and pathogens into these new environments [13]. Species interactions and sharing of resources, including the accidental exposure of human individuals to wildlife, can potentially facilitate the circulation of novel pathogens into new areas and, with this, the risk of zoonotic spillovers (*see* **Zoonosis**), where pathogens are transmitted from wildlife to humans. Predicting the conditions that contribute to spillover events and potential hotspots of zoonotic disease emergence requires the understanding of the environmental circumstances that facilitate the interactions between wildlife and humans and the drivers that can lead to a spillover event [4].

Given their unique dispersal abilities, unconstrained within continents, and distinctive immune system that protects from disease, bats have been suggested to play an important role in the cross-species transmission of many viral infections, including Ebola (EBOV), SARS-2-COVID-19 and Henipaviruses (HNV) [20], and they are likely to share viruses along evolutionary pathways that will facilitate future emergence in humans [13]. Hotspots of novel viral emergence are projected to coincide with areas rich in biodiversity and with high human population density, where the potential risk of zoonotic infections to spillover into humans is expected to be high [13]. Recent studies suggest that these areas are disproportionately likely to occur in tropical Asia such as rural settings with cropland

and human settlements aggregate around biodiversity hotspot [13].

1.2 Bats: Unique mammalian virus host

Bats belong to *Chiroptera*, the second most species-rich order of mammals after rodents, comprising approximately 22% of all named mammal species [21]. Bats can be infected with a high diversity of virus species, relative to other mammalian orders, and they are the natural reservoir host for several emerging viruses of major health concern to humans, including RNA viruses such as Marburg virus, Hendra virus, and Nipah virus [21]. The high richness and zoonotic nature of viral infections they carry [21], combined with the strong adaptation of their immune system to these infections, make bats unique when compared with other mammals [22], [21]. Some of this uniqueness is associated with the biological and behavioural features of the group [23]. For example, the size of colonies can reach the order of millions, individuals have a gregarious social behaviour and habits of mutual grooming; all together, these characteristics can likely facilitate the rapid spread of infectious agents throughout the colony [23]. Moreover, the long individual lifespan, which can facilitate viral persistence, and a history of long co-evolution with many viruses, including their ability to fly over large home ranges and spread or contract viral infections over vast areas, make bats the ideal reservoir host for viral transmission and persistence in the environment [24].

Among the various features listed above, it is particularly important to emphasize the bats' ability for powered flight, which allows them a more rapid and unconstrained range shifts than any other mammals [13]. Even non-migratory bats can regularly travel hundreds of kilometers within a lifetime, moving between roosting sites [13]. Multiple studies have also reported ongoing rapid range expansions in bat species across the world, making these movements extremely relevant for the epidemiology of viral infection under climate change [13]. These movements are frequently associated with changes in resource availability like food constraints, as well as changes in foraging frequency and duration and patterns of social aggregation and colony composition [4]. By understanding the drivers and patterns of these movements it is possible to identify processes of viral transmission and potential viral hotspots of human infection.

Unique immune system The bats' remarkable ability to transmit viruses is closely tied to their unique immune system, which enables them to carry a wide variety of pathogens while remaining asymptomatic [21], [22]. This phenomenon is thought to be linked to their distinct innate and adaptive immune responses to viral infection [21], [22].

A distinctive aspect of the bat's innate immune system is its constant production of interferon-alpha (IFNα) and subsequent induction of IFN-stimulated genes (ISGs), in contrast to other mammals, which produce this protein only in response to infection [22]. This continuous immune activation helps control viral replication and prevent disease, but it may impose high metabolic costs, which are managed through regulatory mechanisms like anti-inflammatory pathways [22].

In addition to innate immunity, bats also possess a distinct adaptive immune system [22], [21]. Although they produce antibodies in response to viral infections, these responses tend to be weaker and shorter-lived compared to those in other mammals [21]. For example, studies have shown that bats generate antibodies to viruses such as Marburg and Nipah, but these antibodies often wane quickly and are insufficient to fully neutralize the virus [21]. This could indicate that bats rely more heavily on antibody-dependent cell-mediated immune responses, which may limit viral replication without fully clearing the infection [21]. The ability to tolerate low levels of viral presence without suffering from disease may play a role in the maintenance and transmission of viruses within bat populations [21]. Furthermore, nutrition appears to have a significant impact on bat immune function [22]. The resource-intensive nature of immune responses, particularly those involving interferon production and antibody generation, means that dietary limitations can influence how effectively bats control infections [22]. For example, vitamin deficiencies and periods of starvation can suppress both innate and adaptive immune responses, potentially increasing susceptibility to infection and the likelihood of viral shedding [22]. This relationship between nutrition and immune function is particularly relevant during periods of physiological stress, such as pregnancy and lactation, when bats are more prone to immune compromise [22].

1.3 South and Southeast Asia hot-spots of infections

Several factors make South and Southeast Asia particularly critical areas for the emergence and re-emergence of zoonoses. Firstly, these areas are species-rich ecosystems, hosting approximately 25% of the world's bat diversity [25]. By serving as a significant reservoir host of pathogens, bat populations from these areas increase the likelihood of novel bat-borne viral zoonosis emergence. Second, the rapid degradation and extensive conversion of bat habitat into anthropogenic usage, cause severe biodiversity and habitat loss while increasing the risk of bat-human interactions and viral transmission. Southeast Asia, homes to nearly 15% of the world's tropical forests, is also among the world major deforestation hotspots, with rates comparable only to that of Latin America [26]. Third, much of this deforestation is converted into extensive monocultures, along with the establishment of human settlements and grazing land for livestock. Fourth, the destruction of forests and resources (roosting and food) for bats has forced animals to move in close proximity to humans and increase the likelihood of closer and more frequent interactions.

The encroachment of humans on bat habitat, combined with the dramatic growth of the human population and often inadequate public health systems, create opportunities for zoonotic infections to spill over from bats into humans while increasing the risk of disease spread in South and Southeast Asia.

Reuters, a global news agency, identified high-risk regions where conditions are ideal for bat-borne virus spillover, potentially triggering future pandemics [11]. These areas, covering 6% of the Earth's landmass, are primarily tropical regions undergoing rapid urbanization and were home to nearly 1.8 billion people in 2020 [11]. India, Indonesia and Laos are flagged as developing countries where the riskiest-areas are experiencing the fastest expansion, while Bangladesh as one witnessing a rapid growth in spillover risk [27]. Notably, pandemic risk is higher in countries with fewer resources to manage it [27].

1.4 Nipah Virus

Among the emerging infectious zoonoses carried by bats endemic to South and Southeast Asia, Nipah virus represents a major concern for public health, and has been listed by the World Health Organization as one of the top ten priority diseases that necessitate urgent investigation and control [28].

Until now, Nipah outbreaks have been confined to these regions, however, its natural reservoir host (i.e., fruit bats) occupies a vast geographic range across the world, encompassing regions home to over 2 billion people [29]. As human interactions with these bats become more frequent, driven by the accelerating pace of anthropogenic ecosystem alteration—including land-use changes and the effects of climate change—the risk of Nipah virus transmission intensifies, significantly increasing the potential for future outbreaks [29], [12].

1.4.1 Etiology

Nipah virus (NiV) is a deadly zoonotic *paramyxovirus* that belongs to the genus *Henipavirus* in the family *Paramyxoviridae*, a group of negative-sense single-stranded RNA viruses [15]. This genus also includes Hendra virus (HeV) and the recently discovered Cedar (CedV), Kumasi (KV) and Mojiang (MojV) viruses [15]. To date, two major strains of NiV have been identified:

- NiV_M strain, isolated from pigs in Malaysia, which shows no evidence of human-to-human transmission [15][30];
- NiV_B strain, more infectious and responsible for human cases in Bangladesh, making it the most severe and deadly form of the infection [15], [30].

1.4.2 Uniqueness of Nipah virus infectious events

The natural reservoir hosts of NiV are bats of the *Pteropodidae* family (*Pteropus* genus), which carry, beside *Paramyxoviridae*, viruses of the families of *Reoviridae* and *Rhabdoviridae*. Unlike most of these bat-borne pathogens that emerged in a single geographic area, NiV uniquely appeared in multiple distant locations over a 26-year period (1998-2024) [17]. Moreover, although zoonotic disease outbreaks are often unpredictable and rare, NiV has repeatedly spilled over from bats to humans and livestock, causing outbreaks with high fatality rates across a wide geographic range [31]. The virus first emerged in Malaysia in 1998, followed by outbreaks in Singapore in 1999, Bangladesh and India in 2001, and the Philippines in 2014 [17]. Among these countries, only Bangladesh experiences regular outbreaks, almost every winter (typically between November and April). The corresponding near-annual spillover events are closely linked to the consumption of raw date-palm sap and exhibit significant variability in locations and frequency (i.e. numbers fluctuate annually) within the central and northwestern regions of Bangladesh, commonly referred to as the "Nipah belt" [31]. Spillover has also been documented outside the typical winter season and region associated with date-palm sap consumption [31]. Interestingly, while no human outbreaks have been reported in eastern Bangladesh, despite similar consumption habits, outbreaks have occurred in Kerala (India) where date-palm sap is not harvested [31]. These observations suggest alternative spillover routes and highlight the importance of understanding the factors driving viral dynamics within bat populations [31].

1.4.3 Why a major concern

Nipah virus is considered to have the potential to cause one of the next pandemics [16], [15]. Many NiV strains have characteristics that make them a significant threat to both human and animal heath:

- They cause an **high human fatality rate**, ranging from 40-75% in various outbreaks (82% in case of acute encephalitis [15]), with a 32% of survival patients reporting residual neurological deficits [30].

- **No vaccines** or specific antivirals are available for either people or animals [30]. The primary treatment for humans is supportive care [30].

- NiV has a **long incubation period** before symptom onset in human, from 4 up to 45 days [18], thus unaware infected hosts have ample opportunities to spread the virus [32]. NiV infection can manifest in a wide array of clinical presentations from asymptomatic to acute respiratory disease, including fatal encephalitis [16]. Cases with mild or sub-clinical infections are likely to go undetected. For this reason, it is believed that at least half of all Nipah virus outbreaks in Bangladesh during the period 2007-2014 were likely gone missed, as encephalitis patients often seek treatment locally, especially if their symptoms are less severe and the surveillance hospital is distant from their home [33].

- **Many domesticated animals**, including pigs, goats, dogs, cats, horses, sheep, and cattle are susceptible to NiV [30], and can serve as intermediate recipient hosts, transmitting the virus to humans [3]. Notably, both *intra-species* (intermediate host-to-intermediate host, human-to-human) and *inter-species* (bat-to-human, intermediate host-to-human) transmission of NiV is possible [34]. Indeed, humans can become infected either directly from bats—through bat consumption or contact with bat secretions or excreta (refer to Section 1.5.3 for further details) [30]-or indirectly through unprotected contact with an infected intermediate host (e.g., via skin abrasions) or by consuming under-cooked meat from a sick animal [30]. Additionally, human-to-human transmission is also possible via respiratory secretions (refer to Section 1.5.3 for further details) [30]. These multiple transmission routes significantly increase the risk of human contagion.

- **NiV may persist in the environment**, outside its reservoir host, for several days, influencing the likelihood of spillover [3]. NiV survival is sensitive to increasing temperature, changes in pH, ultraviolet light and desiccation, thereby surviving longer in winter months rather than in summer [35], [33], [30], [34], [3].

- **Numerous spillovers and outbreaks have frequently occurred** in densely populated and economically disadvantaged regions of Asia, exacerbating the public health impact.

- Fruit bats (*Pteropus genus*), the **natural reservoirs of NiV, are widely distributed throughout Asia**. This extensive distribution increases the so-called “viral chatter,” where repeated human exposures to the virus may lead to transformations that enhance its ability to spread among humans [36].

1.4.4 Natural reservoir of NiV

Bats of the genus *Pteropus* (*Chiroptera: Pteropodidae*), commonly known as flying foxes or fruit bats, are the natural reservoirs for Nipah virus. Flying foxes (*Pteropus medius*, formerly *P. giganteus*), named for their physical similarity to a fox, are the largest bats in the world, having a wingspan of 1.5 metres with a head and body length of about 40 cm [37].

Key stressors affecting bat immune system function *Pteropus* exhibit enhanced innate and adaptive immune tolerance to NiV infections and healthy animals carry the virus without showing clinical disease symptoms, making them effective asymptomatic carriers (refer to Section 1.2 for further details) [21], [22]. However, bats are not immune to the effects of stress, particularly during periods of nutritional deficiency and reproductive phases such as pregnancy and lactation [22]. Immune responses are highly resource-intensive to develop and maintain, relying heavily on adequate nutrition for proper functioning [22]. When dietary shifts limit the intake of essential nutrients and energy, the immune system's ability to control viral replication weakens, leading to higher infection intensity (i.e. viral load) [22]. During times of nutritional stress, bats are more likely to rely on less costly immune defenses, such as their innate immune system, which may be less effective at controlling viral replication [22]. Pregnancy and lactation also impose significant resource demands on bats, further compromising their immune defenses [22]. Research indicates that pregnant bats experience reduced immune function, as their bodies prioritize resource allocation for reproduction over immune defense [22]. This compromised immune state not only increases their susceptibility to infections but may also reactivate latent viruses [22]. Consequently, pregnant and nutritionally stressed bats are more likely to have elevated viral loads and exhibit increased shedding (i.e., excretion of virus), making them more effective transmitters of viruses like NiV [22]. This is particularly concerning because viral load plays a crucial role in spillover, likely correlating with the quantity of pathogen excreted and its potential to infect susceptible hosts [22]. Further details on the conditions that enable the virus to spill over from bats to recipient hosts will be discussed in Section 1.5.3.

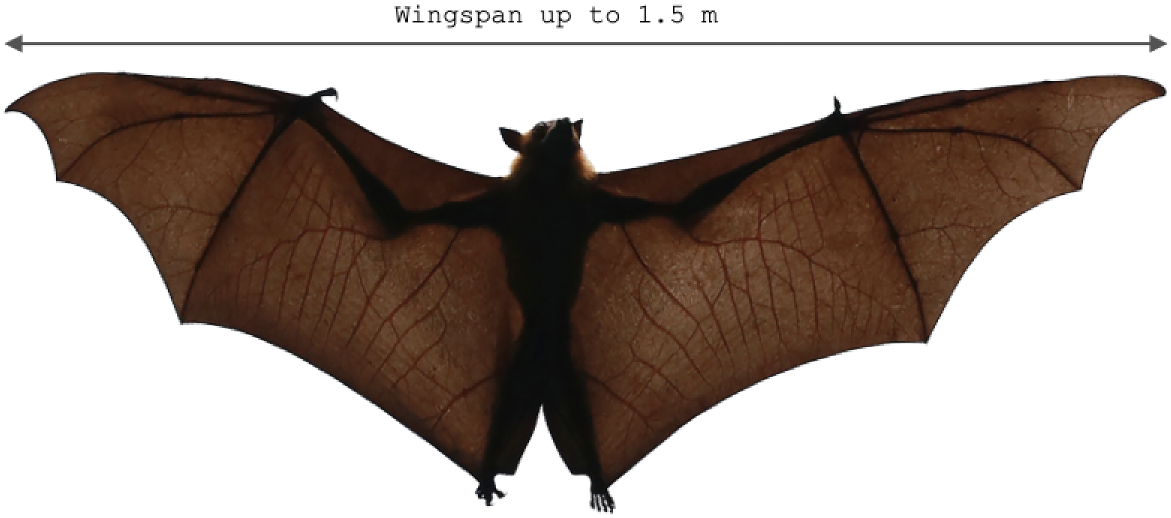
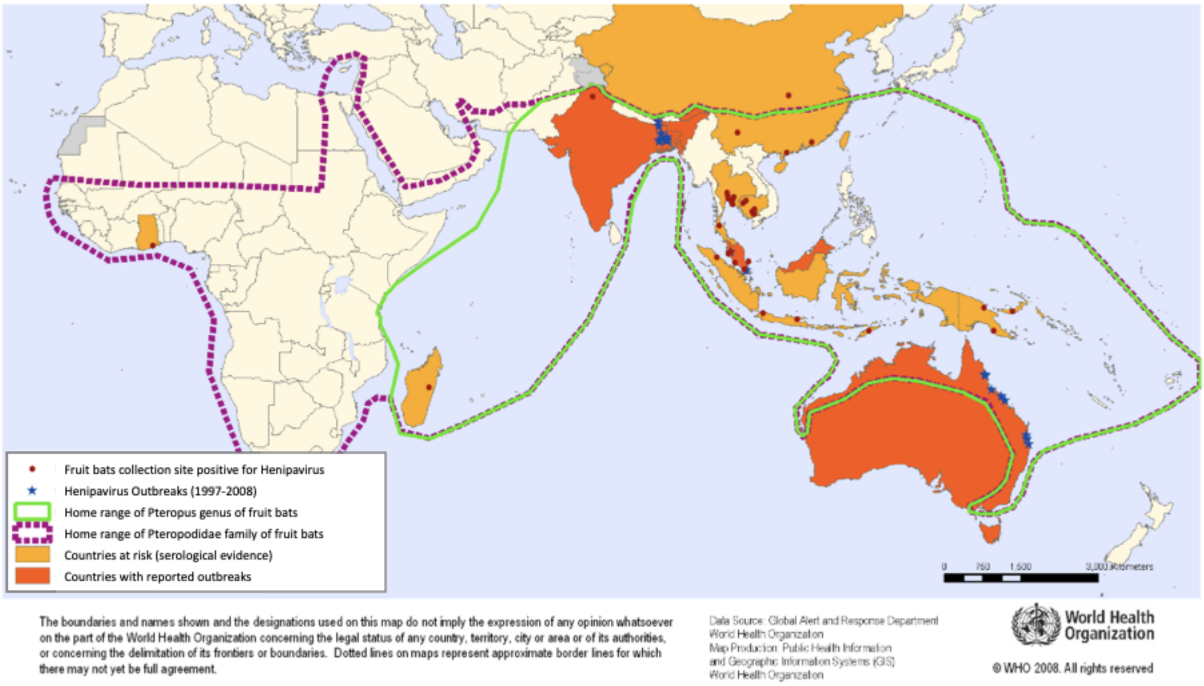


Figure 1.1: A bat from the Family Pteropodidae, Order Chiroptera, and genus Pteropus, commonly known as the flying fox. Source: [38].

Geographic distribution

The geographic distribution of bat populations reservoir hosts of Henipavirus, ranges from continental and insular Australia and Asia to Africa and is assumed to be a proxy of the distribution of this family of viruses [18]. The geographic distribution of Henipavirus bat reservoir hosts, which extends from continental and insular Australia and Asia to

Africa, is assumed to be a proxy of the viral distribution. This assumption is supported by several factors, including serological studies that have demonstrated Henipavirus infection in all tested (i.e. nine) *Pteropus* species across eleven countries within the *Pteropodid* bat distribution [39]. Additionally, Henipavirus **seroprevalence** is high, and these bat species often share overlapping ranges, roost together, and exhibit high mobility [39]. These behaviors promote both intra- and inter-species viral mixing, further supporting the link between the distribution of these bats and the virus [39]. While this assumption offers valuable insight, it provides only a broad indication of areas at risk of spillover. For the virus to spill over from the reservoir bats to a secondary or accidental reservoir species, specific biotic and abiotic conditions must also be met (refer to Section 1.5.3 for further details regarding the enabling condition for spillover events).



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

Data Source: Global Alert and Response Department
World Health Organization
Map Production: Public Health Information
and Geographic Information Systems (GIS)
World Health Organization

© WHO 2008. All rights reserved

Figure 1.2: Geographic distribution of Henipavirus outbreaks and fruit bats of the *Pteropodidae* family [34]. Countries at risk (orange) and countries with reported outbreaks (red) are shown [34]. Violet spots mark bat collection sites positive for Henipavirus (HNVs), and blue stars mark HNVs outbreak locations (1997-2008). Green (solid) and violet (dotted) contours indicate the home ranges of *Pteropus* genus and *Pteropodidae* family of fruit bats, respectively. Data source: Global Alert and Response Department, World Health Organization, WHO; map production: Public Health Information and Geographic Information Systems (GIS). Reproduced and adapted according to WHO permission ID: 390902 [34].

Roosting behaviour

Studies on *P. medius* showed that animals forage at night and rests during the day in tree roosts that can host colonies of many thousands of individuals. Bats tend to select exposed tree branches, in emergent trees that rise above the forest canopy. While this facilitates fly landing and takeoff, given their large size and wingspan, it also makes them particularly susceptible to the effects of extreme temperatures [40].

Roost sites selection and spillover implication Additionally, it is important to note that *Pteropus* roost selection is strongly influenced by food availability and proximity [41]. A significant factor behind *Pteropus giganteus* preferring to roost in forests near densely populated human areas [41] is the combined impact of habitat loss, fragmentation, depletion of native food sources, and the provision of alternative food resources through agriculture. The widespread land-use change, fuelled by urbanization and agricultural expansion, have pushed these bats closer to human settlements, where they can take advantage of abundant food in species-rich urban environments [41], [33]. Growing evidence suggests that these human-induced environmental changes have significantly altered *P.giganteus* roosting behaviour, as well as movements and foraging patterns, with the species now primarily feeding in residential backyards and orchards, rather than in plantations that have replaced their natural habitats [42], [33]. This shift is likely driven by the greater variety and year-round availability of fruit in urban areas, supported by irrigation and diverse planting [43]. Moreover, urban environments provide warmer micro-climates, which are more suitable for these bats, along with potential refuges from predators and post-cyclone effects [44]. These factors collectively demonstrate the remarkable behavioral flexibility of *Pteropus*, allowing them to adapt and thrive in such altered habitats. However, this closer proximity to humans increases the frequency of contact, creating more opportunities for zoonotic diseases to spillover [42], [33]. Further discussion on this topic will be provided in Section 1.5.

Foraging behaviour

Despite their size, these nocturnal mammals are nectivorous and frugivorous. Their diet includes pollen, flowers and leaves, which meet their protein requirements, while nectar, along with the sugar in fruits, are essential sources of energy to power their flight [45]. Bats are generalists with opportunistic foraging behavior, a characteristic that has facilitated the use of more urban settings and the exploitation of alternative food sources within anthropogenic landscapes.

Research into the feeding habits of fruit bats confirms their role as messy eaters. During feeding, fruit bats ingest the juice of fruits and expel the fibrous pulp, while also urinating and defecating [11], [38], [46]. Indeed, their rapid digestive system results in droppings within half an hour of feeding [47]. Foraging areas often bear distinctive marks on the landscape, such as scattered foliage and partially consumed fruits [3].

Messy eaters and food-borne spillover The messy eating habits of fruit bats contribute to nutrient cycling and seed dispersal in tropical ecosystems, playing a crucial role in plant regeneration dynamics and supporting biodiversity [48], [22].
However, these habits also pose health risks, particularly when droppings and partially eaten fruit, contaminated with saliva, carry lethal viruses such as Nipah. Yet, the facilitation of food-born spillover from bats requires a series of interconnected conditions, including the exposure of a secondary host to the contaminated food. This scenario occurred during the first Nipah outbreak in Malaysia in late 1997/early 1998, where mango trees were planted near pig farms to boost agricultural output [36]. The initial transmission of NiV from bats to pigs likely occurred through contamination of pig swill by bat excretions [49]. This was driven by the migration of forest-dwelling fruit bats to cultivated orchards near pig farms, a consequence of the widespread fruiting failure of forest trees caused by the El Niño-induced drought and anthropogenic fires in Indonesia

in 1997-1998 [50].
The decision to boost mango and swine production, together with the establishment of monoculture plantations, in close proximity to pig farms and slaughterhouses promoted spillover [51]. Likewise, the high density of animals on these farms amplified animal-to-animal transmission of the virus. To contain the infection, a massive culling campaign of over a million pigs was necessary, resulting in a devastating economic impact [36], [30]. The emergence of NiV is a clear example of how the loss of natural habitat, together with agricultural encroachment and the increase of livestock abundance, could amplify the likelihood of human-animal interactions and the potential for zoonotic spillover events [36].

Ecosystem services

Flying foxes are highly mobile species, regularly traveling long distances between roosts and foraging sites, exhibiting both frequent and seasonal movements [52]. This behavior provides key ecosystem services such as pollination and long-distance seed dispersal of native and agricultural plants, including economically valuable crops like durian fruit across the paleo- and subtropical regions of Asia [22]. Additionally, they play a key role in forest regeneration by maintaining genetic diversity in fragmented forest habitats [22].

Conservation status

Roosting and foraging behaviour of plant-visiting bats has been severely altered by human pressure on ecosystems, a pattern that is increasing globally [22]. According to IUCN Red List of Threatened species, 60% of *Old World fruit* bat species now occupy disturbed and altered resource landscape, such as gardens, urban areas, plantations [22]. A recent estimate suggests that 75% of species of *Pteropodidae* have declining populations due to these anthropogenic disturbances [22]. The alteration of flying fox habitat is causing changes in their ecology, which in turns is expected to reshape infection and transmission dynamics at both the roost and meta-population level [53].

1.5 Anthropogenic pressure changes bat ecology and increases spillover risk

Deforestation, agricultural intensification, urban development and climate change are removing or irreversibly altering critical wildlife habitat [22]. The loss of native forest poses a significant threat to Flying foxes by destroying both their food resources and roosting habitat [48]. Given the rapid pace of habitat alteration due to human activities, it is crucial to understand how land-use changes impact wildlife species and, in turn, how these changes influence human health [54]. Tropical deforestation, primarily driven by population growth, escalating global food demand, and government policies focused on rural development and agricultural expansion, has led to the extensive conversion of bat habitats into agricultural, pastoral, and urbanized areas [12]. This has resulted in fragmented landscapes characterized by small (often < 50 ha), isolated, and irregularly shaped forest patches [12]. Such habitat degradation, loss, and fragmentation, combined with alternative resource provision, have led to two primary consequences: the redistribution of food resources, and the disruption of bat metapopulation connectivity at the landscape level. Both of them have drastically altered bat roosting, foraging, migratory behaviors and aggregation pattern, with significant implications for zoonotic spillover events.

1.5.1 Redistribution of resources in space and time

Ecosystem alterations have resulted in altered quality, abundance, and timing of available food sources for bats [22]. These shifts have significantly impacted bat ecology in several key ways.

Roosting and foraging behaviour: Bats are increasingly selecting roosting and foraging sites near urban environment (as previously discussed in subsection 1.4.4). Notably, the exploitation of alternative food resources impacts their body condition. In fact, although novel food sources may be abundant, their lower nutritional or energetic value, poor digestibility, and harmful secondary metabolites can weaken or alter bat immunity, with downstream effect on viral dynamics [22]. Moreover, by relying on crop and other cultivated plants, the distribution of reservoir and recipient hosts has overlapped, leading to more frequent human-wildlife conflicts and creating greater opportunities for spillover events.

Bat movement pattern: The rate and distance of bat movements across the landscape have changed. *Pteropus* bats, known for their high mobility, regularly travel long distances to feed. However, the loss of native habitats, and the unintentional provision of novel resources through agriculture, had dislodged these bats from their natural habitat, forcing them to rely on alternative, yet suboptimal, food sources in close proximity to humans [22], [33]. As a result, despite their preference for highly nutritional native food, which provides a higher volume and more concentrated nectar than introduced plant species, bats are forced to make trade-offs [22]. Indeed, they must choose between energetically costly long-distance movements to access ephemeral and uncertain native resources, and the more stable year-round availability of abundant but suboptimal novel food in close proximity [22]. Wildlife responses to changes in the availability and quality of resources across landscape is species-specific [22]. For *Pteropus* bats in Bangladesh, studies suggest that their successful ability to adapt to changes, has led to a reduced migratory propensity and urban habituation. Therefore, despite their ability to cover large distances, these bats now predominantly occupy small, resident roosts where they share food resources with locals [22], [33].

However, both provision of suboptimal food and spatio-temporal redistribution of dietary resources can have several downstream effects on bat health and immunity to infections, as a consequence of nutritional stress, which in turn can affect the **prevalence** and circulation of HNV (as anticipated in subsection 1.4.4). Notably, in Australia and Bangladesh, spillover events are clustered in the cooler and dry winter months, when bats experience nutritional stress and resides in small roosts close to humans [33]. Additionally, bat movement patterns can have significant implications for the prevalence of HNV within bat populations. This is particularly relevant when considering how frequently bats move in relation to the duration of the pathogen's infectious period [22]. Specifically, short infectious periods combined with low movement rates among roosts can result in patchy viral distribution, whereas long infectious periods and frequent bat movement can lead to more homogeneous viral dynamics [3]. Since the length of NiV infectious period in bats is currently unknown, predicting its impact on transmission dynamics remains challenging [22], [33]. Two different hypotheses exist but they will be discussed in details in Section 1.5.2.

Altered bat density: Spatio-temporal patterns of resource availability influence key demographic factors in bat populations, such as births, deaths and dispersal, which in turn affect colony size and population density. Food-shortage and poor quality habitat tend to reduce population density by lowering birth rate, decreasing juvenile survival and by limiting the immigration of new individuals into the population [22]. Conversely, in areas where habitats are of higher quality and resources are abundant, population density tends to increase [22].

During periods of food-shortage-caused by seasonal changes, habitat loss or climate disruption-bats often aggregate in urban and peri-urban areas in search of alternative food sources [3]. For instance, in subtropical regions of Australia, large aggregations of flying foxes occur seasonally in response to transient pulses in fruit and nectar availability [33]. In contrast, bats in Bangladesh are in a permanent state of fission, residing in an increased number of smaller, disconnected, mostly resident roosts in a matrix of anthropogenic food resources [33].

Changes in local bat population density due to resource alterations could significantly affect the spatial dynamics of HNV transmission [22]. Bats in roosts are continuously exposed to low-level viral excretions in urine (see Section 1.5.3 for further details) [22]. If HNV transmission were solely density-dependent, higher population densities would likely lead to increased viral shedding and a greater risk of spillover events [22], [33]. However, it remains unclear how bat density scales with roost size or if HNV dynamics are purely driven by density-dependent processes.[22], [33].

Pathogen transmission can also follow a frequency-dependent pattern, where the rate of infectious contact remains constant regardless of host density [33]. The mode of pathogen transmission is crucial in determining how changes in host density influence pathogen dynamics, persistence within populations, and the likelihood of spillover to new host [33]. Furthermore, seasonal shift in transmission modes may further complicate the relationship between host density and pathogen spread [33].

Understanding how these transmission modes interact with bat population density is crucial for assessing spillover risks. If HNVs transmission is primarily density-dependent, resource-driven increases in population density would likely enhance viral shedding and spillover potential [22], [33]. However, factors such as roost connectivity and bat movement may also play a role, meaning that density alone may not fully explain spillover risks [22]. Indeed, roosts are not isolated systems but are interconnected via bat movement at the landscape level [33], [33]. Further investigation is needed to explore how the relative importance of density- and frequency-dependent factors changes seasonally.

Notably, findings from Bangladesh suggest that spillover events are more closely linked to higher roost density rather than bat population density [55]. Specifically, an increased number of smaller roosts, without a corresponding rise in the total bat population, appears to be associated with a greater risk of spillover events [55], [33]. This highlights the importance of considering roost structure and distribution in addition to population density when assessing HNV transmission dynamics and spillover potential.

1.5.2 Altered connectivity

Habitat fragmentation has disrupted the connectivity between and within bat metapopulations [36]. The landscape is now predominantly characterized by small, isolated, and irregularly shaped forest patches, interspersed with croplands and human settlements—a pattern common across the tropics [12]. This reduced connectivity alters contact rates among bats, affecting infection dynamics and pathogen spread. Consequently, these anthropogenic changes have undoubtedly influenced the frequency of spillover events into human populations [33].

Prevalence, viral load and shedding from infected bats As anticipated, the length of the NiV infectious period in relation to bat movement rates between roosts is unknown but could impact HNV prevalence in bat populations. Two hypotheses have been formulated to explain the temporal and spatial pulses of virus shedding in bat populations [3]:

1. **Acute infection, with or without waning immunity**, leading to transient epidemic waves of bat-to-bat transmission between bat populations [3].
2. **Persistent chronic or latent infection**, causing episodic **viral shedding** due to the recrudescence of chronicity or reactivation of a latent infection [3].

Acute infection and transient epidemics waves: This is characterized by a short infectious period and a low movement rate that promote patchy viral dynamics among roosting sites [3]. A decrease in meta population connectivity, due to urban settling, and a reduced bat migratory behaviour, is comparable to sedentary urban populations with waning population immunity (though not necessary waning of individual immunity) [22]. Therefore, if HNV infection in bats is acute, the disease quickly spreads and then fades out throughout the population in the absence of virus reintroduction [22]. Over time, births and immigration replenish the number of susceptible hosts, leading to extinction-colonisation dynamics of the virus [22]. Consequently, pulses of infection due to local virus extinction and recolonisation across roosts could generate epidemics that travel as waves of infections among hosts [3].

Persistent infection and episodic shedding: The alternative hypothesis is that bats carry persistent infections that do not cause apparent pathology or disease [3]. Therefore, if persistent infections are suppressed by the host's immune response, reactivation and episodic shedding (i.e., excretion of virus) could occur when intrinsic or extrinsic stressor weaken the bat immune response [3].
The bat's innate and adaptive responses to viral infection play a critical role in affecting HNV prevalence within the population, viral load within a host and virus shedding [22]. As previously discussed in Section 1.4.4 and 1.2, immune responses are energetically costly and resource-intensive to develop and activate [22]. Consequently, the maintenance and function of the immune system require allocation of host resources, making immunity dependent on nutrition [22]. Any dietary-shift (refer to Section 1.5.1) or metabolically demanding external stressors (i.e. thermoregulation, pregnancy, lactation, refer to Section 1.4.4) that limit the intake of energy and nutrients necessary to maintain an effective immune response has the potential to increase the individual's viral load [22]. Therefore, bat viral seroconversion results from an impaired immune reaction due to resource-related pressure or stress, which increases susceptibility to HNV infections, recrudescence of latent infections, or opportunities for transmission among bats, particularly if they aggregate in response to resource concentration [22] [3]. Consequently, poor nutritional conditions are more likely to trigger the reactivation of latent infections and recrudescence, either due to reduced immunocompetence or immunity trade-offs during periods of physiological stress, such as pregnancy [22]. In contrast, transmission between hosts is more likely influenced by resource distribution, which affects bat population density and connectivity [22].

1.5.3 Enabling conditions for virus spillover from bats to recipient hosts

The persistence and transmission of viruses operate on multiple levels, from within host cells, individual host, population of hosts to the broader host metapopulations and landscape [3]. At the individual level, hosts provide a habitat for viruses, but for viral propagation across populations, the virus must successfully replicate, exit the host, and be transmitted to new hosts [3]. While the host's immune system—both innate and adaptive—acts to suppress or eliminate the virus, certain viruses, like *henipaviruses, filoviruses, and coronaviruses*, seem to have co-evolved with bats, their natural reservoir hosts [3]. This long-term association may result in minimal pathology or clinical disease in bats, allowing the virus to persist without causing significant harm to its host [3]. In the case of *henipaviruses*, viral replication may also be limited, yet this interaction still enables the virus to survive and spread across bat populations and metapopulations [3]. Furthermore, the bat-borne viruses distribution among roosts is influenced by factors such as the duration of the infectious period and the movement patterns of bats between roosts (refer to Subsections 1.5.2 and 1.5.1) [3]. Antibodies, which reflect the cumulative exposure to viruses, are often widely distributed across bat populations and species communities, indicating consistent viral presence over time [3]. However, viral shedding tends to occur in episodic pulses, suggesting short infectious periods with possible cycles of local virus extinction and recolonization within roosts, or intermittent shedding from persistently infected individuals [3] (refer to Subsection 1.5.2). These episodic shedding events highlight the complexity of viral dynamics in bats, where even widespread antibody presence may not directly correspond to continuous viral transmission [3].

Building on this foundation, the following sections will examine the specific conditions and dynamics that facilitate the transmission of viruses from bats to recipient hosts, commonly referred to as spillover events. These events require a series of hierarchical enabling conditions, including the distribution of the reservoir host, the spatio-temporal dynamics of the pathogen within the host population (such as infection and shedding patterns), the survival or dispersal of the pathogen outside the reservoir host, the overlap between reservoir hosts and susceptible recipient hosts, and behaviors that increase the exposure risk for recipient hosts [22][3].

Within-host virus transmission dynamics

The amount of virus shed by reservoir hosts depends on several factors, including the number of bats in a given site, their infection levels (i.e., viral load), the amount of virus excreted into the environment, and the duration of their infection, which is influenced by their immune response [22]. *Pteropus* bats, which live in dense arboreal colonies, are near-continuously exposed to a low-level of viral shedding trough droplents or virus-laden aerosolized urine or faeces [3]. Indeed, seroprevalence studies suggest that HNVs are predominantly horizontally transmitted among bats (transmitted among individuals), which could plausibly occur through urination in the three-dimensional dense roost structures [22], [3]. This is supported by the fact that, in *Pteropus* bats, Nipah virus has been repeatedly found in urine, while rarely in oropharyngeal and rectal swabs [22]. Although vertical transmission of NiV (from mothers to offspring) has been observed, its significance as a transmission route in natural settings remains unclear [22].

Recipient host exposure

It is important to emphasize that spillover events occur when the reservoir hosts excrete a sufficient infectious dose and susceptible recipient hosts come into contact with the excreted pathogen [22].
Bat execrate the virus through urine, faces, and saliva in a drip zone around trees where they feed or roost [3]. While direct transmission of HNVs (i.e., through bites or consumption of bat birth products [30]) is probably rare, indirect transmission is likely to occur via contaminated food (such as, partially eaten fruit or contaminated Date Palm Sap) or fomites. Intermediate hosts can be exposed to HNVs by consuming contaminated food (i.e. grass, feed, fruit or water) or by touching, browsing or sniffing contaminated surfaces within the drip zone [3]. As already discussed in Section 1.4.3, numerous domestic and peri-domestic animals are susceptible recipient hosts for Nipah virus, including pigs and horses. In Malaysia, discarded fruit pulp was the route of Nipah transmission to pigs, while in Bangladesh, contaminated pots collecting date palm sap are the source of bat-to-human transmission route [3].
The rate and duration of food consumption within the drip zone can influence the accumulation of an infectious dose of bat virus [3]. Consequently, the survival of the virus in the external environment is a crucial factor in determining the zoonotic potential of NiV pathogen [22].

Survival of virus outside reservoir hosts: NiV has a limited survival time outside its natural reservoir host [22]. Typically, it can persists in urine and on contaminated fruit and fomites, with an average half-life of only few hours [22]. Consequently, spillover requires that contact between HNVs and recipient hosts occur shortly after excretion [22].

NiV, like *Hendra virus, filoviruses,* and *coronaviruses*, is highly sensitive to environmental factors such as temperature, pH, ultraviolet light, and desiccation [3]. As temperatures rise, particularly in summer or due to landscape changes like deforestation and urbanization, increased desiccation rapidly reduces the virus's survival [22]. In these conditions, viral particles dry out quickly, shortening the window for potential transmission. Conversely, in cooler months or during periods of higher precipitation and humidity, NiV can survive longer due to slower desiccation [22]. These conditions create a more favorable environment for viral persistence, extending the time during which recipient hosts might encounter and contract the virus from contaminated surfaces or food. Moreover, the impact of local microclimate fluctuations on virus survival also depends on the contaminated medium; if it is liquid, the effects of temperature and desiccation are reduced [22]. This is well represented by the case of virus contaminated collection pots of Date Palm Sap (DPS) in Bangladesh, which acts as the primary route for bat-to-human and human-to-human transmission of NiV [33]. Harvesting methods that do not use bamboo skirts to block fruit bat contamination of pots, along with the seasonal human consumption of raw DPS, has clearly linked spillover events to human behavior rather than environmental conditions [33], [22]. Indeed, even if the human consumption of fermented DPS, known as *Tari*, occurs all year-round, the higher spillover risk is associated with cooler and drier winter months, when DPS is consumed fresh within few hours of collection [56]. To support these findings, more detailed information and numerical data from the literature are provided. Field and laboratory tests have confirmed that:

- NiV can survive for up to 3 days in some fruit juices or on mangoes [30] [34].

- It remains infectious in neutral pH bat urine held at 22°C for up to 4 days, with a half-life of 18 hours [30] [34], [33], [35].
- In artificial date palm sap held at 22°C, consisting of 13% sucrose and 0.21% BSA in water at pH 7.0, the virus can persist for at least 7 days [30] [34].
- The pathogen is relatively stable and remains viable in palm sap held at 70°C for 1 hour. However, it is completely inactivated by heat at 100°C for more than 15 minutes and by common disinfectants such as sodium hypochlorite [30] [34].
- NiV is susceptible to alcohol, and a 60%–70% alcohol solution is recommended for sterilizing contaminated objects[46].

Transmission cycle from bats to humans

Humans can be infected by Nipah virus **through intermediate hosts**, such as domestic, peri-domestic animals, as well as trough **direct bat-to-human** (though less common) and **human-to-human** transmission [16].

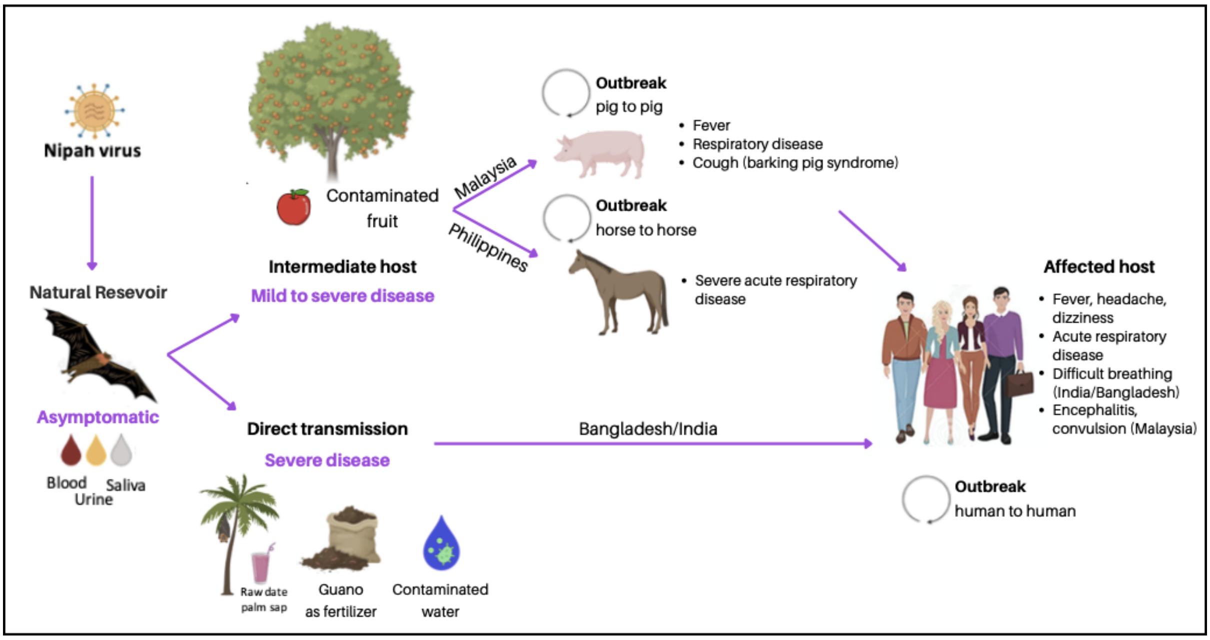


Figure 1.3: Routes of transmission for Nipah Virus (NiV). This figure has been revised and adapted from [16].

Indirect transmission: intermediate host-to-human and human-to-human As anticipated in Section 1.4.3, intermediate secondary hosts for Nipah virus include a wide range of domesticated animals, spanning six mammal orders, including pigs, dogs, cats, horses, hamsters, guinea pigs, sheep, cattle, goats, bats, and ferrets [57]. Notably, among these, pigs and horses have served as key viral amplifiers during outbreaks in Malaysia-Singapore and the Philippines, respectively [57].

Intermediate recipient hosts can be infected trough ingestion of contaminated fruit, water, feed or aborted bat fetuses and birth products (e.g. by pigs) [30], [3]. Humans, in turn, can be infected by sick animals:

- via consumption of undercooked meat [30];
- via direct (unprotected) contact with tissue and body fluids, trough skin abrasions or mucous membranes (e.g. aerosolization of respiratory or urinary secretions) [30].

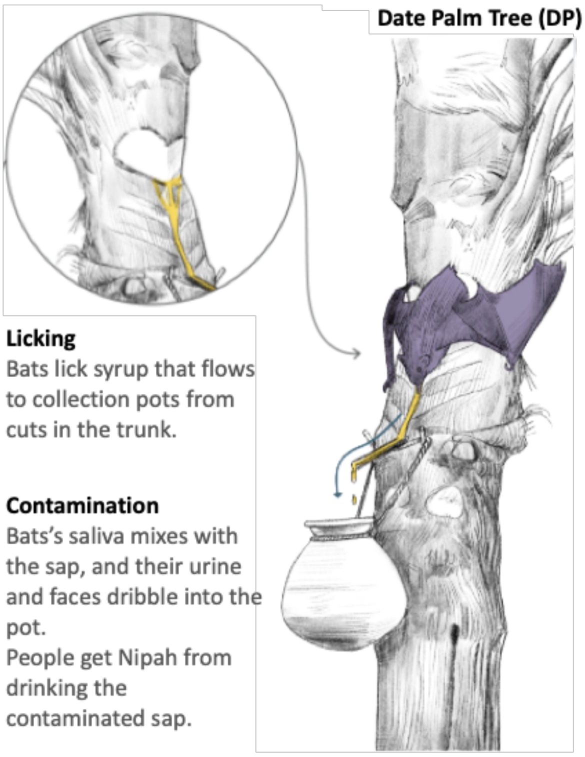
Person-to-person transmission of the NiV_B strain has been documented in Bangladesh and India [30], [57]. Humans can shed the virus in respiratory secretions, saliva, and urine, and are particularly at risk of contracting it after more than 12 hours of exposure to the body fluids of infected patients [30]. Additionally, it has been documented that some individuals have become infected after unprotected contact with deceased patients during the preparation of the corpse for burial [30].

Direct transmission: bat-to-person Humans can be also infected via direct exposure to bat saliva, urine and guano. This can occur through ingestion of contaminated food, such as partially eaten fruit or contaminated beverages (like Date Palm sap or water), or through activities involving exposure to bat secretions or excreta, like guano harvesting [32] or climbing trees [58]. These activities can lead to the inhalation of aerosols or droplets containing bat secretions. Bat feces are a popular fertilizer in Cambodia and Thailand. In rural areas with limited work opportunities, selling bat droppings can be a vital source of income for unaware guano harvesters [32].

Date Palm Sap case In Bangladesh and India, consumption of raw (unpasturized) date palm juice (i.e. Date Palm Sap) is the major route of NiV transmission from bat to humans [36], [33]. The cultivation of date palms has deep historical roots in these regions [36]. It is a seasonal business and a critical component for the local economy, especially for families in rural areas during the winter months when other work opportunities are scarce [36].

Fresh DPS is collected by shaving the date palm tree, placing a tap below the incision, and collecting the sugary fluid as it pours out of the trees into a clay pot overnight [57]. Fruit bats are known to visit these sap streams, and they may even urinate or defecate into the clay pots, thereby contaminating the raw sap with infectious Nipah virus [46].

Date palm sap, is an example of provisioned resource for flying foxes, as it becames available as a food for *Pteropids* during human harvesting [22]. This source attracts large numbers of *P.medius*, especially during winter, a season of natural resource limitation for bats in Bangladesh. These months coincide with increased sap production driven



Licking
Bats lick syrup that flows to collection pots from cuts in the trunk.

Contamination
Bats's saliva mixes with the sap, and their urine and faces dribble into the pot.
People get Nipah from drinking the contaminated sap.

Figure 1.4: A bat feeding on date palm sap. The sap from date palm trees can serve as a vehicle for the Nipah virus, facilitating its transmission from bats to humans. Source: [38].

by the repair mechanism of xylem plants in response to *embolism* caused by specific climatic conditions (i.e., a phenomenon known as *xylem refilling*, see Section A.2.1 in Appendix A for further details) [56].
The increased frequency of *P.medius* visits to date palm trees during winter months have been documented [56], and critically, this period coincides with the primary harvest season for fresh DPS (October to April). Fresh sap is collected during the night and sold in the early morning, often consumed within hours, before natural fermentation occurs and the sweet taste is lost [57], thus increasing the risk of virus spillover [56]. Although Nipah virus is sensitive to pH and alcohol, evidence from India shows that it can remain viable even in fermented sap (tari), which typically contains 5%–8% alcohol and a pH of 4.5–6.0 [46]. This level of alcohol may not fully inactivate the virus, allowing its persistence and potential transmission to consumers, particularly during winter months when temperatures range from 15 to 28 $°C$ [46].

Multiple products are made from date palm sap, including date palm wine (called “Tari”), jaggery (called “gur”) and sugar candy, which are consumed year-round [36]. However, spillover events predominantly occur during the colder and drier winter months [59], and result to be spatially clustered within the so called "Nipah belt", a group of districts in the central and northwest regions of Bangladesh [33]. Despite the association between raw date palm sap consumption and spillover risk, some Bangladeshi villages with high proportion of residents reporting this habits have not experienced outbreaks, suggesting that additional factors, such as local practices or environmental conditions, may influence spillover dynamics [59]. Notably, NiV causes high mortality rates in low-income, rural communities with limited access to healthcare [36].

1.6 Climate change and zoonosis: Nipah virus

Climate change is causing significant alterations in global weather and climate patterns. These extensive changes, driven by human activities such as the burning of fossil fuels and changes in land cover, are affecting the entire climate system, including the atmosphere, hydrosphere, cryosphere, and biosphere, as well as the complex interactions among them. Human-driven increases in greenhouse gases (GHGs) and aerosols have played a key role in the climatic changes observed during the twentieth century [60]. These alterations are expected to persist, leading to further changes in climate throughout the twenty-first century and beyond [60]. The resulting shifts in the composition and properties of atmospheric constituents, altering the energy budget of Earth, are likely to impact temperatures, precipitation patterns, sea levels, and the frequency of extreme events, all of which are critical to the stability of natural environments and human systems [60]. Moreover, the accelerated pace of these changes suggests that societies worldwide may face significant challenges in adaptation, with potentially severe consequences for specific regions.

According to the Sixth Assessment Report (AR6) of Intergovernmental Panel on Climate Change (IPCC), South Asia, with over one-fifth of the world’s population, is among the most vulnerable regions to climate change impacts, as indicated by the increasing frequency and severity of extremes such as cyclones, droughts, floods and heatwaves [61]. This vulnerability arises from a combination of factors, including widespread poverty, heavy reliance on natural resources and ecosystem services, and the region’s relatively narrow temperature ranges [62]. These challenges are further exacerbated by high annual population growth rates (2%, compared to the global average of 1.4% [63]) in what is already the most densely populated region in the world, along with increasing poverty and food insecurity [60]. Notably, Bangladesh is highly vulnerable to both natural disasters and climate change, ranking as the seventh most disaster-prone country in the world, according to the *Global Climate Risk Index* 2021[1].

In addition to its socio-economic vulnerabilities, South Asia is recognised as a deforestation hotspot, with some of the most impaired ecosystems in the world [26]. The extensive land-use change, fuelled by agricultural expansion and rapid urbanisation, further drive climate change with downstream effect on human health, and societal resilience [60]. These complex interactions between land-use alterations, biodiversity loss, and climate change, set within South Asia's unique geographic, demographic, and socio-economic landscape, create a vicious cycle that increases the risk of zoonotic disease spillover. This makes South Asia a hotspot for the emergence, re-emergence, and spread of zoonotic infectious diseases [64], [65], [66]. Addressing this requires an integrated approach, recognizing that protecting forests, reducing land-use change, and mitigating climate change are essential for reducing the risk of future zoonotic outbreaks [67].

The following sections will examine the current and projected impacts of climate change on South Asia, focusing on its relationship with zoonotic diseases. Key attention will be given to how rising temperatures and altered precipitation patterns reshape host, vector, and pathogen dynamics, along with their interactions [10]. By exploring the intersections between climate change, human health, and ecological conditions that influence zoonotic disease emergence, the following analysis will shed light on the importance of understanding these complex relationships underlying spillover events to better prepare for future health challenges driven by a changing climate.

1.6.1 Climate Change in South Asia

In South Asia, a significant portion of the population, particularly those living below the poverty line, depends heavily on climate-sensitive sectors such as agriculture, forestry, and fisheries [60], [63], which are particularly affected by rising temperatures, rising sea levels and changing rainfall patterns. While the region has experienced rapid economic growth in recent decades, lifting millions out of extreme poverty, this progress has come with new challenges. Land use changes, land degradation, urbanization, and pollution have disrupted ecosystems and altered climate dynamics [60]. As a result, many communities in this disaster-prone region remain highly exposed and vulnerable to weather and climate-related hazards [63].

South Asia is already experiencing the impacts of climate change, and these effects are expected to intensify in the coming decades, exacerbating the region's vulnerabilities. The key observed and projected climate changes for this region are outlined in the Sixth Assessment Report (AR6) by the IPCC, as detailed below.

Temperature trends: South Asia has already seen rising air temperatures, and this warming is expected to accelerate significantly throughout the 21st century [68], [60]. Projections suggest that, under a high-emissions scenario, average annual temperatures, could increase by more than 2 $°C$ across most of the region by mid-century, compared

---

[1] Climate Vulnerability Index (CVI)

to the 20th-century average [68]. Increases could exceed 3 °C, with high-latitude areas potentially seeing rises over 6 °C by the century's end [68]. Under a low-emissions scenario, temperature increases would be more moderate, remaining below 2 °C in most areas, except higher latitudes, where warming could reach 3 °C [68]. This rise in temperatures will have severe consequences for agriculture, particularly in tropical parts of South Asia, where crops are already grown near their temperature tolerance limits [60]. Climate-induced declines in crop yields will reduce food productivity, impacting livelihoods, exports, and exacerbating poverty [68]. For instance, in Bangladesh, these factors are projected to increase poverty levels by 15% by 2030 [68]. As will be further discussed throughout this thesis, the decrease in fruit yields will also influence the emergence of zoonotic diseases like Nipah virus. Nutritional stress in the reservoir host, driven by reduced fruit availability, can weaken its immune system, thereby increasing the likelihood of pathogen shedding and the risk of disease spillover.

Monsoon alterations: South Asia's climate is primarily governed by the monsoons, which consists of a seasonal reversal of the prevaling winds (in the direction and speed) in the lower troposphere, accompanied by dramatic shifts in precipitation regime between very dry to very wet [69]. The transition from dry to wet conditions typically happens suddenly around mid-June, but the timing, duration, and intensity of the monsoon rains fluctuate each year, leading to either floods or droughts [69]. The South and Southeast Asian monsoons (SAsiaM) cover a broad region, including countries such as India, Bangladesh, Nepal, Myanmar, Sri Lanka, Pakistan and parts of Southest Asia (i.e. Thailand, Laos, Cambodia, Vietnam, and the Philippines). The unique geographical features of this areas-including the Tibetan Plateau, Himalayas, Western Ghats, Arakan Yoma mountains, and the nearby Indian Ocean—play a crucial role in shaping monsoon dynamics [70].
The South Asian summer monsoon (SASM), a key component of the SAsiaM, brings abundant precipitation, from June to September, after a period of hot and dry weather contributing more than 75% of the annual rainfall in many parts of the region, including Bangladesh [71].

a) Observed: over the past century, the SASM has undergone significant changes, with far-reaching impacts on natural ecosystems, food security, and socio-economic conditions across the region [71]. One major driver of these changes has been the increased air pollution (i.e., anthropogenic aerosol forcing) over the Indian subcontinent and the rest of South Asia [72], [71]. These aerosols have acted as a barrier, lowering the temperature contrast between land and sea, a key factor in driving monsoon circulation [72], [71]. The cooling effect from aerosols diminishes both the intensity and frequency of monsoon rains by dampening atmospheric circulation and reducing rainfall [72], [71]. While some recovery in monsoon activity has been observed in recent decades, the long-term weakening during the 20th century continues to affect regional climate patterns [72], [71].

b) Projected: looking ahead, while near-term changes in monsoon precipitation are expected to be influenced by internal variability, long-term projections indicate that both South and Southeast Asian monsoon precipitation will increase during the 21st century, with enhanced inter-annual variability [72]. However, despite the anticipated rise in precipitation, the weakening of monsoon circulation due to various complex factors—such as reduced meridional land-sea thermal contrast and changes in atmospheric stratification—remains a critical concern for the region's climate stability and future sustainability [71].

Increased weather extremes: a) observed: according to the IPCC (2021) report, over the past few decades, South Asia has witnessed a marked increase in the frequency and intensity of extreme weather events. These include humid heat stress and heatwaves, characterized by a rise in the number of hot days and warm nights, alongside a decrease in cold days and nights [63], [72]. The region has also experienced more heavy precipitation events, leading to river floods and landslides, with significant damage to property, assets, and human life [63]. Additionally, the warming of the world's oceans has resulted in seawater thermal expansion, contributing to a faster-than-average rise in sea levels around Asia [72]. This has caused severe coastal erosion, shoreline retreat, and increased the frequency of coastal flooding, particularly during typhoons, which have led to substantial economic losses and fatalities [63], [60]. Saltwater intrusion has been reported up to 100 km inland along tributary channels of the Bay of Bengal during the dry season. Moreover, significant portion of mangroves, which play a crucial role in preventing saltwater intrusion, have been reportedly lost along the South Asian coastlines over the past 50 years, largely due to human activities [60].

b) Projected: looking ahead, the AR6 of IPCC projects a continued rise in mean surface air temperatures across all regions of Asia, with a shift towards more frequent and intense heat extremes [72]. Climate change is expected to amplify the urban heat-island effect, particularly in South and East Asian cities. Under a high-emissions scenario, extreme temperatures and precipitation are projected for nearly all cities, impacting freshwater availability, regional food security, human health, and industrial outputs [72]. The frequency of heavy precipitation events and related river floods is also expected to increase, along with annual mean precipitation and precipitation extremes.

In addition, glacier runoff is projected to increase until mid-century before declining due to reduced glacier storage [72]. Rising sea levels will continue to threaten coastal regions, exacerbating issues such as coastal erosion and saltwater intrusion. The ongoing climatic changes are likely to further constrain agricultural production, intensify water shortages, and heighten the risk of forest fires and degradation [72], [63]. These developments will also pose increased threats to coastal and marine resources and elevate the risk of outbreaks of infectious diseases [63].

1.6.2 Focus: climate's role in driving zoonotic disease emergence and transmission

Climatic changes play a critical role in the emergence and re-emergence of many serious infectious diseases (*see* **Emerging** infectious diseases and **Re-emerging** infectious diseases), alongside various human, biological, and ecological drivers [73].

Zoonotic pathogens are particularly climate-sensitive compared to those affecting only humans or animals, posing a significant challanges to public health and disease prevention efforts, especially in the context of a rapidly changing climate [74]. Each of the following points highlight a specific pathway through which climate change can increase the risk of disease transmission, setting the stage for the subsequent focused discussion on the role of bats in this context [75].

Host and vector distribution: As anticipated in Chapter 1, climate change exacerbates the risk of zoonotic virus emergence by altering ecosystems and shifting the geographical and altitudinal range of hosts and vectors [76], [13]. Indeed, warmer temperatures and changes in precipitation patterns can transform previously unsuitable regions into hospitable environments for these EID's carriers, leading to their expansion into new areas [76], [13], [75]. The resulting increased proximity with new species, facilitates pathogen cross-species transmission [32], [13], and amplifies the likelihood of viral spillover to humans [13].

Examples of these climate-driven shifts are already evident. In the U.S., the hispid cotton rat (*Sigmodon hispidus*), a reservoir for *Black Creek Canal hantavirus*, has expanded its range northward and to higher altitudes due to rising minimum temperatures, increasing hantavirus pulmonary syndrome (HPS) transmission risks [76]. Warmer conditions in southern Europe have similarly increased the abundance of *Culex mosquitoes*, vectors for *West Nile virus*, extended their active seasons and expanded their range into northern regions [76]. Additionally, due to climate change rodents that host *bubonic plague* and *tularemia* have shifted their ranges northward in North America, expanding areas at risk for these zoonotic diseases [76].

Vector reproduction, activity and pathogen transmission efficiency: Climate change can significantly accelerate the reproduction rates and activity levels of vectors, such as mosquitoes and ticks. Warmer temperatures can lead to more frequent breeding cycles, faster life-cycles (resulting in shorter lifespans), substantial increases in overall abundance, extended active seasons, and heightened biting rates, all of which raise the likelihood of pathogen transmission between vectors and hosts [77], [76], [78]. Additionally, climate conditions, particularly warmer temperatures, can accelerate the pathogen development within vectors (reducing the time needed for them to become infectious), and when interacting with humidity, can influence vector survival, vectorial capacity, and enhance the efficiency of pathogen transmission among hosts and vectors [77], [78].

Therefore, with more vectors active and reproducing rapidly, the transmission of diseases to humans, such as *West Nile virus, Dengue*, and *Zika* can intensify [76]. Moreover, faster pathogen development within vectors can lead to more rapid and extensive outbreaks. For instance, during warmer periods, the rate at which mosquitoes transmit *Malaria* can increase, resulting in larger and more frequent epidemics [76].

Pathogen persistence in the environment: Climate change can also impact how long pathogens survive in the environment outside of their hosts [77], [76]. Factors such as temperature, humidity, and precipitation influence the survival rates of pathogens in water, soil, and other **Fomites**. Prolonged pathogen persistence in the environment can lead to higher exposure risks for humans and animals (refer to Section 1.5.3 for the NiV case).

Zoonotic disease epidemiology: Climate change, including global warming and geoclimatic shifts, significantly alters the spread of zoonotic diseases by influencing the dynamics of hosts, vectors, and pathogens, as well as their interactions [10]. Many zoonoses are highly sensitive to climate variability and anomalies [17]. For instance, a strong link has been observed between the extreme weather events related to El Niño Southern Oscillation (ENSO) and outbreaks of *Rift Valley fever* (RVF), *Malaria, Cholera, Plague, Hantavirus*, and other emerging diseases [10]. ENSO (refer to Section A.1 for further details) has also been tied to the emergence of bat-borne viruses such as Hendra (HeV) in Australia and Nipah (NiV) in Malaysia [17]. Similarly, the North Atlantic Oscillation (NAO) has been associated with zoonotic diseases like *Lyme borreliosis* and *Tularemia*. Another example from the Western United States shows that *Plague* outbreaks have been linked to higher-than-average temperatures and the Pacific Decadal Oscillation (PDO), which leads to increased precipitation [10]. This, in turn, boosts food supplies for small mammals and promotes favorable conditions for flea populations, contributing to the spread of the disease [10]. As climate models predict more intense oscillation cycles in the future, the spread and impact of various zoonoses—including vector-, water-, food- as well as rodents and bat-borne diseases—are expected to increase [10].

Ecosystem degradation: Climate change coupled with land-use change (i.e. forest clearing for agriculture, urban development and road construction) irreversibly alter and deteriorate disease-carrying wildlife habitat. Notably, deforestation not only affect ecosystems but also drive climate change by releasing stored carbon and reducing the natural carbon sequestration capacity of forests [67]. As natural habitats are destroyed, wildlife species—particularly those that can adapt to living near humans, such as bats—are forced into closer contact with densely populated areas, heightening the risk of zoonotic disease transmission (refer to Chapter 1 for further details) [67]. Furthermore, the stress caused by habitat loss and poor living conditions in these disturbed environments can increase the likelihood of infected animals shedding pathogens [67].

1.6.3 Drivers of bat-borne zoonosis in Bangladesh: Focus on the Nipah virus

The **SARS-CoV-2 pandemic**[2] has highlighted that zoonotic diseases vary in their level of threat, with some posing particularly high risks due to their elevated **virulence** and **transmissibility**, resulting in a substantial **death burden** [1].

Bats, in particular, have been identified as reservoirs for several of these high-risk zoonoses, such as Nipah and Hendra viruses, Ebola, and various coronaviruses, including SARS, MERS, and SARS-CoV-2 [1]. The significant danger posed by bat-borne viruses underscores the importance of focusing on zoonoses originating from bats, making the **Nipah virus** a critical subject of this thesis.

As previously discussed, various stressors can negatively affect bats, leading to poor body condition, and ultimately, increased viral shedding and a higher risk of spillover. These stressors include land-use changes such as habitat loss, degradation, and fragmentation, as well as climatic factors (which will be addressed in the following paragraph). Another significant driver, particularly in Asia, is human behavior. In Bangladesh, activities like wildlife consumption, the widespread wildlife trade, bat harassment, hunting, and the consumption of date palm sap during the winter months, without measures to prevent bat access, have been identified as key factors facilitating cross-species transmission and the emergence of bat-borne viruses [33]. Moreover, consuming fruit partially eaten by bats and engaging in activities that bring humans into close contact with bat excreta and secretions—such as tree climbing and guano harvesting—further elevate the risk of Nipah virus (NiV) spillover [67], [32], [58].

---

[2] The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for the coronavirus disease 2019 (COVID-19) pandemic.

Climate-related drivers for Nipah virus

Anthropogenic climate change, which has already resulted in increased temperatures, altered precipitation patterns, and more frequent and intense extreme events like heatwaves, droughts, floods, and wildfires, not only affects bat distribution, with the implications outlined above, but its effects can also induce chronic stress and impaired body condition in bats, altering their immune system functioning and the dynamics of virus transmission among bat metapopulations (refer to Paragraph 1.4.4).

Bats' physiological sensitivity to climate stress: Bats are particularly sensitive to climate change effects due to their high surface-to-volume ratios, which make them prone to dehydration, a consequence of their small body mass and large wing and tail membranes [14]. Moreover, the absence of sweat glands in fruit bats hampers their ability to dissipate heat efficiently, making them particularly vulnerable to heat-induced stress and mortality [79]. To combat hyperthermia, which occurs when ambient temperatures (T_a) exceed their body temperature ($T_b \in [35 - 39]°C$) and reach lethal levels above 40 $°C$, large bats rely on thermoregulatory behaviors such as wing fanning, fluttering, licking their wings and chest, panting, and belly-soaking [80],[40], [79]. Although these behaviors are essential, they impose significant physiological stress due to their high metabolic demands. However, both extreme heat and cold temperatures present substantial energetic challenges, and the stress induced by these thermoregulatory responses can compromise their immune system function [80].

Rising temperatures and shift in precipitation pattern: Climate change also affect the redistribution of bat-dietary food sources across space and time altering plant phenology (refer to Section 1.5.1). This lead to **nutritional stress** in nectarivorous bats, especially when shifts in precipitation patterns alter flowering cycles [14]. This disruption compromises the availability of essential nutrients needed to sustain bats' immune systems, increasing the likelihood of viral recrudescence and transmission. Pregnant bats are particularly vulnerable to this nutritional stress due to the high metabolic demands of gestation and lactation. Notably, in Australia, the combined effects of nutritional and physiological stress have been linked to mass abortions and premature births in flying-foxes, highlighting the critical and indirect role that precipitation plays in the fertility and **reproductive success** of these mammals [81].

Additionally, climate-induced food shortages may also force bats to seek alternative food sources (often sub-optimal in their nutritional value, refer to Section 1.5.1 for further details), increasing their interactions with humans. Habitat fragmentation further exacerbates these challenges, as bats must travel greater distances between forest patches to forage, which not only increases their exposure to predators but also requires them to expend more energy and time to meet their dietary needs [54]. Additionally, the declining quality of these fragmented habitats can further limit available food resources, intensifying the strain on bat populations [54].

Chronic stress caused by these external factors has the potential to lead to an impaired body condition and to permanently disrupt the homeostasis of animals [54]. An weakened immune system may then led to increased pathogens susceptibility, higher rates of pathogen shedding, and reduced survival. These compounded stressors increase the likelihood of viral transmission and the potential spillover to other species, including humans [54].

Weather extremes: Furthermore, extreme weather events like thypoons can greatly diminish the concealment of roosting bats due to storm-related defoliation, thereby increasing their exposure to post-storm hunting and harsh abiotic factors, such as wind and rain [48]. After intense tropical cyclones, bats have been observed foraging on the ground or among fallen trees. However, their inability to take flight from the ground makes them highly vulnerable to predation during these periods, significantly increasing storm-related mortality rates [48].

The role of climate change mitigation: While the need for climate change mitigation is critical, emerging evidence suggests that even successful mitigation efforts may not necessarily reduce the risk of climate-driven zoonotic spillover events [13]. Recent studies indicate that milder global warming scenarios, where species can successfully track shifting climate optima, may actually lead to more first encounters between species, thus creating more opportunities for viral sharing across species [13]. These findings underscore the urgent need to pair wildlife disease surveillance systems and tracking bat range shift especially in tropical regions, as well as enhancing the public health infrastructure as essential components of climate change adaptation strategies, even if global warming is limited to below $+2\,°C$ above pre-industrial levels [13].

Ecological drivers of Nipah virus in Bangladesh

In Bangladesh, the winter months witness a convergence in the consumption of raw sap from date palms by both fruit bats and local communities [56] (refer to Section 1.5.3). So far, this phenomenon has been identified as the primary ecological driver of Nipah virus transmission in the region. However, drawing parallels from the Hendra virus case in Australia [82], it is plausible that nutritional stress, due to a scarcity of optimal food, and physiological stress associated with thermoregulation, lactation and gestation, play crucial roles in the increased viral shedding in fruit bats during the winter season in Bangladesh as well. Indeed, the convergence in palm sap consumption arises from several factors that collectively heighten the likelihood of transmission (refer to Figure 1.5 for timing details):

1. **Food availability and xylem refilling:** During the winter months, fruit availability for bats reaches its annual low (i.e. only 12 fruits available out of 49) [33], whereas the availability of date palm nectar peaks, owing to the climate-induced xylem refilling phenomenon (refer to Section A.2.1 for further details), which also enhances its sugar content [56]. Date palm sap closely resembles the caloric density and macronutrient profile of other fruits in the diet of *P. medius*, offering high levels of certain vitamins and minerals [56]. As such, date palm sap may serve as a crucial dietary supplement during a season marked by reduced availability of other fruits, leading to more frequent visits by bats to date palm trees, as documented in [56]. However, despite its nutritional content, DPS is not an optimal food source for bats, potentially compromising their nutritional status and immunity, thereby increasing their susceptibility to pathogens and the recrudescence of viruses like Nipah [22].

2. **Feeding behavior and contamination:** Fruit bats exhibit messy foraging behaviors, often dropping and dispersing partially eaten fruits and chewed leaves in the surrounding environment. Additionally, these bats commonly urinate and defecate while foraging (refers to Section 1.4.4 for further details). As discussed in paragraph 1.5.3, bats feed on the sap of date palms by lapping the nectar with their long, flexible tongues [38]. During this feeding process, their saliva mixes with the syrup, and urine

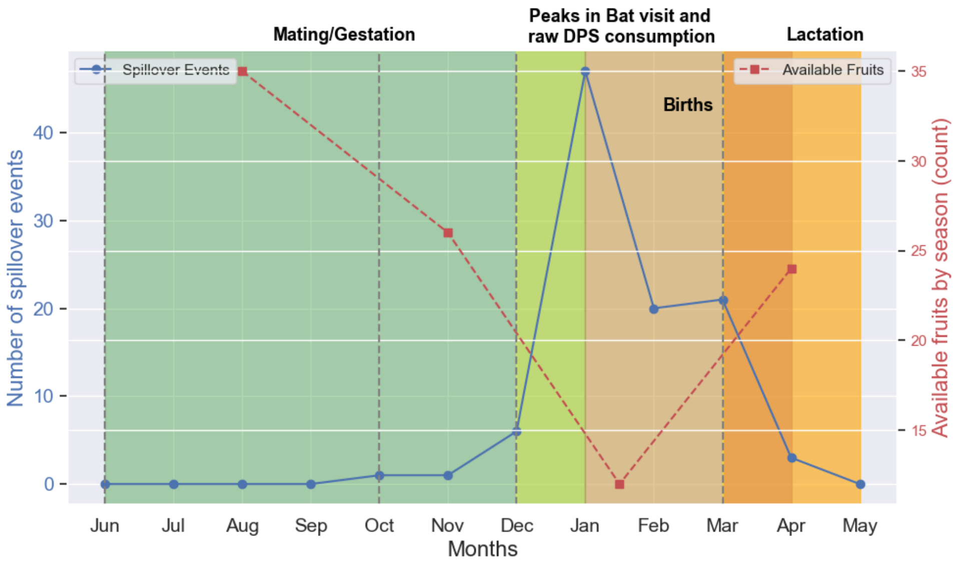


Figure 1.5: Temporal overlap of ecological drivers of spillover events in Bangladesh (monthly total across 18 years), highlighting the interplay of bat's nutritional stress (fruit availability [56]), physiological stress (gestation/lactation [48]), thermoregulatory stress (seasonal changes), and the alignment of bat visits with human consumption of raw date palm sap [56].

and feces can drip into the collection containers, contaminating the raw sap with Nipah virus [38], thereby heightening the risk of food-borne transmission to humans.

3. **Virus survival in low temperatures:** During the winter months, the virus has enhanced survival in the external environment (DPS) due to lower temperatures (refers to Sections 1.4.3, 1.6.2 and 1.5.3).

4. **Human consumption practices:** In winter, date palm sap is consumed raw and within hours of collection, without undergoing any fermentation. As the Nipah virus is sensitive to pH variation, alcoholic content and rise in temperature, this seasonal practice increases the likelihood of the virus remaining infectious in the juice (refer to Section 1.5.3).

5. **Breeding season and immunosuppression:** According to data on the reproductive cycle of *P.giganteus* in Bangladesh, the months from January to March coincide with the birthing period [48]. Literature suggests that climatic factors such as temperature, precipitation, and humidity during the six to nine months preceding birth—encompassing the gestation period, which lasts between 4.6 to 6.3 months—affect food availability and consequently contribute to nutritional stress during this energetically demanding phase [83], [82].

   Moreover, the dry winter season in Bangladesh, from December to February, overlaps with the end of the gestation period and the beginning of lactation for *Pteropus* bats. As previously discussed, a significant association between pregnancy and increased susceptibility to pathogens in bats has been observed [84]. Notably, the high metabolic demands of gestation can be further exacerbated by the nutritional and thermoregulatory challenges typical of winter. This combination of stressors during the pre-winter and winter periods can increase bats' vulnerability to infections such as Nipah virus, rising the likelihood of viral spillover during pregnancy or lactation phases.

1.7 Limits identified in NiV research and their influence on thesis choices

An in-depth review of the existing literature has revealed several critical limitations in the current understanding of the relationship between climatic factors and the Nipah virus (NiV) spillover phenomenon in Bangladesh. These limitations, along with potential improvements, have guided the methodological choices and the focus of the subsequent analysis in this thesis. Specifically, the study tackles gaps in data resolution, the choice of climate variables, and the temporal scope of analysis, which are essential for accurately assessing NiV spillover risks. The following key points outline the identified limitations and the corresponding improvements introduced in this thesis:

1. **Knowledge gaps and research biases:** Despite the crucial role that bats play in ecosystems and their significance as vectors of zoonotic diseases, our understanding of how climate and climate change specifically impacts bat ecology, and the dynamics of viral transmission remains limited. Research on bats is significantly underdeveloped compared to studies on other mammals and birds [85]. The challenges of studying bats, such as their tendency to select inaccessible roosting sites, their nocturnal activity, and their high mobility—allowing them to rapidly cross geopolitical boundaries—complicate data collection and monitoring [86]. These challenges have led to limited information on their actual distribution ranges and population dynamics. Moreover, research effort are geographically biased towards Europe, North America and Australia, and temperate and Mediterranean biomes, living significant knowledge gaps, in regions like Asia, where the risk of zoonotic spillover is most pronounced [14]. For instance, in the extensive literature review cited in [14], only 6% of studies where focused on Asia. Notably, over 80% of the papers analysed (published between 1930 and 2020) were released in the last decade, partly driven by the heightened demand for bat-related data following the COVID-19 pandemic. Furthermore, these studies exhibit significant taxonomic disparities, with only 7% of articles focusing on *Pteropodidae* and a mere 0.014% on *Pteropus*, the primary NiV reservoir [14]. Therefore, one of the key reasons this thesis focuses on studying *Pteropus* species in Bangladesh in relation to climate and NiV spillover events is to address these significant gaps in the existing research, recognizing the substantial potential for advancing our understanding in this critical area.

2. **Climate variables and modeling limitations:** To date, modelling studies primarily address how changes in average temperature and precipitation affect bat species distribution and behavior. However, they often overlook the impact of extreme weather events such as droughts, heatwaves, and cyclones, which are increasingly frequent due to climate change, as previously discussed. To address this gap, another key improvement in this work involves exploring the correlation between extreme weather events and NiV spillover incidents. This approach implicitly account for the potential impact of extremes on nutritional and physiological stresses, and the subsequent increased viral shedding by fruit bats.

3. **Seasonal focus in Nipah virus research:** Research on the Nipah virus in

Bangladesh has predominantly concentrated on the winter season, coinciding with the majority of spillover event occurrences. However, this narrow seasonal focus overlook the continuous year-round consumption of date palm sap, the primary ecological driver in the transmission of NiV int his region. Additionally, as seen with the Hendra virus in Australia, this approach fails to account for the potential delayed effects of external stressors on bats' immune systems, which, when impaired, can lead to viral recrudescence and increased shedding. To address these gaps, this thesis introduces an expanded temporal analysis to better understand the year-round dynamics of spillover.

4. **Data sources for climate impact studies:** Many studies on bat (e.g., current and predictive modeling of habitat suitability, range shifts, and mapping of spillover risk areas) rely on data from sources like WorldClim or ground-based meteorological stations, which may not capture the full complexity of climate-bat-spillover interactions, especially at tropics (refer to Section 3.1.1 for further details). This thesis, therefore, utilizes ERA5 reanalysis data, which offers improved spatial and temporal resolution and greater accuracy, allowing for a more detailed analysis of how climate may shape zoonotic spillover.

Chapter 2

The Bangladesh case

2.1 Description of the study region

Geography and topography Bangladesh, extending between latitudes 22 and 27°N, is a sub-tropical country bordered by India to the west, north and east, Myanmar to the south-east and the Bay of Bengal to the south [87]. Except for the hilly terrain in the southeast, the majority of the country (approximately 80%) consists of low-lying floodplains and deltas formed by deposition of sediments carried by the major river systems (the Ganges, Brahmaputra, and Meghna), which originate in the Himalayas[87], [88]. The topography is marked by minimal elevation variation between adjacent ridge tops and center depression (ranging from 1 to 6 m a.s.l), which renders this country highly prone to flooding events (Refer to Figure 2.1a) [87].

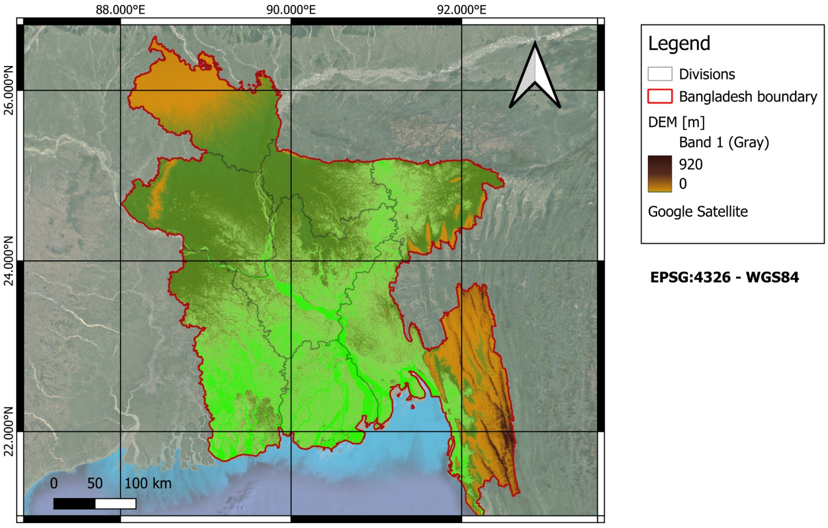
2.1.1 Climate of Bangladesh

Bangladesh experiences a typical monsoon climate, characterised by high humidity, moderately warm temperature and marked seasonal variations in rainfall [89]. From a meteorological perspective, four distinct seasons can be identified in the country with the following general characteristics [87], [89]:

- **Winter** (December to February), is characterized by cold and dry conditions, with average temperature ranging from a minimum of 7.2 to 12.8 $°C$ to a maximum of 23.9 to 31.1 $°C$ [87]. As illustrated in Figure 2.3a, a clear south-to-north thermal gradient exists, with southern districts being generally 5 $°C$ warmer than northern ones [87]. Sylhet, located in the northeast, is the coldest district, where minimum temperatures occasionally drop below 5 $°C$, though frost is extremely rare [87]. January stands as the coldest month, with an average temperature of 17.9 °C over the reference period 2000-2018 (refer to Figures 2.2a and A.3a).
  Winter is also the driest period, with the lowest precipitation levels observed in the southern regions, particularly in the Cox's Bazar district (refer to Figures 2.3b and A.3b).

- **Pre-monsoon** (March to May) is characterized by hot summer conditions, with average maximum temperatures reaching 36.7 °C and very high evaporation rates [87]. The temperature gradient during this season generally moves from the southwest toward the northeast (Figure 2.3a), and in some areas, temperatures can occasionally rise above 40.6 °C [87]. May is typically the hottest month, with an average maximum

(a) Digital Elevation Model



(b) Population density (2001)

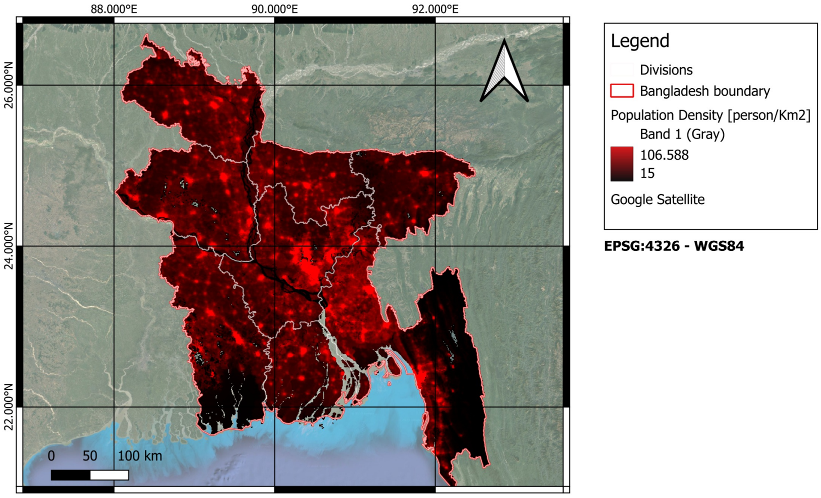
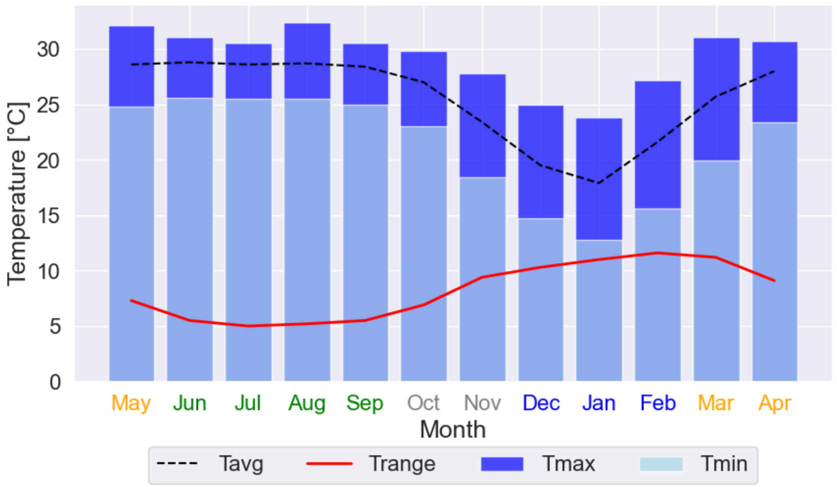


Figure 2.1: (a) Digital Elevation Model (DEM) of Bangladesh elaborated in QGIS software. Data source: SRTM DEM, from NASA web site, 1 arc sec (approximately 30 m) resolution. (b) Red areas represent the population density concentration across Bangladesh in 2001. Source data: WorldPop, 1 km resolution.

temperature of 32.1 °C during the reference period (Figure 2.2a). This season also experiences erratic rainstorms, often accompanied by heavy rainfall [87].

- **Monsoon** (June to early-October) is both hot and humid, bringing heavy torrential rainfall throughout the season. More than 75% of the mean annual rainfall occurs during this period, driven by the moisture-laden south-west trade winds which are drawn to the Indian sub-continent by the intense heat and consequent low pressure over Punjab (in Pakistan and India) and the Upper Ganges Valley [87]. During the 18-year period, the average monsoon temperature across Bangladesh was 28.6 °C (refer to Figure 2.2a),with the central districts consistently experiencing higher

(a) Monthly average temperature



(b) Monthly precipitation

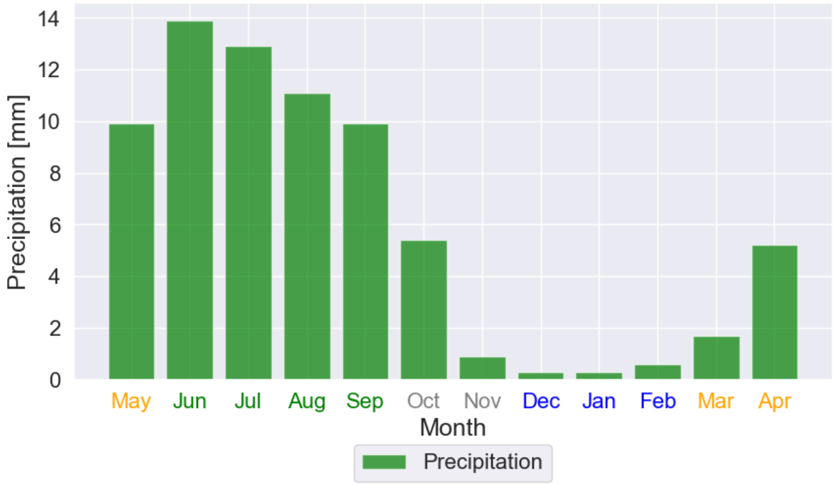


Figure 2.2: (a) Monthly temperature (average, maximum, minimum, and range) and (b) precipitation averaged across Bangladesh during the reference period from 2000 to 2018.

temperatures compared to other regions, as shown in Figure 2.3a.

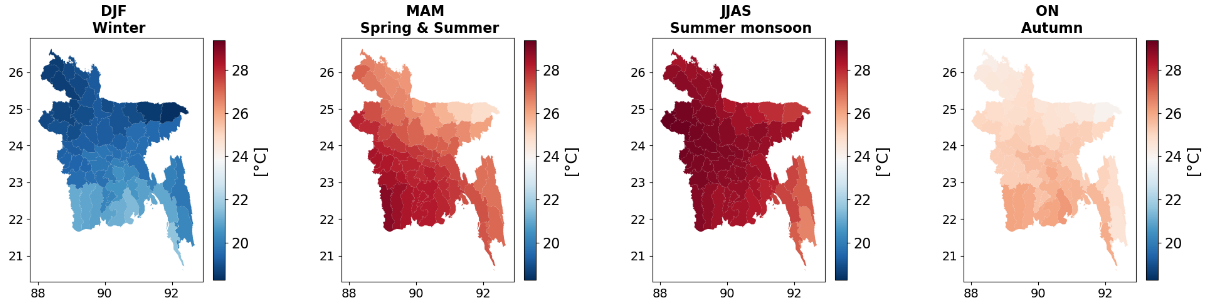
- **Post-monsoon** (late October to November) is a brief season characterized by withdrawal of rainfall and a gradual decrease in nighttime minimum temperatures [87].

The mean annual rainfall exhibits significant spatial and temporal variation, ranging from 1400 mm in the west to more than 4400 mm in the est, with a west-east gradient of almost 7 mm/km [89]. The higher precipitation levels in the Sylhet division are attributed to the orographic effect of the Meghalaya Plateau [89].

2.1.2 Climate change exposure

Bangladesh faces a range of climate-induced hazards, including climatological events such as droughts, hydro-meteorological events like tropical cyclones, tornadoes, storm surges, and floods, as well as geophysical threats (e.g., landslides, and coastal and riverbank erosion). The country's coastal regions, with an average elevation of less than 1.5 m a.s.l., are particularly vulnerable to storm surges, coastal flooding, and salinity intrusion [87]. Therefore, understanding the impact of climate change is crucial in assessing how it exacerbates these already significant hazards, which have historically caused considerable destruction. Notable examples include the severe floods of 1988, 1998, 2004, and 2007, and the cyclones and tidal surges of 1991, 1998, 2000, 2004, and 2007 [89].

(a) Average temperatures



(b) Average precipitation

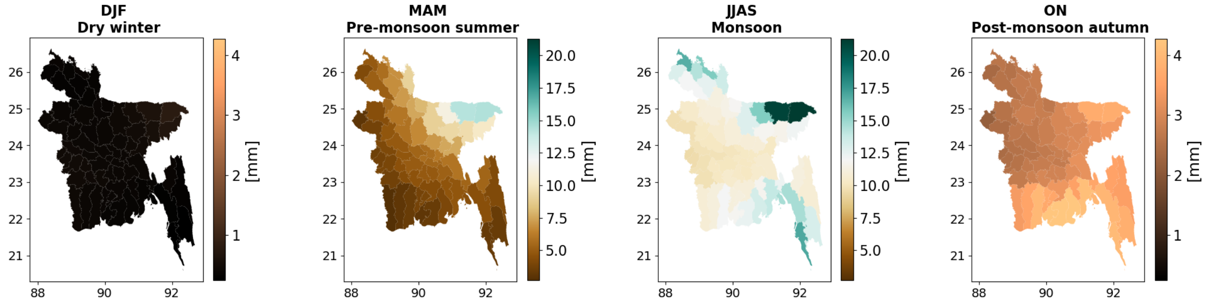


Figure 2.3: Spatial distribution of district-level seasonal averages for (a) mean temperature and (b) precipitation across Bangladesh during the reference period (2000-2018).

2.1.3 Demographics and economy

Bangladesh, the most densely populated country in the world and the eighth-most populous, has experienced significant population growth during the last decades (Our World in Data). From 129.19 million people in 2000, the population rose to 163.68 million by 2018 (refer to Figure 2.4), and reached 166.3 million in 2020 (World Bank, 2020). Projections indicate that this number will increase to 192.6 million by 2050 (FAOSTAT, 2020). A corresponding rise in population density was also been observed, climbing from 992 to 1,257 of people per km^2 between 2000 and 2018. Rapid and massive urbanization has shifted the proportion of the population living in urban areas from 5% to 28% over the past 40 years, with approximately 45 million people now residing in urban settings [90]. Dhaka, the capital, stands out as a megacity with a disproportionately higher population density compared to other cities, as clearly illustrated by the dense red area at the center of Figure 2.1b. Despite its exceptionally high population density, Bangladesh remains moderately urbanized (refer to Build-up thematic class in Figure 2.5).

In this overpopulated country, marked by a continual influx of people toward cities driven by poverty, climate change, and the promise of better economic opportunities, unprepared urban areas with deteriorating infrastructures worsen the already challenging living conditions of urban poor people [90]. GDP per capita is low, with 24.3% of population living below the poverty line in 2016/2017 (Wold Bank, 2023). Although GDP has risen rapidly in recent years, it remains significantly lower than in other countries, living a considerable portion of the population underemployed, with limited access to healthcare [90].

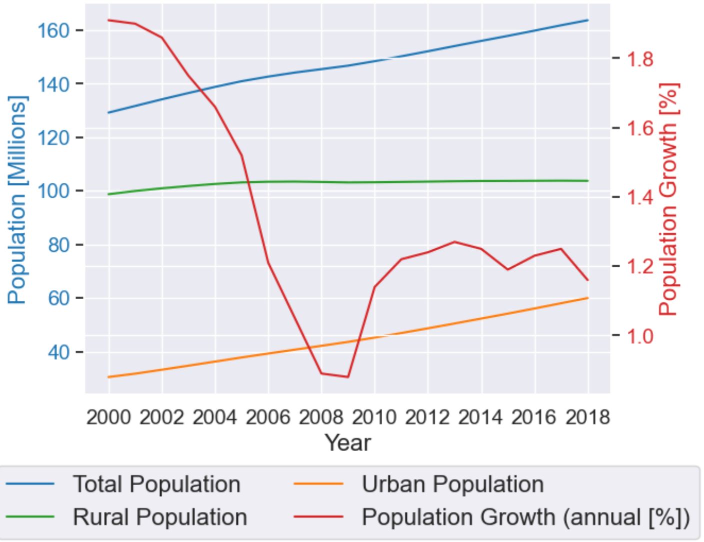


Figure 2.4: Population trends and growth rate in Bangladesh (2000-2018). Data source: World Bank.

Bangladesh covers an area of 130,170 km^2 (World Bank, 2020), with over 50% of this land dedicated to cropland (for detailed percentages, refer to Figure 2.5 and for spatial distribution to Figure 4.1). The agriculture sector contributes 14.74% to the country's GDP, providing employment about 41% of the labor force (Ministry of Finance, 2017), playing a crucial role in sustaining the livelihoods of the poor.

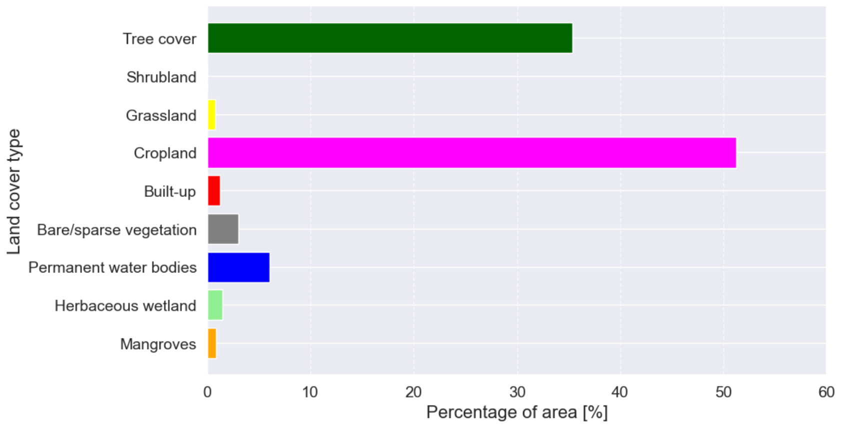


Figure 2.5: Land cover and land-use in Bangladesh for 2020. Data source: ESA WorldCover.

2.1.4 Land-use and Nipah belt

In Bangladesh, forest loss and its conversion to mostly cultivated areas, have drastically reduced the native habitat for *Pteripus medius.* These bat now persist year-round in an increasing number of smaller roosts on or near human dwellings, opportunistically feeding on cultivated food resources [36], [33]. The villages in central and northwestern Bangladesh, where most Nipah spillover events occured (the so called **Nipah Belt**) are characterised by high population density, high forest fragmentation and land cover dominated by irrigated and rainfed croplands, intersparsed with grassland and forest [36], [33]. All these factors are notably associated with the increased transmission of zoonoses, as discussed earlier.

Moreover, Nipah belt villages exhibit less forest cover, higher forest patch density and less canopy cover, all of which positively correlate with NiV events [55]. Although no significant differences in the total number of bats, bat population density, or proximity of roosts to human settlements have been detected within and outside the Nipah Belt, villages affected by spillover events report a greater abundance of *Pteropus giganteus* roosts [55], [33]. This altered roosting behavior, likely a response to habitat fragmentation, could have implications for Nipah spillover, increasing the likelihood that fruit bats feed on human resources, such as fruits from home gardens and date palm sap (DPS) collection containers, particularly in regions with higher population density [55], [33].

Additionally, roosts within Nipah belt are characterised by greater biordiversity compared to those outside this area, likely due to species-rich household gardens, orchards, and agroforestry [55]. This may support larger bat populations, suggesting that flying foxes within the Nipah Belt could expand alongside the human population [55]. Notably, two tree species, the *silk cotton* and *Indian mast trees*, have been associated with Nipah virus spillovers [55]. These trees are critical food resources for bats during their brief flowering phase in winter, a period of food scarcity for bats. As a result, bats congregate around these limited resources, increasing the likelihood of Nipah virus transmission within bat populations [55]. Subsequently, the virus can spill over to humans when infected bats drink from DPS containers [55].

As Bangladesh's population continues to grow, urban expansion will increasingly encroach upon renmant forested areas, significantly contributing to the emergence and re-emergence of infectious diseases in humans, including zoonoses such as Malaria, Yellow fever, Hantavirus, and Hemorrhagic fevers [55]. Notably, like other zoonotic diseases globally, Nipah virus causes high mortality rates in impoverished, rural communities where health systems are inadequate, fragmented, and ill-equipped to handle sanitary emergencies [36], [38].

Spillover events are often sporadic in both space and time, with repeated outbreaks being rare, further complicating the understanding and managment of spillover risk [33]. However, over the past 22 years, the Henipavirus strain, NiV, has emerged almost annually in Bangladesh and sporadically in other neighboring regions, causing multiple outbreaks and resulting in significant loss of life [15]. Although these outbreaks have been confined to South and Southeast Asia, the rapid expansion of spillover-prone areas—driven by human activities discussed earlier—has led scientists to speculate that **NiV could be the next pandemic agent after COVID-19 [15], [16]. This underscores the critical need for a deeper understanding of the mechanisms underlying NiV spillover events, which is the primary focus of this thesis**, with the ultimate goal of supporting the development of targeted and informed mitigation strategies.

2.2 Spatio-temporal reconstruction of Nipah virus spillover events in Bangladesh (2001-2018)

This study used Nipah virus case data from Bangladesh publicly available from two authoritative institutions: the World Health Organization (WHO) and the Institute of Epidemiology, Disease Control and Research (IECDR) of the Government of Bangladesh. The initial goal was to map the spatial and temporal distribution of NiV cases in the country during the period 2001-2018 and provide a detailed assessment of case counts at the district level.

In Bangladesh, detection of **Nipah-cases** relies on a hospital-based surveillance system (established in 2007 and made up of a network of three public tertiary-care hospitals [91]) and outbreak investigations conducted by the **icddr,b** in partnership with the Institute of Epidemiology, Disease Control and Research (IECDR) [92]. The annual totals of NiV cases (N_c) from 2001 to 2018 are summarized in Table 2.1. These NiV cases are classified into two main types of cases, defined as follows [92]:

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
N_c	13	0	12	67	12	0	18	10	4	18	43	17	31	37	15	0	3	4

Table 2.1: Annual count of NiV cases for year range 2001-2018. Sources: WHO, IECDR.

- **Primary cases**, refer to individuals who contracted the virus directly from bats, also known as **spillover events**, or where no human-to-human transmission pathway has been identified [93]. In this category are also included **Index cases**, defined as distinct and independent initial infections within a given outbreak year [33], that do not include cases resulting from secondary human-to-human transmission.

- **Secondary cases**, identify cases that occur when Nipah virus symptoms develop 5-15 days after close contact with a confirmed or probable case, corrected by the assumed incubation period [92],[93].

Additionally, the definition of NiV cases includes both [92]:

- *Laboratory-confirmed NiV cases* are defined as cases of meningo-encephalitis with detectable serum Nipah IgM antibody or detectable DNA in serum, throat swabs or cerebrospinal fluid [92].

- *Probable NiV cases* refers to meningo-encephalitis cases that meet the criteria for suspected Nipah virus infection and had epidemiological links to a confirmed NiV case within three weeks of illness onset but whose blood was not collected due to death, or where an the initial negative result was not followed by a second test due to death [92].

As a second step, the temporal distribution of the total yearly spillover events was reconstructed for the reference period 2001-2018, using data from three key sources: Nikolay et al. (2019) [92] for the 2001-2014 period, Cortes et al. (2018) [59] for 2007-2013 and McKee et al. (2021) [33] for 2001-2018. While McKee et al. (2021) [33] served as the primary reference, adjustments were made for 2012 and 2019 to align the counts with those reported in other sources. The final annual counts used in this study are shown

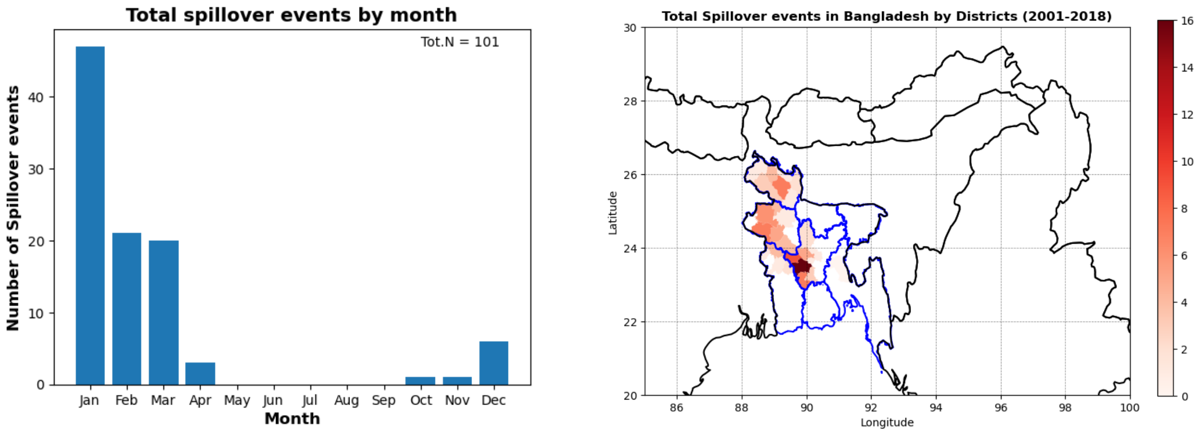
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Sources
N_s	1	0	1	15	2	0	5	2	2	12	9	10	17	11	9	0	2	3	[33],[59],[94]

Table 2.2: Annual count of spillover events considered in this study. Sources: [33], [59], [94].

in Table 2.2. After establishing the annual count of NiV spillover events, the analysis proceeded by collecting additional information through a thorough year-by-year review of each case from the available literature to reconstruct their spatial distribution across Bangladesh and gain higher temporal resolution. Spillover events were classified as ‘real’ when corroborated by multiple sources or explicitly reported in the literature. In contrast, events were classified as ‘dubious’ when precise information was lacking or based on general assumptions. In this case, an average incubation period of one week was assumed to estimate the spillover date from the onset of symptoms. The methodology used to reconstruct and assign these events is detailed in Appendix A (refer to Section A.2.2), with the final results presented in Tables A.2 and A.3.

2.3 Preliminary data analysis

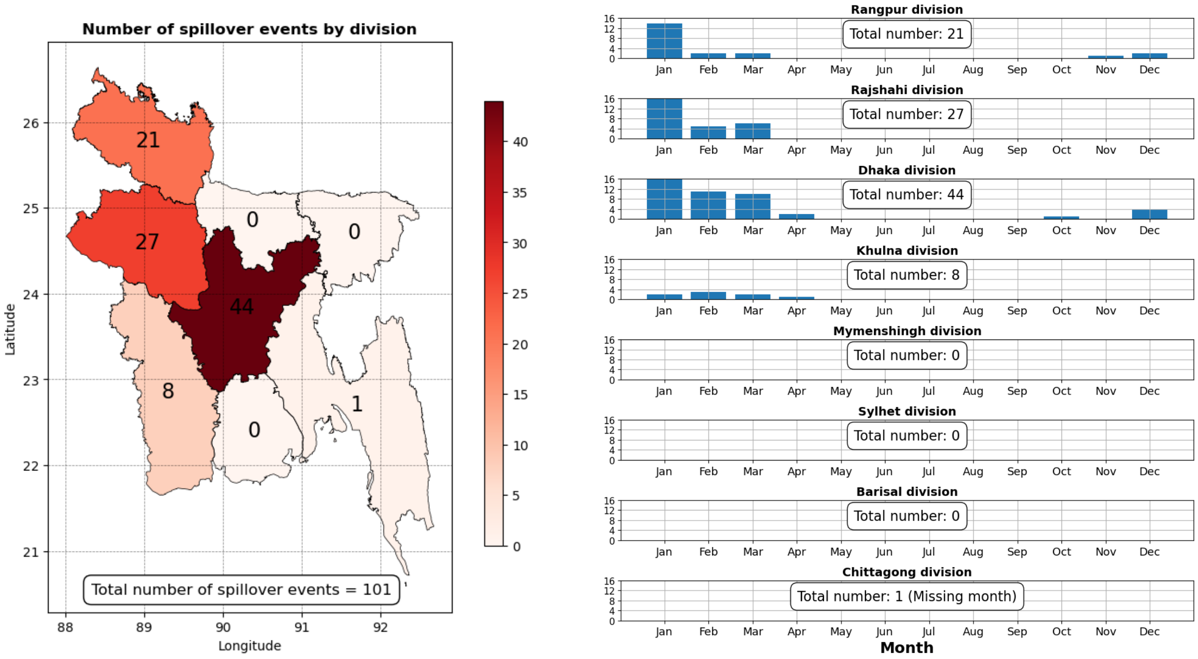
Exploring spatio-temporal trends The aim of this analysis was to pinpoint the specific months, districts, and divisions in Bangladesh that were most impacted by NiV spillover events. The subsequent Figure 2.6 provides a temporal (2.6a) and spatial (2.6b) illustration of the results. The highest count of events is observed in January, followed by a drop to half both in February and March. The remaining months show a sharp decline, with no occurrences reported from May to September and only minimal activity during the autumn months (October-November). This pattern suggests a concentration of spillover events in the early part of the year and a seasonal trend where the outbreaks typically occur between December and April, with winter (December- February) as the primary season for spillovers; a pattern as also reported in other studies from Bangladesh [33], [59].



(a) Total spillover events by month. **(b)** Total number of spillover events by district.

Figure 2.6: Comparison of total spillover events by month and by district.

These events also appear to be spatially clustered in the central and northwest districts of the country, as similarly reported in previous studies [33], [59]. The Dhaka division ($n = 44$) reported the highest number of spillover events across the eighteen-years period analysed, followed by the Rajshahi ($n = 27$) and Rangpur ($n = 21$) divisions, respectively (Figure 2.7a). Notably, within the Dhaka division, the Faridpur district stands out, experiencing a disproportionately high number of spillover events—16 in total—compared to all other affected districts. Other divisions show relatively lower counts, highlighting regional variations in spillover occurrences.



(a) Total spillover events by division. **(b)** Total spillover events by division and month.

Figure 2.7: Comparison of total spillover events by month and by division.

The bar plots (Figure 2.7b) confirm the consistent seasonal trend across all divisions, with peak events occurring in January, February, and March. This similarity across regions underscores the seasonal nature of spillover events, including local patterns highlighted by a division-specific analysis within each region.

Definition of two year groups: High and Low spillover years Based on these spatio-temporal patterns and to facilitate data analysis, the reconstructed annual number of spillover events from 2001 to 2018 were adjusted to fall within the same time frame. Specifically, the annual count of events for each year is now defined as spanning from May to the following April. This adjustment, detailed in Table 2.3, ensures that the annual count of spillovers aligns within the end of the considered year. This approach also facilitates the investigation of the pre-winter climatic conditions that may have influenced the number of events in the subsequent months, accounting for potential lag effect.

A second adjustment was to cluster the data in two distinct groups: 7-years with a high number of annual spillover events and 11-years with a low number. This classification aims to determine whether years with a high frequency of spillover events exhibit specific patterns in temperature (T) and precipitation (Pr) that could account for differences in spillover counts compared to years with lower occurrences. To ensure the inclusion of all

Year	2000/ 2001	2001/ 2002	2002/ 2003	2003/ 2004	2004/ 2005	2005/ 2006	2006/ 2007	2007/ 2008	2008/ 2009	2009/ 2010	2010/ 2011	2011/ 2012	2012/ 2013	2013/ 2014	2014/ 2015	2015/ 2016	2016/ 2017	2017/ 2018	Total
N_s	1	0	1	14	3	0	5	2	2	9	12	8	17	13	9	0	2	3	101

Table 2.3: Annual number of reconstructed spillover events for each year from 2000/2001 to 2017/2018. The year is defined from May to April, rather than January to December. The numbers highlighted in green represent the annual count variations resulting from the shift from the January-December year to the May-April year.

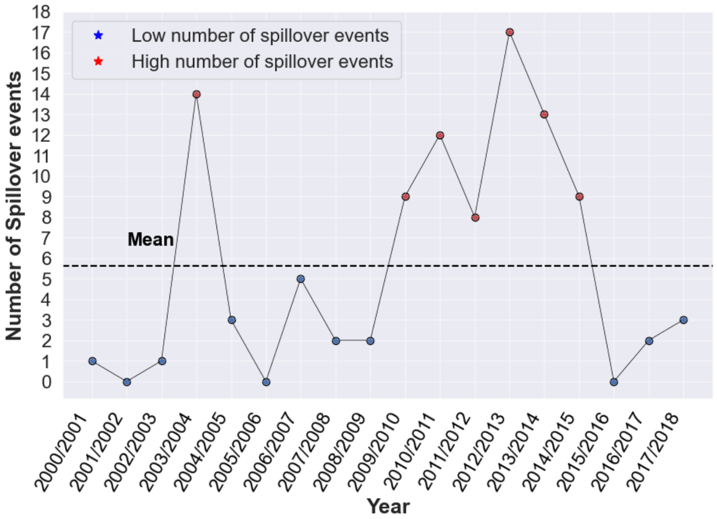


Figure 2.8: Yearly distribution of spillover events (N.tot = 101), with years above the mean (dashed line) marked in red and below the mean in blue.

Statistic	Value
Minimum	0
First quartile (Q1)	1.25
33rd percentile	2.00
Median	3.00
Mean	5.611
66th percentile	8.22
Third quartile (Q3)	9.00
Maximum	17
Inter Quartile Range (IQR)	7.75
Standard deviation	5.35

Table 2.4: Descriptive statistics of annual spillover events (year from May to April).

available data, the mean value was chosen as the reference cut-off between the two groups of years (Figure 2.8) and defined as follows:

- **High spillover years:** Seven years, primarily clustered between 2009/2010 and 2014/2015, along with 2003/2004, each reporting more than 8 annual spillover events.
- **Low spillover years:** The remaining eleven years, each recording fewer than 5 annual spillover events.

To ensure better data management and consistency in the analysis, the years will be then labeled sequentially: 2000/2001 will be referred to as Year 0, 2001/2002 as Year 1, and so forth up to 2017/2018, which will be designed as Year 17.

Statistical inference Finally, to identify the distribution that best describes the annual spillovers, data were fitted to both Poisson and Negative Binomial distributions. The detailed tools and methods employed are reported in Section A.2.3 in Appendix A. The goodness of fit was evaluated using various statistical methods, including Chi-square tests and the Kolmogorov-Smirnov (KS) test. Additionally, a QQ plot was generated to visually assess the fit.

Results indicated that the Negative binomial distribution well explains the annual data (Figure 2.9a). The dispersion parameter α is found to be statistically significant (p-value < 0.01), indicating that the mean and variance of the underlying distribution are not

equal.
This was further supported by a *Kolmogorov-Smirnov* test which yielded a KS = 0.1602 and a p-value = 0.6866, therefore the null hypothesis that the data follow a Negative Binomial distribution cannot be rejected) and QQ-plot (Figure 2.9b) indicates that the observed data closely follow the theoretical Negative Binomial distribution.

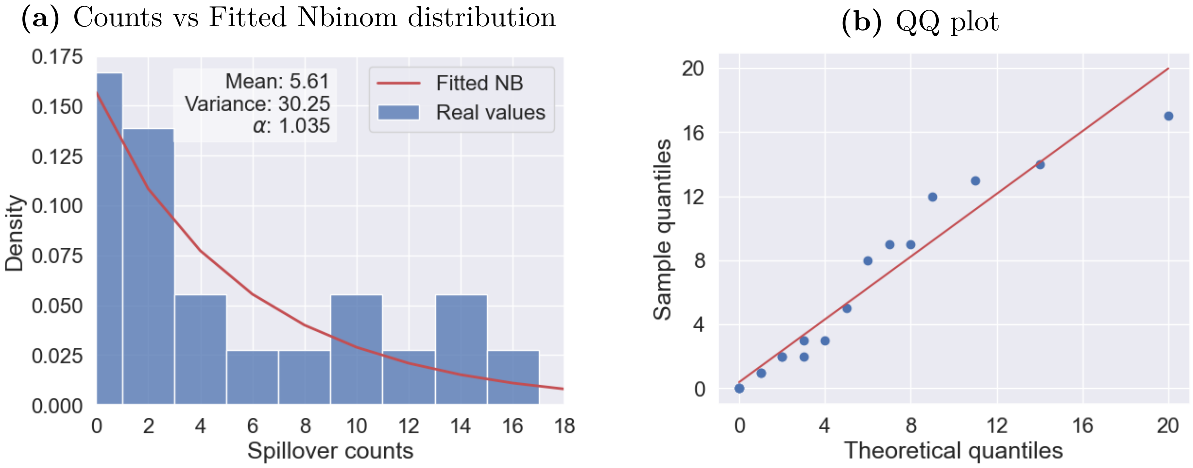


Figure 2.9: (a) Histogram of annual (May to April) spillover events with the fitted Negative Binomial distribution (red line). Total spillover counts across affected districts (2001-2018) are compared to the fitted Negative Binomial model. Key statistics, including mean, variance, and dispersion parameter α, are reported. (b) Quantile-Quantile plot (QQ-plot) as a graphical method for determining if the dataset follows a negative binomial probability distribution.

Chapter 3

Climate analysis

3.1 Analysis of climate anomalies

Empirical evidence suggests that most zoonotic and vector-borne diseases are sensitive to climate, particularly to variations of temperature and precipitations (refer to Section 1.6.2) [17]. Many of these diseases may also be influenced by climate anomalies and extreme events (see Section 1.6.2) [17]. As already discussed, climatic patterns exert numerous cascading effects on ecosystems ecology, agricultural productivity and reservoir hosts behaviour, by operating at different temporal and spatial scales [95]. These effects, in turn, can trigger various mechanisms that alter host, vector and pathogen dynamics, facilitating the onset of spillover events. Consequently, gaining a deeper understanding of the influence of climate on disease transmission dynamics is essential.

This chapter explores the relationship between temperature, precipitation, and the frequency of spillover events in Bangladesh during years with varying spillover counts (i.e., High and Low spillover years as defined in Section 2.3). The primary objective is to determine whether specific climatic patterns can explain the increase in spillover events during high-spillover years compared to low-spillover years. Two main research questions guide this analysis:

- **Seasonal spillover patterns:** The majority of annual spillover events occur during the winter months, with tails extending into March and April, and occasional occurrences in October and November. This observation prompts the question: *Do years with a high number of spillover events exhibit specific winter anomalies in temperature and/or precipitation that could explain differences compared to years with low spillover events?*

- **Preceding climatic conditions:** A related key question is: *Can anomalies outside the spillover period influence the number of events with a certain time lag?* This investigation is based on the hypothesis that pre-winter climatic conditions may affect food availability and bat behavior, leading to delayed impacts on their health status and ability to control infections (for example, changes in the quality and intensity of the immune response to infections, refer to Section 1.4.4). Previous studies on the relationship between climate and spillover events in Bangladesh have predominantly focused on the winter months, when these events are typically detected [33], [59]. However, virus shedding by bats and the collection and consumption of date palm sap – the primary transmission route of NiV in Bangladesh - occur all year-round.

This thesis aims to expand the analysis by examining a broader temporal frame, and exploring how climatic conditions affect spillover frequency beyond the winter period.

To address these questions, preliminary analysis identified the optimal spatial and temporal resolution for interpreting the results and explaining the phenomenon as follow:

- **Spatial Resolution:** the ERA5 climate data were aggregated by district to match the resolution of the spillover data. This adjustment was necessary to ensures a coherent framework, as the location of most spillover events is only known at the district-level.
- **Temporal Resolution:** non-overlapping seasonal windows were selected to align with Bangladesh's climatic seasons, namely, monsoon (June-September, JJAS), post-monsoon (October-November, ON), winter (December-February, DJF), and pre-monsoon (March-May, MAM). These seasonal windows provided clear initial trends, which were further refined using monthly data. The monthly resolution offered a more detailed understanding of climatic patterns within each season, allowing the identification of critical months that were particularly significant for distinguishing between high- and low-spillover years across the 18-year study period.

3.1.1 Climate data

The ERA5 reanalysis dataset, the fifth generation of global climate reanalyses from the European Centre for Medium-Range Weather Forecasts (ECMWF), serves as the source of observational data for this study. The key technical details of these reanalysis datasets are summarised in Table 3.1.

Reanalysis dataset characteristics	**ERA5**
Data type	Gridded
Horizontal coverage	Global
Horizontal resolution (atmosphere)	$0.25° \times 0.25°$
Temporal coverage	1979 to present
Temporal resolution	Hourly

Table 3.1: Main technical details of the ERA5 reanalysis datasets used in this study, produced by the Copernicus Climate Change Service (C3S) [96].

Reanalysis combines model data with observations from around the world into a globally complete and consistent dataset [97]. This process, known as data assimilation, integrates observations from various sources-including ground-based measurements, satellite observations and other remote sensing data-with a numerical weather prediction (NWP) model to produce a best estimate of the atmospheric state, known as reanalysis when applied to past data. This results in a comprehensive, high-quality, and high-resolution dataset that spans several decades.

One key advantage of reanalysis is that its outputs are not directly dependent on the density of ground-based observational networks [98]. This allows a reconstruction of climate variables in areas with sparse or no surface observations, making it particularly valuable for regions with limited observational coverage [99], [98].

In contrast, the WorldClim dataset, which has been extensively used in studies related to the Nipah virus, provides high-resolution GIS climatic layers derived from spatially interpolated climate surfaces using weather station data [99]. WorldClim data includes 19 bioclimatic variables derived from precipitation and temperature records for the period 1950 to 2000 [100].

However, WorldClim and weather stations, while valuable data sources, have several limitations in tropical regions, primarily associated with the reliance on spatial interpolation from sparse ground weather station data [101], exacerbated by topographic heterogeneity, inter-annual variability, and the distance to the closest weather station, which can compromise the accuracy of climate-related analyses [101]. To address these limitations, this study opts for ERA5 reanalysis data, which provides better coverage and higher spatial and temporal resolution across several decades.

3.1.2 Data pre-processing steps

The ERA5 reanalysis datasets were freely downloaded from the Copernicus Climate Change Service platform through the Climate Data Store web interface v.1.0, in Network common data form (NetCDF) format for the entire Bangladesh domain. The hourly variables of interest were:

- Air temperature at 2 meter above the surface of land ('2m_temperature', t2m, [K]);
- 3-hours precipitation ('precipitation', tp, [m]).

To focus on the specific region of interest (latitudes 20° N to 30°N and longitudes 85°E to 100°E) and the relevant time period (January 1, 2000, to December 31, 2019), a precise spatial and temporal subset was extracted using the Climate Data Operators (CDO)[1]. Both precipitation and temperature data were then aggregated at a daily time step. Specifically, from the original temperature data, daily averages ($T_{avg,day}$), daily minimums ($T_{min,day}$), daily maximums ($T_{max,day}$), and daily temperature ranges (DTR_{day}) (i.e. maximum-minimum) were derived. For precipitation, the daily cumulative precipitation was calculated ($Pr_{cum,day}$).

After temporal aggregation, the daily climatic data were further processed at the district level to align with the intended spatial resolution of the analysis. This spatial aggregation process was efficiently carried out using the advanced capabilities of the `xagg` package [2], which facilitates the computation of a weighted spatial average of gridded data values within each administrative unit.

---

[1] **CDO** is a collection of command-line Operators developed by the Max Planck Institute for Meteorology to manipulate and analyse climate and NWP model data. For more information, see https://code.mpimet.mpg.de/projects/cdo/

[2] `xagg` is a package to aggregate gridded data in `xarray` to polygons in `geopandas` using area-weighting from the relative area overlaps between pixels and polygons. For more information, see https://pypi.org/project/xagg/

3.1.3 Methods

In this study, climate anomalies are employed to distinguish years with high and low occurrences of spillover events at the district level. Climate anomalies provide a standardized frame of reference that enables more meaningful comparisons between regions, facilitating the identification of patterns and trends that might not be apparent with absolute values alone.

Anomalies are calculated by subtracting climatological values (i.e. long-term averages typically calculated over a reference period of 20-30 years [102]) from observed data, indicating whether a variable is above (positive anomaly) or below (negative anomaly) this baseline. Standardized anomalies, also referred to as normalized anomalies, were used, and calculated by dividing anomalies by the climatological standard deviation. This standardization removes the influence of location-specific factors and the effects of signal dispersion (such as the difference in seasonal ranges), allowing for more meaningful comparisons across different regions and time periods [102]. As a result, standardized anomalies provide clearer insights into the magnitude of climate variations, making it easier to distinguish normal conditions from unusual ones and to assess their potential impact on spillover events. The units are standard deviations (σ) from the mean of a normal distribution [103].

Differences between normalised climate anomalies: High vs. Low spillover years

This section addresses the calculation of climate anomalies. For each climate variable, the dataset consists of daily climate data aggregated at the district level from January 1, 2000, to December 31, 2018. To focus on conditions preceding spillover events, the year was redefined to span from May 1 to April 30, and relabeled from 0 to 17 for easier data management. Monthly resampling produced 216 mean values per district, with seasonal averages (18 values per district) calculated to capture overall conditions. After identifying seasons with significant anomalies, a month-by-month analysis was conducted to pinpoint specific months within these seasons that may have had notable impacts on the surge of spillover events.

Before calculating climate anomalies, a trend analysis was performed using two different approaches for temperature and precipitation data to uncover underlying patterns.

Trend analysis For temperature data, which satisfy the normality assumption, a *linear least-squares regression method* was used to detect trends (i.e., slopes) in seasonal and monthly temperature data at the district level. The trend slope was then removed from the original time series, regardless of its statistical significance, which was however assessed using a *two-tailed Student's t-test* at a 5% significance level, to ensure that the anomalies reflect variations around a stationary mean.

For precipitation data, which inherently exhibits greater variability and cannot be expected to follow a normal distribution, a more robust approach was needed to assess monotonic changes at the district level. The *Theil-Sen estimator* (also known as *Sen's slope estimator*) was used to provide an estimate of the trend, combined with the *non-parametric Mann-Kendall test* to evaluate the significance of the trend at the 5% level. For further details on these non-parametric methods, which are less sensitive to outliers and thus well-suited for analyzing hydro-climatic data, refer to Section A.3.1 in Appendix A.

The de-trending process was applied to each climate variable X from the set $\{T_{avg}, T_{min}, T_{max}, T_{range}, Pr\}$ and for each district d within the set D, representing the 64 districts of Bangladesh ($D = \{d \mid d \in \{0, 1, 2, \ldots, 63\}\}$). The process is defined by the formula:

$$\hat{X}_d(y) = X_d(y) - \hat{\beta}_d \cdot y \tag{3.1}$$

Here, y denotes a year within the research period $Y = \{y \mid y \in \{0, 1, 2, \ldots, 17\}\}$, covering the 18 years of study. The variable $\hat{X}_d$ represents the de-trended series, obtained by subtracting the estimated trend slope $\hat{\beta}_d$ from the original series for variable X. The original series is represented by $\{X\}_{y=0}^{N-1} = \{x_0, x_1, \ldots x_{N-1}\}$ where $N = 18$, indicating the number of years.

Normalized anomalies In this analysis, climatology is defined over the available 18-year period Y (from May 2000 to April 2018), rather than a **WMO** recommended 20- or 30-year reference period. This shorter time frame aligns with the spillover event data, enabling a more targeted examination of climatic conditions linked to spillover frequencies. The climatology for each month m and season s is calculated as the average over this period, providing a tailored baseline for analyzing the impact of climate on spillover events. Specifically, the normal of a season s or a month m, for each district d, is given by the following formulas:

$$\bar{X}_d(s) = \frac{1}{N}\sum_{y=0}^{17}\hat{X}_d(s) \quad s = 1, ..4\,; \quad \bar{X}(m) = \frac{1}{N}\sum_{y=0}^{17}\hat{X}_d(m) \quad m = 1, ..12 \tag{3.2}$$

where N is the number of years in the reference period (i.e. 18), and $\{\hat{X}_d\}$ represents the de-trended series of the climatic variable X under analysis.
Subsequently, for each district d in D, the seasonal or monthly anomaly of the year y are computed by subtracting the climatological baseline from the actual observed values, following the detrending process:

$$X'_d(s, y) = \hat{X}_d(s, y) - \bar{X}_d(s, y)\,; \quad X'_d(m, y) = \hat{X}_d(m, y) - \bar{X}_d(m, y) \tag{3.3}$$

Then, the annual normalised anomalies $\tilde{X}'_d(s, y)$ (or $\tilde{X}'_d(m, y)$) are calculated dividing the annual anomalies by the climatological standard deviation:

$$\tilde{X}'_d(s, y) = \frac{X'_d(s, y)}{\sigma_d(s)}\,; \quad \tilde{X}'_d(m, y) = \frac{X'_d(m, y)}{\sigma_d(m)} \tag{3.4}$$

where:

$$\sigma_d(s) = \sqrt{\frac{1}{N-1}\sum_{y=0}^{17}(X_d(s) - \bar{X}_d(s))^2}\,; \quad \sigma_d(m) = \sqrt{\frac{1}{N-1}\sum_{y=0}^{17}(X_d(m) - \bar{X}_d(m))^2}$$

Finally, the annual standardised anomalies are averaged separately within the two categorically defined groups of years: those with high spillover event frequencies (i.e. a 7-year period) and those with low frequencies (i.e. an 11-year period). The following formulas are used for this analysis:

$$\forall d \in D: \quad \tilde{X}'_d(s, Y_H) = \frac{1}{n_H}\sum_{y \in Y_H}\tilde{X}'_d(s, y) \quad ; \quad \tilde{X}'_d(m, Y_H) = \frac{1}{n_H}\sum_{y \in Y_H}\tilde{X}'_d(m, y) \tag{3.5}$$

$$\forall d \in D : \quad \tilde{X}'_d(s, Y_L) = \frac{1}{n_L} \sum_{y \in Y_L} \tilde{X}'_d(s, y) \quad ; \quad \tilde{X}'_d(m, Y_L) = \frac{1}{n_L} \sum_{y \in Y_L} \tilde{X}'_d(m, y) \tag{3.6}$$

where:

- $Y_H = \{y \mid y \in \{3, 9, 10, 11, 12, 13, 14\}\}$ represents the 7-years subgroup of to the reference period Y that includes years characterised by a high number of annual spillover events;
- n_H is the number of years y in Y_H;
- $Y_L = \{y \mid y \in \{0, 1, 2, 4, 5, 6, 7, 8, 15, 16, 17\}\}$ represents the 11-years subgroup within the reference period Y that encompass the years characterised by a low number of annual spillover events;
- n_L is the number of years y in Y_L;

As a final step, the difference between the standardised anomalies of these two groups is calculated as follows:

$$\forall d \in D : \quad \Delta\tilde{X}'_{d,obs}(s) = \tilde{X}'_d(s, Y_H) - \tilde{X}'_d(s, Y_L) \tag{3.7}$$

The goal of this comparative analysis is to identify and highlight the distinct climatic conditions that characterize years with high spillover events compared to those with low spillover events, evaluated on a district-by-district basis. The final step is to quantify these climatic differences to gain a deeper understanding of the factors that may contribute to variations in spillover occurrences between the two groups of years and across districts.

Temporal shuffling to assess statistical significance of results

After calculating the observed differences in standardized anomalies between high and low spillover years (i.e. $\forall d \in D : \Delta\tilde{X}'_{d,obs}(s)$ or $\Delta\tilde{X}'_{d,obs}(m)$), a *temporal shuffling method* (i.e., a random permutation without replacement) was employed to assess whether these differences were statistically significant or due to random variations.

Permutation test: random resampling without replacement The analysis covered 18 years (11 high and 7 low spillover years) and involved the following steps for each variable X and district d:

1. **Step 1: Random permutation (or shuffling) without replacement** Randomly assign 11 years to a synthetic "low" group ($Y^{(i)}_{L,synt}$) and 7 to a "high" group ($Y^{(i)}_{H,synt}$), repeating this process 1000 times to create different synthetic groupings, ensuring that each year y was assigned to only one group per iteration i.
2. **Step 2: Normalised anomaly differences** For each iteration, calculate the standardized anomaly differences between synthetic high and low groups across all districts. This produces a distribution of anomaly differences for comparison:

$$\forall i, d : \quad \tilde{X}'_d(s, Y^{(i)}_{H,synt})^{(i)} = \frac{1}{n_H} \sum_{y \in Y^{(i)}_{H,synt}} \tilde{X}'_d(s, y) \tag{3.8}$$

$$\tilde{X}'_d(s, Y^{(i)}_{L,synt})^{(i)} = \frac{1}{n_L} \sum_{y \in Y^{(i)}_{L,synt}} \tilde{X}'_d(s, y) \tag{3.9}$$

$$\Delta\tilde{X}'_{d,synt}(s)^{(i)} = \tilde{X}'_d(s, Y^{(i)}_{H,synt})^{(i)} - \tilde{X}'_d(s, Y^{(i)}_{L,synt})^{(i)} \tag{3.10}$$

The same method applies to monthly anomalies, replacing the seasonal indicator s with the monthly indicator m.

3. **Step 3: Statistical comparison:** For each district d, the observed difference in standardized anomalies between actual high and low spillover years were then compared against the permutation-based distribution. If these differences fell within the 5% or 10% tails of the distribution, they were considered statistically significant:

$$\Delta\tilde{X}'_{d,obs}(s) \text{ significant at 5\%} \iff \Delta\tilde{X}'_{d,obs}(s) < P_{2.5} \text{ or } \Delta\tilde{X}'_{d,obs}(s) > P_{97.5} \tag{3.11}$$

$$\Delta\tilde{X}'_{d,obs}(s) \text{ significant at 10\%} \iff \Delta\tilde{X}'_{d,obs}(s) < P_{5} \text{ or } \Delta\tilde{X}'_{d,obs}(s) > P_{95} \tag{3.12}$$

where $P_{2.5}, P_5, P_{95}, P_{97.5}$ represent the distribution percentiles. A result within these extreme tails, as exemplified in Figure 3.1, indicates that the observed difference is unlikely due to random chance, suggesting significant climatic distinctions between high and low spillover years.

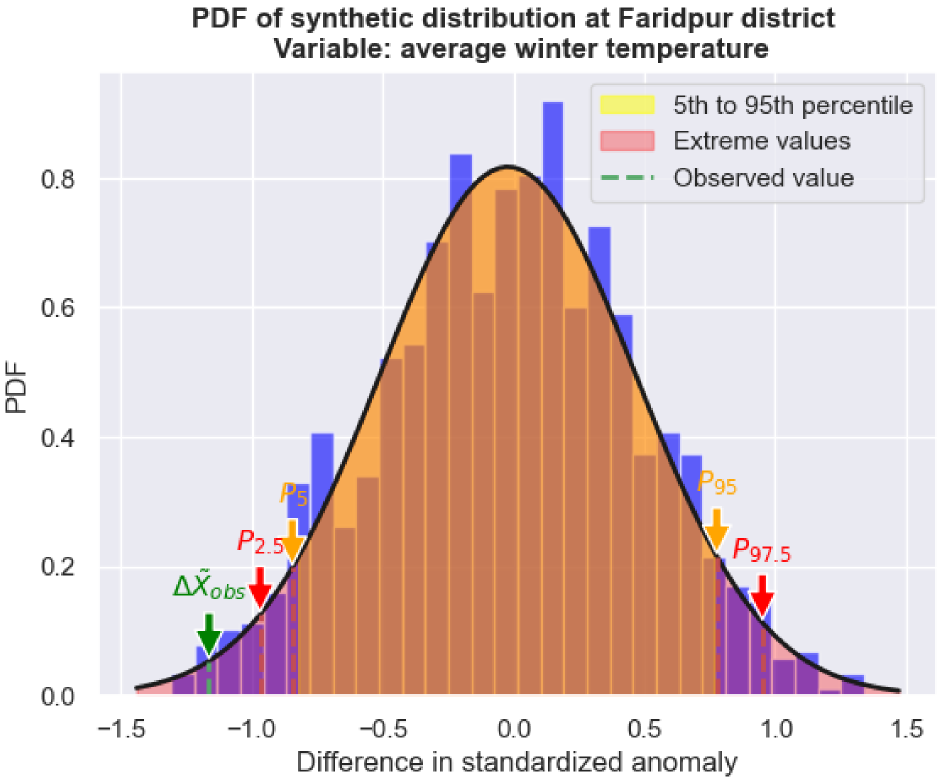


Figure 3.1: Probability distribution of synthetic differences in standardized anomalies for average winter temperature in Faridpur District. The extreme tails are defined by the percentiles $P_{2.5}, P_5, P_{95}, P_{97.5}$. The observed value, marked with a green dashed line, exceeds the lower $P_{2.5}$ threshold, indicating it falls within the extreme tails of the distribution. This placement confirms the anomaly is statistically significant at the 5% level, demonstrating a significant deviation from the expected climate conditions.

This permutation test, applied to each district d and variable X, randomizes observations to break any inherent dataset structure, thereby quantifying patterns that might emerge purely by chance [104], [105]. Unlike bootstrap methods, which resample with replacement to estimate the distribution of a statistic, the permutation test disrupts correlations among data and rigorously tests the null hypothesis that observed differences are due to random variation alone [104], [105].

3.1.4 Results

Climate trends

The evaluation of seasonal climatic trends identified statistically significant changes during the monsoon (June-September, JJAS) and winter (December-February, DJF) seasons. These results are consistent with long-term observations reported in the literature from 1966-2015, which also highlight a prevalent warming trend during the monsoon and a cooling trend during winter [89].

Specifically, the significant increase in average temperatures during the monsoon season across all districts in this analysis (p-value < 0.001) mirrors the 0.75 $°C$ long-term warming trend observed over the 50 years reported in the literature [89]. Notably, the most substantial increases in July and August align with sustained warming observed from May to October in this study, affecting over 20% of districts.

For winter, the observed cooling trend matches the reported 0.1 $°C$ decrease between 1966-2015 [89]. The narrowing temperature range (TR) across over half of the districts (p-value < 0.01) further supports this cooling pattern, likely driven by changes in daytime temperatures.

Additionally, the literature reports a warming trend during the pre-monsoon (MAM) period, particularly in maximum and minimum temperatures in coastal areas, linked to rising sea surface temperatures (SST) [89]. In contrast, no significant trends were detected for the overall MAM period in this study, though a significant increase was observed in May for average and minimum temperatures, possibly reflecting regional or short-term fluctuations.

Regarding precipitation, historical data shows increasing trends in pre-monsoon, monsoon, and post-monsoon seasons, with the largest rise (3.3 mm/year) during the monsoon, particularly in July and September [89]. This is attributed to warmer land temperatures and a deepened low-pressure system [89]. However, monsoon precipitation trends fluctuate over decades [89], and no significant long-term trends were detected in this study (2001-2018), reflecting these variations.

For winter, no significant long-term precipitation trends were found, consistent with the literature, which also reports no overall trend, though decreasing trends were noted in individual DJF months [89]. These discrepancies may result from the shorter time frame of this study, which may not capture broader variability.

Such changes in seasonal temperature and rainfall pattern might have a number of implications for water resources, agriculture (i.e., irrigation demand, crop phenology) and public health (more extreme weather events and related disasters; EIDs spread).

Normalised anomaly differences and statistical significance

The analysis of differences in standardized anomalies between years with high and low spillover events provides critical insights into the climatic conditions associated with varying spillover frequencies. Both seasonal and monthly evaluations revealed important patterns in temperature and precipitation anomalies. Maps illustrating these anomalies across districts were generated, with statistically significant differences at the 5% and 10% level highlighted using red and blue color-coded dots, respectively.

Temperature Anomalies During the monsoon season (June-September, JJAS), high spillover years showed positive average temperature anomalies in 25% of districts, half of which reported spillover events (see Figure 3.2b). Notably, July stands out as the only month with statistically significant positive anomalies across all temperature variables (average, minimum, maximum and temperature range) in more than 25% of districts, reflecting a substantial increase in heat during both daytime and nighttime.

In the winter season (December-February, DJF), significant negative anomalies were observed in 94% of districts for average temperature and in 75% of districts for minimum temperature, indicating colder nights during high spillover years. December, in particular, showed marked decreases in average, minimum, and maximum temperatures, while February exhibited an increased diurnal temperature range, indicating greater variability between daytime and nighttime temperatures.

No significant anomalies were detected at the seasonal level during the pre-monsoon period (March-May, MAM). However, the monthly analysis identified March as a key month, with significant increases in temperature range in over 28% of districts, suggesting notable climatic deviations during high spillover years.

Precipitation anomalies In terms of precipitation, high spillover years during the monsoon season were associated with drier conditions, with 31% of districts showing negative anomalies, particularly in the southern region (refer to Figure 3.3a). The most significant precipitation deficits were recorded in July, with 84% of districts experiencing reduced rainfall.

In the post-monsoon season (October-November, ON), negative precipitation anomalies were observed in 27% of districts, with November emerging as a critical month, confirming a drier post-monsoon period (see Figure 3.3b).

No significant anomalies were detected during the winter and pre-monsoon seasons overall. However, March showed a notable reduction in precipitation in over 50% of districts, during high spillover years.

3.1.5 Discussion

The results of this analysis, detailed in Table 3.2, clearly indicate that, over the 18-year study period, years with high spillover events are consistently characterized by warmer and drier monsoon periods—most notably in July, where the most pronounced temperature and precipitation anomalies were observed—followed by drier post-monsoon season (especially in November), and colder winters (with significant temperature deviations in December and February). In the pre-monsoon period, March exhibits notable climatic shifts, including an increased temperature range and reduced rainfall. These distinct climatic patterns appear to be linked to the frequency of spillover events in Bangladesh, suggesting potential environmental triggers that could heighten the risk of spillovers. These findings are further supported by several key factors, detailed in the following considerations, which provide additional context and potential explanations for the observed links between climate and spillover dynamics.

Colder winter and increased spillover activity: The analysis suggests a correlation between colder winters (lower average and minimum temperature) and heightened spillover activity, as shown in Figure 3.4. A plausible explanation for this link is the extended survival of the NiV virus in the environment at lower temperatures. Colder conditions can significantly prolong the virus's viability on contaminated fomites, such as date palm sap (DPS), increasing the risk of human exposure (see Section 1.6.3 for details). Additionally, DPS is consumed raw and within hours of collection during the winter months. As discussed in Section 1.5.3, the combination of these climatic factors and human behavior plays a crucial role in facilitating spillover events.

Flying foxes, which roost on exposed tree branches, are highly sensitive to extreme temperatures [40]. In response to cold, *Pteropus* bats activate thermoregulatory mechanisms, such as shivering, increased muscular activity, and wing flapping, to maintain their body temperature between 33 and 37.5 °C even when ambient temperatures drop to 17.5 °C [80]. This thermoregulation, especially during colder months in Bangladesh, adds physiological stress that could weaken their immune system, increasing the likelihood of viral recrudescence and shedding (refer to Section 1.6.3 for further details). The stress is further compounded by nutritional challenges, as winter is also when fruit availability is at its lowest in Bangladesh (see Section 1.6.3), leading to heightened vulnerability among bats [33], [56].

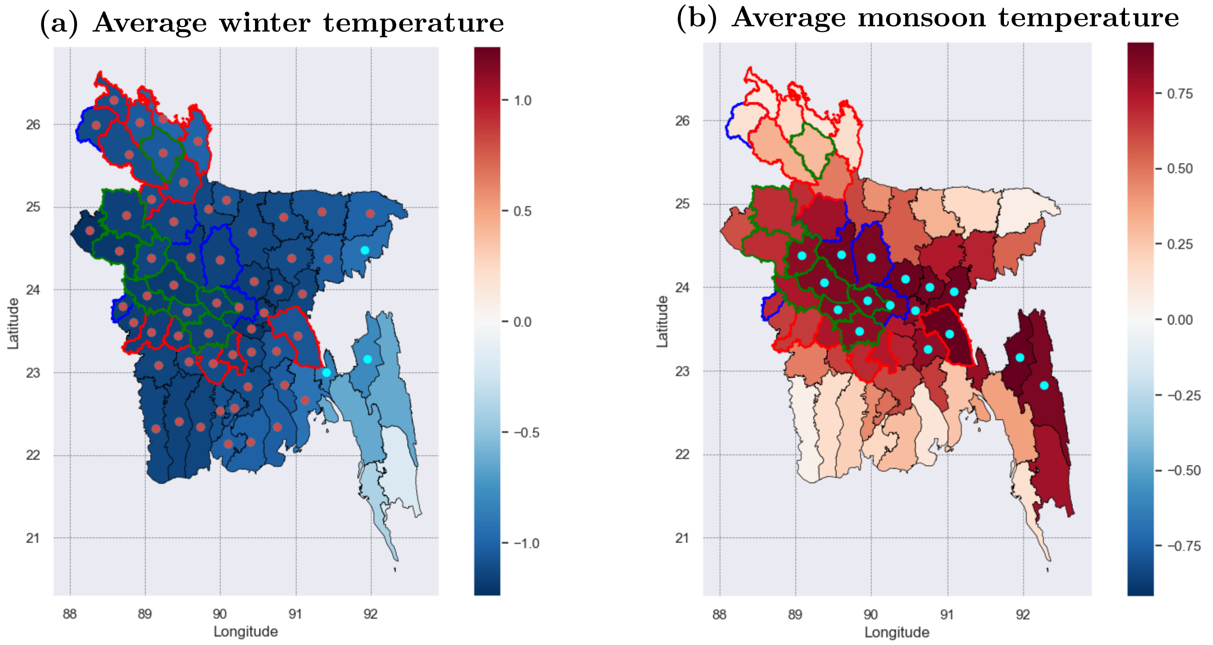
Moreover, colder nighttime temperatures (i.e. minimum temperatures) could increase sap production and leakage from cuts made by harvesters on date palms. As discussed in Sections 1.6.3 and A.2.1, the xylem refilling process in plants, like date palms, is influenced by lower nighttime temperatures, when the plant's evapotranspiration (ET) is minimal. This creates ideal conditions for sap flow and leakage, attracting wildlife, including bats, and increasing the chances of viral transmission. As anticipated in paragraph 1.5.3 and documented in [56], there is a significantly higher frequency of bat visits to date palm trees during winter compared to other seasons, especially on colder nights. Thus, the convergence of food scarcity, increased sap production and sugar content of DPS, the peak season of fresh sap consumption by humans, and cold-induced physiological stress in bats forms a set of ecological conditions that likely drive spillover events during colder winter months over the reference period.

Warmer and drier monsoon, drier post-monsoon and higher spillover activity in the subsequent winter: The analysis suggests a correlation between warmer and drier monsoon periods, drier post-monsoon conditions, and heightened spillover activity during the subsequent winter season (see Figure 3.4). Notably, the combination of higher temperatures and reduced rainfall pinpoints July as a critical month within the monsoon season, where the observed anomalies were most severe and statistically significant. While overall monsoon anomalies show some mismatches with expected spillover frequency patterns (refer to Figure 3.4), the anomalies observed in July more closely align with spillover frequency (refer to Figure 3.5), suggesting a stronger correlation during this key period. However, these preliminary observations remain speculative and will be further examined through a statistical regression analysis in Chapter 5.

Drier conditions and rising temperatures during these seasons may significantly contribute to environmental stressors that disrupt flowering and fruit production cycles, leading to pollen desiccation and reduced pollinator activity [106]. Shortened pollination windows and inadequate moisture availability can also result in premature fruit drop and reduced fruit set [106]. These climatic shifts can also advance flowering and fruiting times, creating mismatches with optimal growing conditions and reducing fruit yield and quality (i.e., size, color, and taste - all known attracting factors for bats) [62], [106]. Tropical fruits like mango, papaya, guava, litchis, and jackfruit, which are critical food sources for *Pteropus* bats, are particularly sensitive to such changes [62], [107].

The cumulative effect of these environmental stressors may disrupt the availability of food resources for frugivorous bats, potentially leading to nutritional stress that promotes spillover. Furthermore, monsoon and post-monsoon months coincides with the energetically demanding pregnancy phase for *Pteropus* bats in Bangladesh [48], increasing their vulnerability to nutritional stress (refer to 'Breeding season and immunosuppression' in paragraph 1.6.3). Additionally, this climate-driven food scarcity may drive bats to seek alternative food, altering their foraging behavior and increasing contact with human environments, thereby elevating the risk of zoonotic spillover events. These ecological stressors underscore the need for adaptation strategies to mitigate climate change impacts on both agricultural productivity and public health.

Interestingly, these conditions, seems to mirror observations of the Hendra virus in Australia, where acute food shortages, typically occurring in spring, are followed by increased viral shedding in the subsequent winter, with a delayed viral recrudescence in bats 6 to 9 months later [83], [82]. The exact mechanism driving this time-lag is unclear, but prolonged nutritional stress, thermoregulation challenges, pregnancy, and lactation likely contribute to delayed viral shedding [83]. This finding is particularly noteworthy because, to the best of the author's knowledge, a correlation between monsoon climate conditions (especially July month) and winter spillover events in Bangladesh has not been previously identified, offering new insights into how seasonal factors influence viral dynamics in bats. This hypothesis will be further examined in Chapter 5 through regression analysis using Generalised Linear Models (GLMs) to explore the potential relationship between monsoon climate conditions and spillover frequency.


(a) Average winter temperature

(b) Average monsoon temperature

Figure 3.2: Maps of observed differences in standardised average temperature anomalies between years with high and low spillover events during: (a) winter, and (b) monsoon seasons. **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both.

Variable	Monsoon				Post-Monsoon		Winter			Pre-Monsoon		
	J	**J**	**A**	**S**	**O**	**N**	**D**	**J**	**F**	**M**	**A**	**M**
T_{avg}	-	**Positive** [+0.9,+1.7] 89% (46%)	-	-	-	-	**Negative** [-0.9,-1.4] 92% (44%)	-	**Negative** [-0.8,-0.9] 52% (33%)	-	-	-
	Positive [+0.8,+0.9] 25% (50%)				**-**		**Negative** [-1,-1.2] 94% (100%)			**-**		
T_{min}	-	**Positive** 75% (46%)	-	-	-	-	**Negative** [-0.8,-1.2] 50% (50%)	-	**Negative** [-0.8,-1] 66% (29%)	-	-	-
	-				**-**		**Negative** [-0.9,-1.1] 75% (48%)			**-**		
T_{max}	-	**Positive** 67% (44%)	-	-	-	-	**Negative** [-0.9,-1.5] 66% (57%)	-	-	-	-	-
	-				**-**		**-**			**-**		
TR	-	**Positive** [+0.8,+1.2] 25% (6%)	-	-	-	-	-	-	**Positive** [+0.8-+1.2] 40% (15%)	**Positive** [+0.9,+1.1] 28% (44%)	-	-
	-				**-**		**-**			**-**		
Pr	-	**Negative** [-0.8,-1.6] 84% (32%)	-	-	-	**Negative** [-0.7,-1] 19% (0%)	-	-	-	**Negative** [-0.8,-1.1] 56% (47%)	-	-
	Negative [-0.9,-1.3] 31% (15%)				**Negative** [-0.9,-1.2] 27% (6%)		**-**			**-**		

Table 3.2: Results of the analysis of differences in standardized anomalies between years with high and low spillover events. "Positive" or "Negative" indicates that high spillover years are characterized by statistically significant positive or negative anomalies compared to low spillover years. The range of statistically significant values is shown in square brackets. The first percentage indicates the proportion of districts across Bangladesh reporting statistically significant values. The second percentage, in parentheses, indicates the proportion of districts affected by spillover events over the 18 years analyzed. An hyphen (-) indicates that none or fewer than 19% of districts reported statistically significant values. This 19% threshold was specifically chosen because all other percentages fall below 5%, thereby establishing a clear and significant boundary between relatively rare and more frequent occurrences of statistically significant reports.

3.2 Analysis of weather extremes

This section explores the selection and analysis of climate extreme indices in relation to the ecological drivers of the Nipah virus (NiV) in Bangladesh, focusing on two main aspects: the climate-driven xylem refilling phenomenon in date palms (detailed in paragraph A.2.1) and climate-induced food shortages, both of which are linked to increased human cases of NiV.

- **Xylem refilling and date palm sap production:** The xylem refilling phenomenon in date palms, which occurs during the dry and cold winter months, leads to increased sap production, closely tied to NiV transmission. This process is driven by specific climatic conditions, such as high evaporative demand on sunny, low-humidity days followed by cold, clear nights [56]. To capture this, the number of dry days followed by cold nights, referred to as "Dry and Cold" days (DCD), is calculated both annually (denoted as DCD_y) and for the winter season (denoted as DCD_{winter}). This approach aims to determine if this climate-driven phenomenon occurs at other times of the year beyond the documented winter period. The analysis also assesses the maximum consecutive number of these days ($CDCD$). The mean DCD for the year and winter is then compared between years with high and low spillover events to determine whether this phenomenon significantly contributes to the increased virus transmission observed in certain years and districts. Additionally, the study investigates whether the frequency on annual scale, or accumulation of these days over time influences

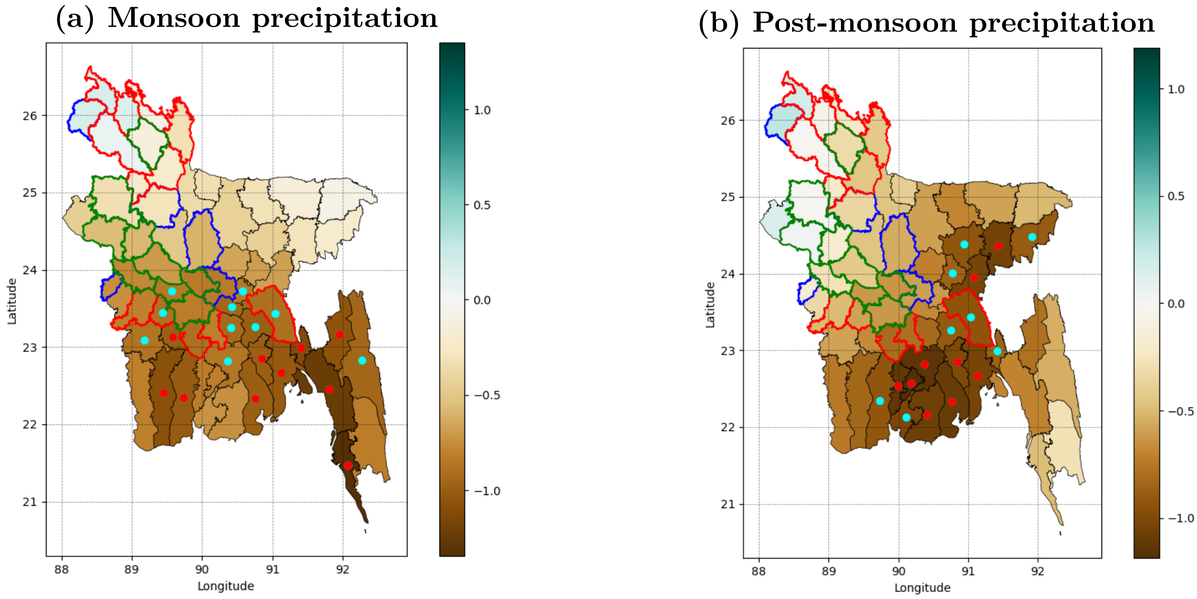


Figure 3.3: Maps of observed differences in standardised precipitation anomalies between years with high and low spillover events during: (a) monsoon and (b) post-monsoon seasons. **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both.

NiV transmission dynamics.

- **Climate-driven food-shortage:** Reduced precipitation in the six to nine months leading up to the birthing period of *Pteropus* bats can lead to food shortages, which, through nutritional stress, can weaken their immune systems and increases NiV recrudescence and shedding. The analysis quantifies the frequency of minimal precipitation levels (i.e. dry days) and the maximum duration of drought periods (i.e. consecutive dry days) annually and during the winter season to evaluate their direct and indirect impacts on virus transmission. This assessment addresses key questions such as the correlation between dry conditions and increased spillover events, identifying the districts most affected by prolonged droughts, and the role of these climatic patterns in NiV dynamics.

- **Literature findings:** Given the significant positive correlation identified between the percentage of winter days with average temperature below the 17 $°C$ and the frequency of spillover events [33], this analysis also includes the calculation of the total number of cold days annually (i.e. CD_y) and during the winter season (i.e. CD_{winter}). Additionally, it quantifies the maximum number of consecutive cold days within a year (i.e. CCD_y) and specifically during winter (i.e. CCD_{winter}). This approach seeks to clarify whether this climatic condition, when accumulated over time or on an annual basis, influences the transmission dynamics of the virus.

The selected climate indices of extreme events are detailed in Table 3.3.

Index name	**Index symbol**	**Definition**
Dry Days	DD	A day d is called ’Dry’ $\iff$ $Pr_{cum,d}(d) < 1\ \frac{mm}{day}$
Consecutive Dry Days	CDD	CDD index is largest number of consecutive “dry days” during the period of interest (year or season)
Cold Days	CD	A day d is called ’Cold’ $\iff$ $T_{min,d}(d) < 17\ °C$
Consecutive Cold Days	CCD	CCD index is largest number of consecutive “cold days” during the period of interest (year or season)
Dry days followed by Cold nights (i.e. Dry and Cold Days)	DCD	A day d is called ’Dry and Cold’ $\iff \begin{cases} Pr_{\mathrm{cum,d}}(d) < 1 \dfrac{\mathrm{mm}}{\mathrm{day}} \\ T_{\mathrm{min,d}}(d+1) < 17\,°\mathrm{C} \end{cases}$
Consecutive Dry and Cold Days	CDCD	CDCD index is largest number of consecutive “dry and cold days” during the period of interest (year or season)

Table 3.3: The selected indices of extreme events: name, symbol and definition.

3.2.1 Methods

Calculation of extreme indexes

The calculation of both dry and cold days followed similar methodologies.

Dry Days were calculated using daily cumulative precipitation data from May 1, 2000, to April 30, 2018. A dry day ($Pr_{\text{cum, d}}$) was defined as a 24-hour period with precipitation below 1 mm/day, using the dry day indicator d_i (where $d_i = 1$ for dry days, $d_i = 0$ otherwise). This threshold of 1 mm/day was chosen as it is a standard value below which precipitation is considered negligible.

Monthly dry day counts ($DD_d(m)$) were summed to obtain annual dry days ($DD_{y,d}$) for each district. Winter dry days ($DD_{winter,d}$) were calculated by summing values for December, January, and February (DJF).

Cold Days were similarly calculated using daily minimum temperature data. A cold day was defined as a 24-hour period where the minimum temperature fell below 17 °C, based on the cold day indicator c_i (where $c_i = 1$ for cold days, $c_i = 0$ otherwise). The choice of the 17°C threshold was based on McKee et al.’s sensitivity analysis [33], which found the strongest correlation between spillover events and the number of winter days with temperatures between 16-18 °C. Though statistical significance persisted across the 15-20 °C range, 17 °C was chosen as an upper bound to effectively capture cold days without producing zeros in the dataset.

Monthly cold day counts ($CD_d(m)$) were summed to compute annual cold days ($CD_{y,d}$) and winter dry days ($CD_{winter,d}$) for each district.

In both cases, to explore the relationship between these climatic conditions and Nipah virus spillover events, the annual averages of dry and cold days were calculated for two groups: years with high spillover frequency ($y \in Y_H$) and years with low spillover frequency ($y \in Y_L$). The difference between these averages ($\Delta DD_{obs,d}$ for dry days, $\Delta CD_{obs,d}$ for cold days) was then computed for each district, identifying distinctive climatic patterns linked to spillover events. Additionally, tracking the specific dates of dry and cold days allows for analysis of whether these conditions extend beyond the winter months. See Section A.3.2 for further details on the formulas used.

Dry days followed by cold nights To derive the "dry days followed by cold nights" index, a systematic approach was employed, analyzing the sequences of dry days (DD) and cold nights (CD) for each district d and year y. Since cold nights correspond to cold days (as minimum daily temperatures are typically recorded at night), the method identifies consecutive occurrences of dry and cold conditions by shifting dry day dates and matching them with cold night dates.

The analysis was performed annually and for the winter season (December to February), and the results were organized into matrices for all districts and years. Average counts of "dry days followed by cold nights" were calculated for years with high (Y_H) and low (Y_L) spillover events, allowing for a comparison between these periods. The differences between the two groups were computed for each district, highlighting specific climatic patterns potentially influencing spillover events.

By recording the dates when dry days and cold nights occur in sequence, this method precisely identifies overlapping periods and offers insights into whether these patterns extend beyond the winter season, deepening the understanding of their spatial and temporal distribution and their potential impact on spillover events.

Further details on the steps and formulas are provided in Section A.3.2 in Appendix A.

Consecutive Indexes: CDD, CCD, CDCD The analysis of consecutive dry days (CDD), consecutive cold days (CCD), and consecutive "Dry and Cold" days ($CDCD$) aims to identify, count, and record the longest annual sequences of these conditions across districts. This process utilizes daily precipitation and temperature data from May 1, 2000, to April 30, 2018, aggregated at the district level.

A systematic approach was employed, as detailed in Algorithm 1, where consecutive days that meet a specified condition (i.e. "Dry" or "Cold") are tracked by incrementing a counter. When the condition is no longer satisfied, the counter resets, and the longest sequence is updated if the current sequence exceeds the previous maximum. The corresponding dates for both the current and longest sequences are also recorded. For further details, refer to the helper function **`FindConsecutiveDays`** (Algorithm 2) within the **`ConsecutiveDaysDetection`** procedure (Algorithm 1) in Appendix A.

Next, dry day sequences are shifted by one day and matched against cold night sequences to identify consecutive "Dry and Cold" days, using the **`ShiftDates`** and **`FindMatches`** procedures (refer to algorithm 3 in the Appendix A). The number of such matched sequences is counted for each district and year, resulting in a comprehensive dataset of consecutive dry, cold, and "Dry and Cold" days. These results are organized into matrices representing districts and years (i.e.64×18), allowing for detailed analysis of climatic patterns. To assess the potential influence of these patterns on spillover events, the counts of consecutive dry, cold, and "Dry and Cold" days are averaged separately for years with high (Y_H) and low (Y_L) spillover frequency. The differences between these averages are then calculated for each district, providing insight into whether climatic extremes correlate with spillover events.

Additionally, by extracting the unique months associated with these dates, this methodology allows for a precise identification of the periods when these conditions occur, enabling a comprehensive assessment of environmental factors that may influence the transmission dynamics of the Nipah virus across various districts.

Temporal shuffling to assess statistical significance of results

After computing the observed differences in the number of dry days (ΔDD_{obs}), consecutive dry days (ΔCDD_{obs}), cold days (ΔCD_{obs}), consecutive cold days (ΔCCD_{obs}), dry days followed by cold nights (ΔDCD_{obs}) and "Consecutive Dry and Cold days" ($\Delta CDCD_{obs}$) between high and low spillover years across the research period, it is crucial to determine whether these differences are statistically significant or simply due to random variation. To assess this, the temporal shuffling method, as detailed in Section 3.1.3, is applied. The same permutation process described for anomalies is adapted for day count variables ($DD, CD, DCD, CDD, CCD, CDCD$). Rather than anomalies, the mean number of days is computed for each synthetic high and low group of years, and the difference between these means is calculated. The shuffling, iterations, and statistical comparison using percentile thresholds follow the same procedure, maintaining consistency with the approach used for anomalies.

This method provides a robust statistical framework for evaluating the significance of the observed differences, ensuring that the results are not merely artifacts of chance.

3.2.2 Results

The results in Table 3.4 show that three out of six indices — Annual Dry Days (DD_y), Annual Cold Days (CD_y), and Annual Dry days followed by Cold Days (DCD_y) — show significant differences between high and low spillover years. Specifically, years with more spillover events had notably higher counts of dry days, cold days, and dry days followed by cold nights.

Annual Dry Days (DD_y)	**Consecutive annual Dry Days** (CDD_y)	**Annual Cold Days** (CD_y)	**Consecutive annual Cold Days** (CCD_y)	**Annual Dry and Cold Days** (DCD_y)	**Consecutive annual Dry and Cold Days** ($CDCD_y$)
Higher [+8.9,+12] 47% (57%)	-	**Higher** [+9.7,+12.2] 25% (11%)	-	**Higher** [+7.06,+11.81] 41% (54%)	-

Table 3.4: Results of the analysis of differences in extreme events indexes (i.e. count day variables) between years with high and low spillover events. "Higher" or "Lower" indicates that high spillover years are characterized by statistically significant higher or lower number of days of the corresponding count variable compared to low spillover years. The range of statistically significant values (ΔN_{days}) is shown in square brackets. The first percentage indicates the proportion of districts across Bangladesh reporting statistically significant values. The second percentage, in parentheses, indicates the proportion of districts affected by spillover events over the 18 years analyzed. An hyphen (-) indicates that none or fewer than 19% of districts reported statistically significant values. This 19% threshold was specifically chosen because all other percentages fall below 6%, thereby establishing a clear and significant boundary between relatively rare and more frequent occurrences of statistically significant reports.

Statistical analysis reveals that 47% of districts show significant differences in the DD_y index, followed by 41% for DCD_y and 25% for CD_y. Notably, 65% of districts with significant differences in DD_y experienced at least one spillover event over the 18-year period, compared to 54% for DCD_y, and 44% for CD_y, highlighting a strong link between climatic extremes and spillover dynamics.

Notable result: The spatial distribution of these districts, shown in Figure 3.6, reveals that most of the significant differences cluster within the 'Nipah belt'—the area that experienced at least one spillover event at the district level during the 18-year period.

This clustering suggests that climate extremes, such as increased dry days, cold days, and sequences of dry days followed by cold nights, may have played a key role in triggering spillover events. Given the strong alignment between these climatic patterns and high-risk districts, the indices (DD_y, CD_y, DCD_y) will be explored further in Chapter 5.

Deeper temporal analysis of DD_y, CD_y and DCD_yindexes To gain deeper insights into the temporal patterns of extreme conditions, the analysis explored the monthly distribution of annual dry days, cold days, and dry days followed by cold days. While the indices are based on annual counts, examining monthly trends helped reveal whether these conditions extended beyond the winter months and how they differed between high and low spillover years.

The analysis involved the following steps:

1. **Extraction of unique months:** The unique months associated with these extreme events were identified to understand when they occurred and whether they spanned other seasons beyond winter.

2. **Quantification of monthly frequencies:** For each district and month, the total number of extreme days over the 18-year period was calculated using:

$$\forall d \in D, \forall m \in M : \quad DayCount\_tot_d(m) = \sum_y DayCount_d(m, y) \tag{3.13}$$

   The monthly totals were then averaged across all districts to capture regional trends:

$$\forall m \in M : \quad \begin{aligned} SumDayCount(m) &= \sum_{d=0}^{63} DayCount\_tot_d(m)\,; \\ AvgDayCount(m) &= \frac{SumDayCount(m)}{64} \end{aligned} \tag{3.14}$$

3. **Quantification of monthly differences:** For each district, monthly averages were computed separately for years with high (Y_H) and low (Y_L) spillover events.

$$\forall d \in D, \forall m \in M : \quad \begin{aligned} AvgDayCount_d(m, Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} DayCount_d(m, y)\,; \\ AvgDayCount_d(m, Y_L) &= \frac{1}{n_L} \sum_{y \in Y_L} DayCount_d(m, y) \end{aligned} \tag{3.15}$$

   Differences between high and low spillover years were then calculated, for each district $(d \in D)$ and month $(m \in M)$:

$$\Delta DayCount_d(m) = AvgDayCount_d(m, Y_H) - AvgDayCount_d(m, Y_L) \tag{3.16}$$

   A spatial average of these differences across all districts was computed to highlight broader trends:

$$\forall m \in M : \quad <\Delta DayCount_d(m)>= \frac{1}{64} \sum_{d=1}^{64} \Delta DayCount_d(m) \tag{3.17}$$

The results of this analysis revealed distinct monthly patterns for each index:

- **Annual dry days:** As illustrated in Figure 3.7a, dry days are predominantly concentrated between October and April, with frequencies exceeding those from May to September (including monsoon period, JJAS) by more than an order of magnitude. This concentration of dry days outside the monsoon period overlap significantly with the spillover periods. The results in Figure 3.7b show that years with high spillover events had more dry days in every month, particularly between October and April, with increases ranging from 12 additional dry days in October to 28 in December compared to low spillover years.

- **Annual cold days:** Cold days are mainly concentrated between November and March, as shown in Figure 3.7a. This pattern closely aligns with the monthly distribution of total spillover events over the 18-year period. The analysis of monthly differences (Figure 3.7b) further supports this finding, indicating that high spillover years experienced more cold days during these months than low spillover years, suggesting that cold days may play a key role in the increased incidence of spillover events during these critical months.

- **Annual dry days followed by cold nights:** The occurrence of dry days immediately followed by cold nights extends beyond the winter months, spanning from October to April as shown in Figure 3.7a, perfectly overlapping with the spillover period. This highlights a significant correlation between these climatic conditions and the increased risk of spillover events. The detailed examination in Figure 3.7b reveals a pattern similar to the Cold Days index but adds the critical insight that a cold night follows a dry day. This sequence may be linked to the xylem refilling phenomenon, which increases Date Palm Sap (DPS) production. As discussed earlier, this increase in DPS can attract bats during periods of food scarcity, such as winter, thereby raising the likelihood of spillover events.

3.2.3 Discussion

Overall and summarizing the insights from these findings and Figure 3.7b, it is evident that from October to April, high spillover years consistently exhibit a greater occurrence of all three extreme climate conditions, with notable peaks in January and December. This suggests a strong seasonal pattern where these extreme conditions not only coincide with but may also be involved to exacerbate spillover risk. Dry days show the most significant overall differences between high and low spillover years, particularly in January and December, while the combined impact of "dry and cold" days was more variable, especially during the early winter months. This variation suggests that the consecutive occurrence of these extreme weather conditions could significantly amplify spillover outbreaks. This may be tied to embolism repair processes in date palms, where xylem refilling, triggered by such climatic conditions (refer to Section A.2.1 for further details), increases sap flow attracting bats, particularly during times of food scarcity.

Another key observation from this analysis is the progressively higher number of dry days from July to December in high spillover years compared to low spillover years, which aligns with the fruit bats' gestation period in Bangladesh [48]. This temporal pattern could disrupt the phenological stages of key dietary plants by affecting their flowering and altering food availability during critical periods for bats. These disruptions can lead to nutritional stress that could weaken the bats' immune systems and increase the likelihood of viral recrudescence (refer to Section 1.4.4 for further details). Although this hypothesis requires further validation, it offers a plausible explanation for the observed findings and the evidence that it coincides with the critical pregnancy phase for bats. Furthermore, short, intense climatic events—such as sudden temperature drops or brief dry spells—may cluster temporally, altering bat behavior, such as increased foraging or movement patterns, ultimately leading to heightened contact with humans or domestic animals.

These findings underscore the importance of considering temporal clustering of climate extremes and threshold events in the study of spillover dynamics. Understanding which months are most critical for extreme conditions that correlate with spillover events allows for more strategic planning of interventions and a better allocation of resources to monitor and mitigate potential outbreaks.

The lack of significance in indices calculating the maximum consecutive number of dry or cold days associated with high vs. low spillover years suggests that it is not the prolonged duration of specific conditions but rather the occurrence of particular climatic events at critical times of the year that influences the number of spillover events. In other words, certain short-term climatic events may be more critical for facilitating Nipah virus transmission than sustained climatic conditions. This highlights the role of threshold events, where crossing a certain climatic threshold, even briefly, can trigger behavioral changes in bats or increase their interactions with humans, thereby facilitating virus spillover.

These findings seem to suggest that it is the timing and intensity of these threshold events, rather than their persistence, that play a pivotal role in the transmission dynamics of the Nipah virus.

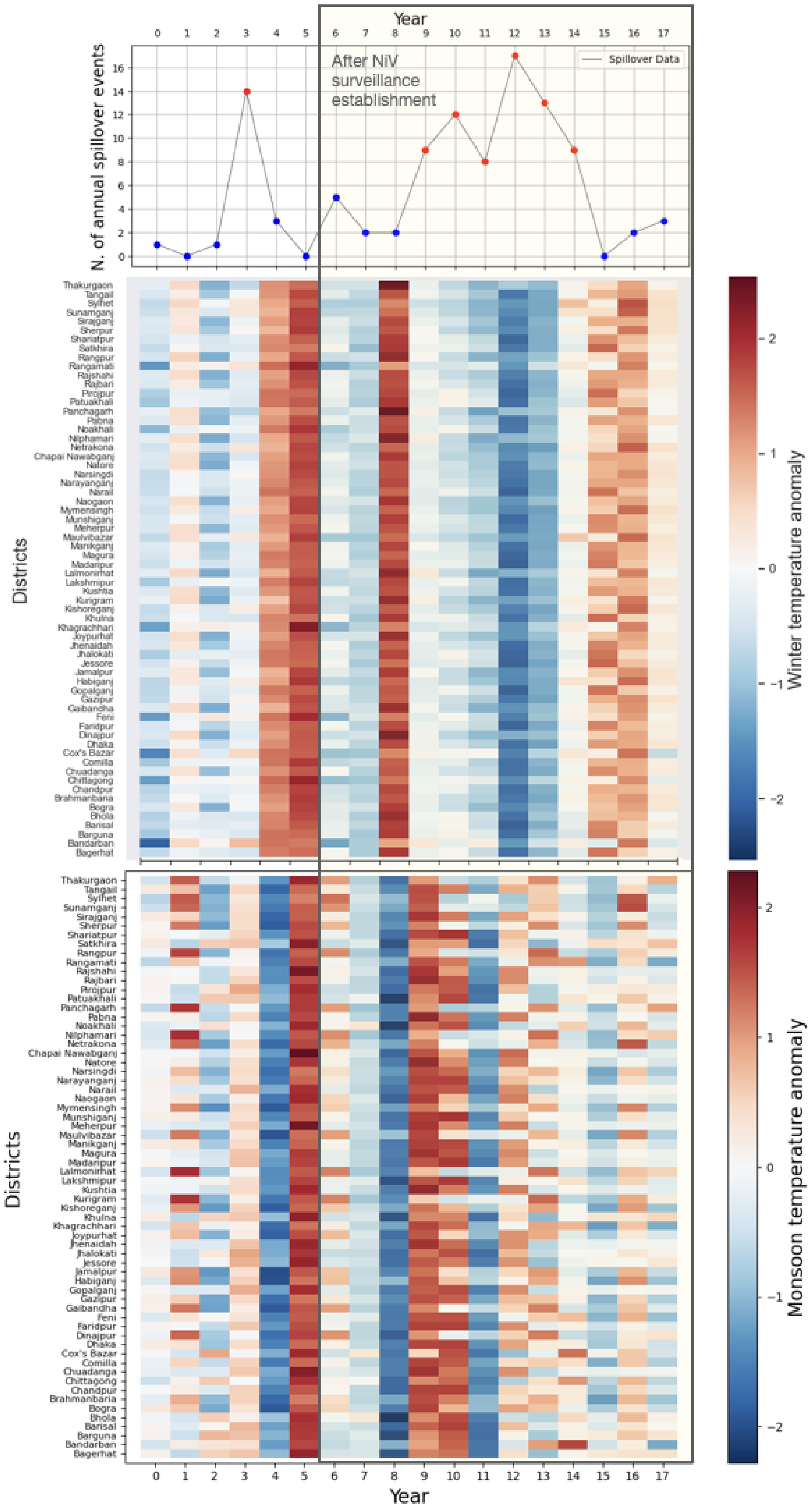


Figure 3.4: Anomalies in average temperatures during the winter and monsoon seasons across various districts of Bangladesh from year 0 to year 17 under analysis. The top panel plots the annual number of spillover events, distinguishing periods before and after the establishment of Nipah virus surveillance. Red dots signify years with high spillover events, while blue dots represent years with low spillover events. The heat maps below show the temperature deviations from the norm for each district during the winter and monsoon seasons, respectively. This visualization aims to explore potential correlations between temperature anomalies and the frequency of spillover events.

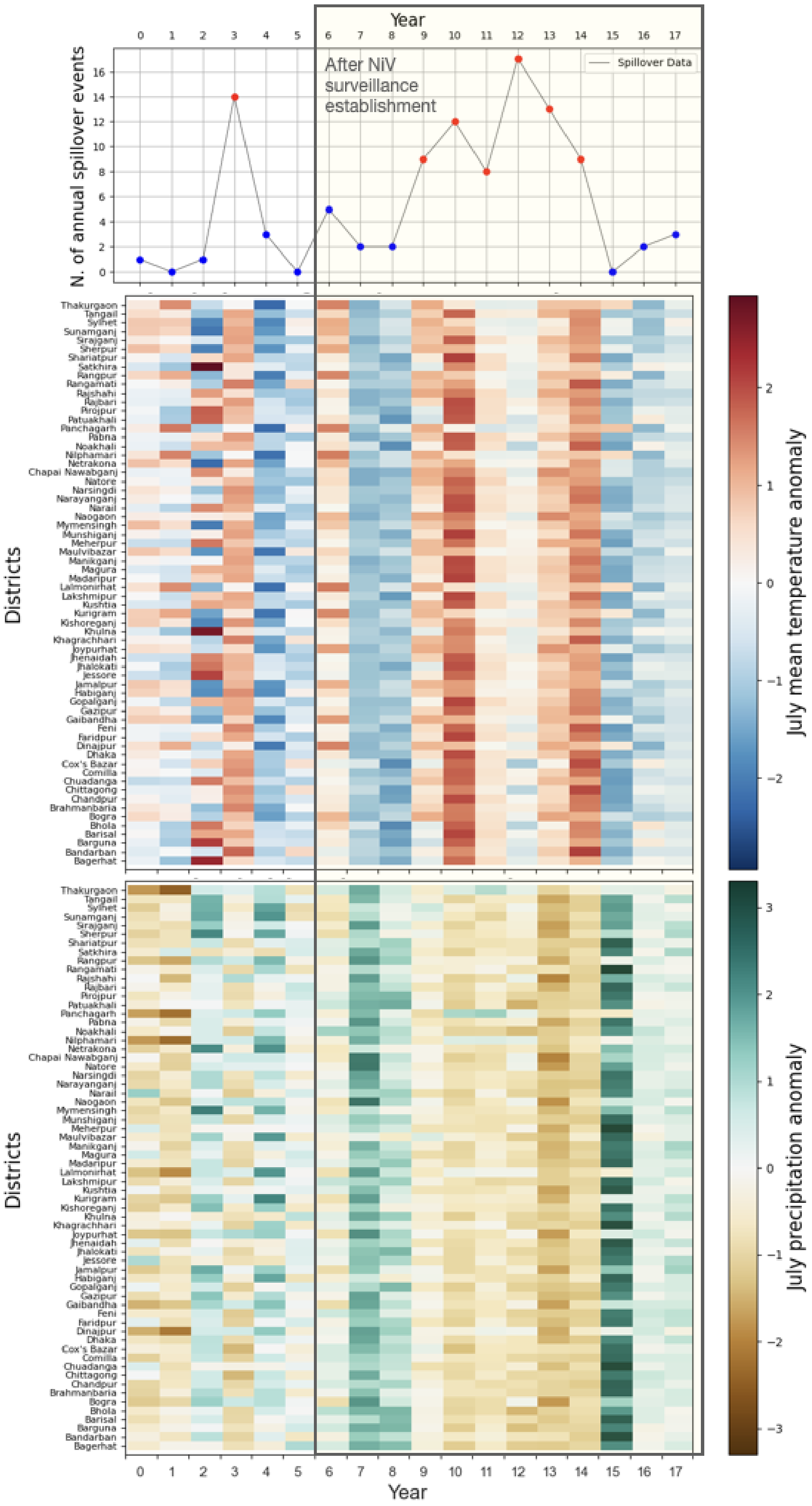
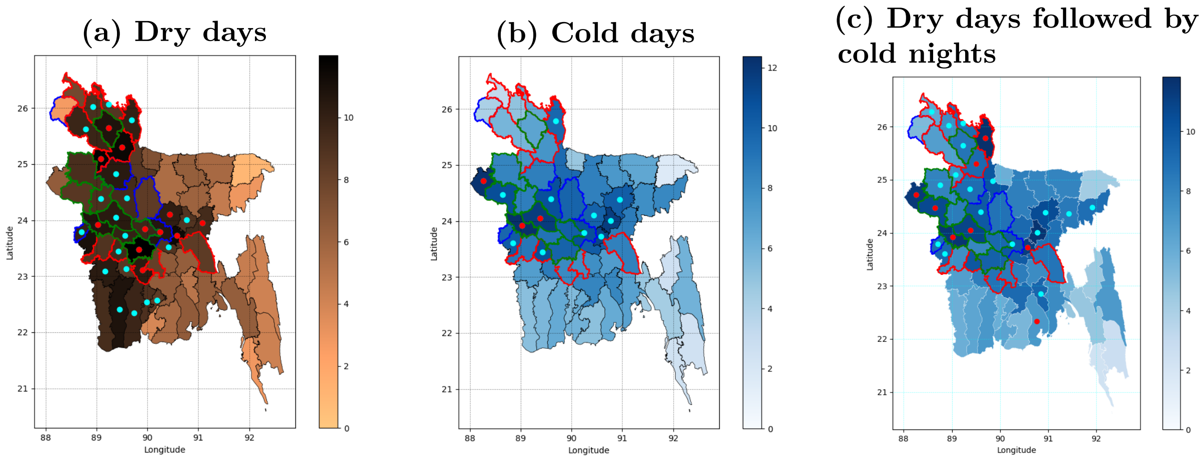


Figure 3.5: July anomalies in average temperatures and precipitation across various districts of Bangladesh from year 0 to year 17 under analysis. The top panel plots the annual number of spillover events, distinguishing periods before and after the establishment of Nipah virus surveillance. Red dots signify years with high spillover events, while blue dots represent years with low spillover events. The heat maps below show the temperature deviations from the norm for each district during the winter and monsoon seasons, respectively. This visualization aims to explore potential correlations between temperature anomalies and the frequency of spillover events.



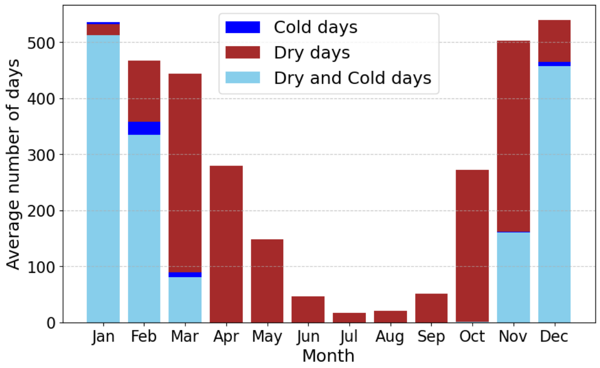
(a) Dry days

(b) Cold days

(c) Dry days followed by cold nights

Figure 3.6: Maps of observed differences in day counts between years with high and low spillover events for indexes: (a) dry days (b) cold days (c) dry days followed by cold nights). **Legend:** In both figures districts reporting statistically significant values are highlighted with colored dots. Blue dots refers to significance at 10% level, red at 5% level. District outlines indicate when a district was affected by spillover events: red for high spillover years, blue for low spillover years, and green for both.

(a) Monthly totals



(b) Observed monthly differences

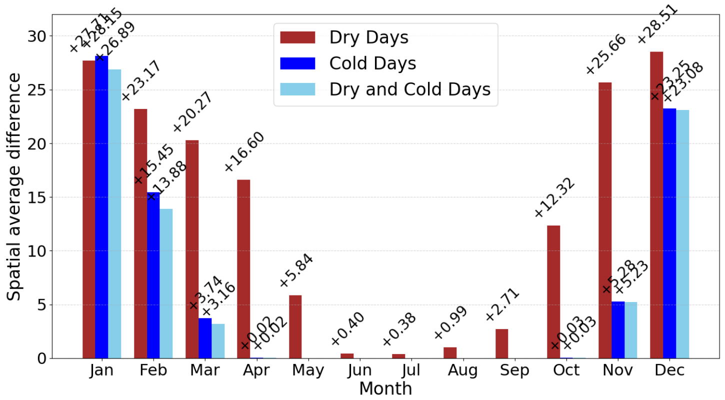


Figure 3.7: Spatial averages of: (a) Monthly totals for all statistically significant count variables, aggregated across districts and years: DD_y, CD_y, DCD_y; (b) Observed monthly differences in the number of dry days, cold days, and dry days followed by cold nights between years characterized by high and low spillover events.

Chapter 4

Spatial analysis of environmental stressors

4.1 Spatial distribution of environmental stressors

As extensively discussed in previous chapters, Bangladesh presents a convergence of numerous risk factors conducive to NiV disease outbreaks. Up to this point, the focus has been on analyzing temporal climate-driven stressors. However, the spatial distribution of human-induced ecosystem alterations also plays a crucial role, a factor that is not overlooked in this thesis. This chapter aims to reconstruct and analyze the spatial distribution of additional external stressors that are relatively stable over time or exhibit minimal variation, and are linked to land use and land cover.

While it is well established that the primary transmission route of NiV in Bangladesh is through the consumption of date palm sap (DPS), other exposure routes exists and have been considered to account for the complexity of NiV transmission dynamics (refer to Chapter 1.5.3 for further details). These include:

- **Intermediate-host transmission: Livestock** A wide range of domesticated animals (including pigs, goats, sheep, and cattle) are susceptible to NiV, and a direct contact with these infected intermediate hosts can potentially transmit the virus to humans (refer to Chapter 1.5.3 for further details). The role of these viral amplifier hosts has been examined by reconstructing their abundance and distribution at the district-level across Bangladesh.

- **Direct bat-to-human transmission: Food-borne sources** Humans can be infected through the consumption of food contaminated by bat secretions and excreta. Therefore, in addition to the date palm sap, bat dietary fruits—whether cultivated in garden orchards or plantations—are also considered [45], [42]. The abundance and distribution of these food sources have been meticulously reconstructed at the district level to assess their potential role in NiV transmission.

- **Human-to-human transmission: Population density** NiV can be transmitted from person to person through respiratory secretions and saliva, and high human population density has been associated to an increased risk of NiV spillover [55]. Consequently, the spatial distribution of population density has been analyzed to identify areas with elevated spillover risk.

- **Human-induced ecosystem alteration:** Extensive habitat loss and fragmentation, driven by deforestation for agricultural extension and urbanization, have significantly altered bat roosting and foraging behaviours (refer to Section 1.5). These changes have, in turn, increased bat-human interactions, thus the associated spillover risk. Therefore, the abundance of forest cover, built-up areas, cropland and rangeland have been analysed at the district level.
- **Biological and external climate-driven stressors:** Weather patterns can directly impact bat body conditions through heat or cold stress, and indirectly by altering the flowering and fruiting cycles of their dietary plants, leading to nutritional stress (see Section 1.6.3). These factors can weaken the bats' immune systems, increasing viral shedding (refer to Section 1.4.4 for further details). Therefore, the spatial distribution of temperature and precipitation has been analyzed to assess their effects on nutritional, thermoregulatory, and physiological stress in bats, which in turn exacerbate other ecological drivers.

The ultimate goal is to quantify and map the abundance of these established drivers, to identify areas of heightened risk, and to understand how different factors influence various regions, potentially clustering districts that share similar characteristics.

4.1.1 Data and data pre-processing

The methodologies in this chapter are designed to meticulously map and quantify the ecological and anthropogenic factors discussed earlier, assessing the abundance and spatial distribution of key environmental variables crucial to understanding NiV transmission dynamics in Bangladesh. Detailed descriptions of the data sources, along with their spatial and temporal resolutions, are provided in Table A.4 located in the Appendix A.

Land use and land cover by district

Land use data play a pivotal role in understanding how human activities have transformed natural ecosystems at the district level. To analyze these changes, four `.GeoTiff` raster tiles from the year 2018, which cover the entire region of Bangladesh, were sourced from the *Esri Land Cover - ArcGIS Living Atlas* website. Each raster has a spatial resolution of $10m$ and pixels with assigned values from 1 to 11, each representing a distinct land use category: water (pixel value 1), trees (pixel value 2), flooded vegetation (pixel value 4), crops (pixel value 5), built areas (pixel value 7), bare ground (pixel value 8), snow/ice (pixel value 9), clouds (pixel value 10), and rangelands (pixel value 11).

These tiles were imported into QGIS and merged into a single raster layer using with the `Merge Raster` tool and then reprojected from the WGS 84 - EPSG:4326 geographic coordinate system to the Gulshan 303 / TM 90 NE - EPSG:3106, which is specific to Bangladesh, using the `Warp` tool. Layer properties were adjusted by symbolizing the single band file using paletted/unique values, assigning specific colors to each pixel value. Additionally, the raster was clipped by the Bangladesh mask layer, to ensure the analysis was confined to the relevant geographical area (refer to Figure 4.1). To analyze the frequency distribution of land use and land cover classes within each districts, the QGIS `Zonal Histogram` tool was employed. This tool counts the pixels for each thematic class within polygons that represent district boundaries in a shapefile, facilitating the calculation of the area allocated to each land use type. This is achieved by multiplying the number of pixels by the area each pixel covers. Areas are then normalized to percentages of the

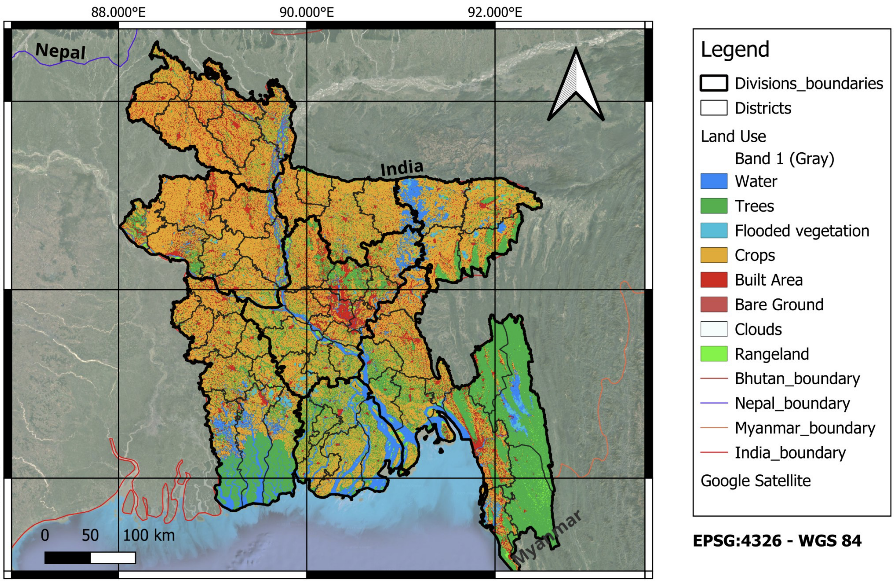


Figure 4.1: Land Cover and Land Use thematic map of Bangladesh pre-processed in QGIS (Resolution: 10 m). Raster source: ESRI Land Cover - ArcGIS Living Atlas of the World (Sentinel-2 L2A).

total district area (calculated using the dedicated `Field calculator` tool) to yield the proportional representation of each land use type within the districts.

Percentage of flood prone areas by district

Triggered by heavy monsoon precipitation, flooding is a recurrent hazard in Bangladesh, causing significant damages to lives, crops and infrastructures [87]. As discussed in Section 1.6.2, extreme weather events like floods and typhoons threaten *Pteropus* bats by affecting their roosting sites (i.e. defoliation) and increasing their exposure to post-storm hunting and predation. These external physiological stressors can contribute to impaired body conditions and enhance the recrudescence and transmission of NiV.

In line with this discussion, flood hazard maps in shape file format were retrieved from *ArcGIS hub*. These maps were used to quantify the percentage of each district that is prone to flooding, thereby deepening the undestanding of environmental challenges potentially impacting bat populations. Within the QGIS software, the 'Selection by Attribute' feature was employed to select areas categorised at "Severe" and "Moderate risk" levels, without distinguish among the different causes of flooding, such as tidal surges, riverine floods, or flash floods (refer to Figure 4.2). The `Dissolve` tool, was then used to merge these risk-specific areas into single polygons per district. Subsequently, the `Field Calculator` tool was employed to retrieve total and percentage area of flood-prone regions within each district.

Population density by district

Population density is another well-recognized trigger for NiV spillover. The *WorldPop* project provides gridded population data for Bangladesh in `.GeoTIFF` format, with a resolution of 30 arc seconds (about 1 km at the equator). This dataset, detailing the number of people per square kilometer, covers the years of interests from 2000 to 2018. The `Zonal Statistics` tool in QGIS was used to calculate the weighted average population density for each district over all years in the dataset (refer to Figure 2.1b).

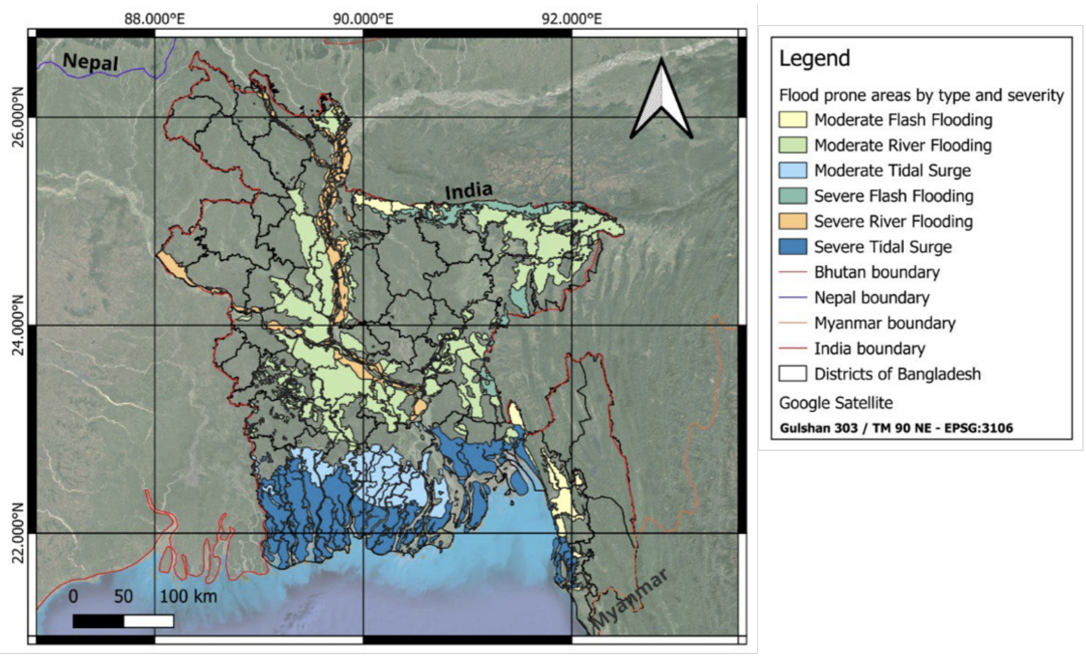


Figure 4.2: Hazard map of flood prone areas categorised by severity (Moderate and Severe) and type (tidal surge, river and flash flooding).

Density of intermediate hosts by district

To assess livestock abundance at the district level, gridded data with a resolution of 0.0833 decimal degrees were retrieved from the *Harvard Dataverse* website. These data belongs to the 2015 Gridded Livestock of the World database (GLW v4), collaboratively developed and maintained by FAO's Livestock Information, Sector Analysis and Policy Branch (NSAL), the Spatial Epidemiology Lab (SpELL), and Université Libre de Bruxelles.

The data for Bangladesh, derived from the Bangladesh Bureau of Statistics (BBS) for 2014, includes `.GeoTIFF` files, with pixels values indicating the absolute number of various livestock species. Among the available rasters, only those pertaining species susceptible to Nipah virus (i.e. pigs, cattle, sheep, and goats) were selected. Additionally, each dataset is provided together with a corresponding raster file, detailing the square kilometer area represented by each pixel.

To transform these raw counts into a measure of livestock density, the QGIS `Raster Calculator` tool was employed. This tool processed each paired set of raster files - one indicating the number of heads per pixel and the other the area per pixel - to create a new raster displaying the density of livestock per pixel.

Given the coarser resolution of these files (approximately 10 km at the equator) compared to the population density (1 km) and land use (10 m) data, enhanced data processing techniques were employed to achieve accurate areal density estimates matching the precision of the other datasets. Using *Python* in *Visual Studio Code*, a weighted average of livestock density by district was computed with the advanced capabilities of the `xagg` package (see Section 3.1.2 for details). The pre-processed data and the final output for cattle density across Bangladesh districts are presented in Figure 4.3. The corresponding figures for pigs, sheep, and goats are reported in Appendix A (refer to Figure A.4).

To derive a unique value representing the total livestock abundance in each district, the headcounts of each relevant species were summed, and an overall percentage was calculated.

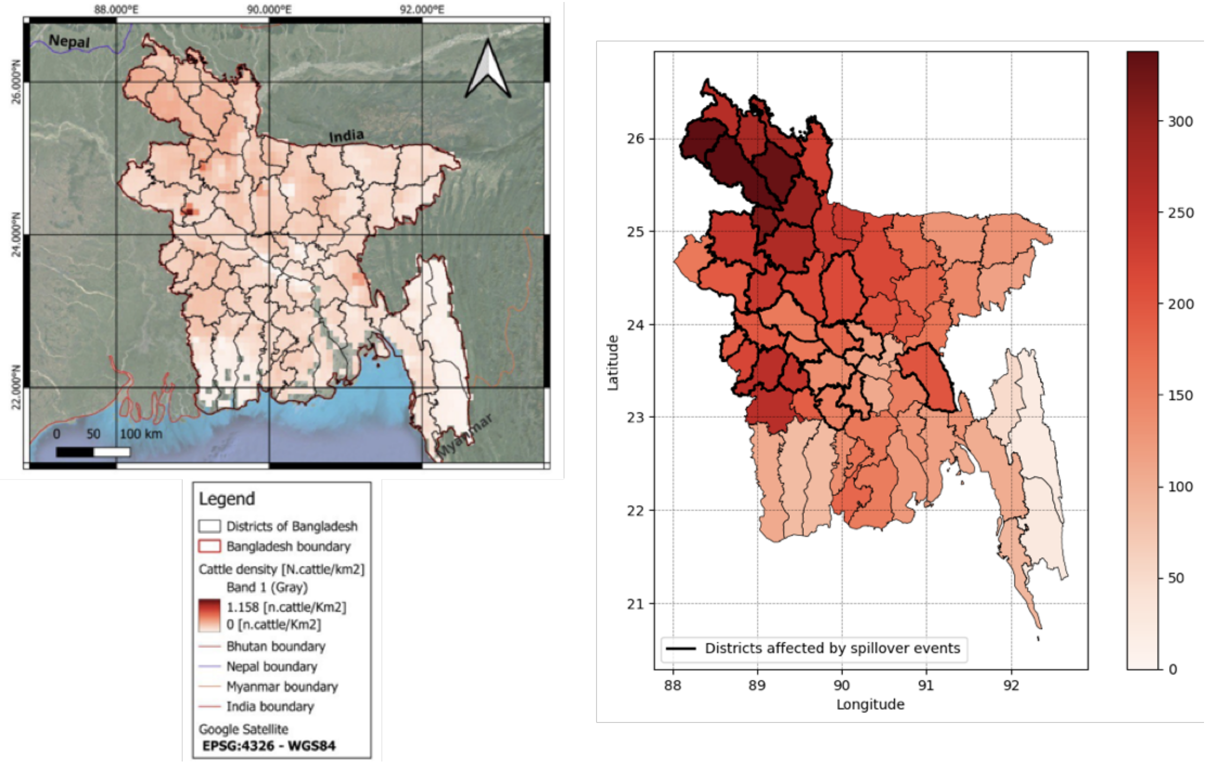


Figure 4.3: Pre-processed gridded data show cattle density per square kilometer (left). Average cattle density per district, color-coded to indicate different density levels (right). Bold black lines on district boundaries denote areas affected by NiV spillover events, identifying regions within the Nipah belt.

Percentage of fruit and sugar crops by district

To analyze the agricultural land use within districts of Bangladesh, specifically for fruit and sugar crops, data were extracted from the Bangladesh Bureau of Statistics via the ‘District Statistics Reports’ and the ‘Yearbook of Agricultural Statistics’, published in 2013 and 2022, respectively.

Data for major permanent fruit crops—Ripe Papaya, Guava, Jackfruit, Banana, Pineapple, Mango, Litchi, and Blackberry—were sourced from district-specific reports based on the 2010/2011 agricultural census, detailing the area of each crop per district.

Additional data were retrieved from the latest ‘Yearbook of Agriculture’ to include other crops vital to local bat diets. Thus, relevant crops and flowering plants such as *Jujube (Boroi/Kul), Star Apple (Jamrul), Ata (Custard Apple), Dewya, Lotkon, Carambola Apple (Kamranga), Wood Apple, Green Coconut (flowers), Dragon Fruit, Dalim Fruit, Safeda (Sapodilla), Sharifa*, and *Strawberry* were cataloged. The 2019/2020 records listed these crops in acres as ’area under garden’, reflecting cultivation practices in household gardens, orchards, and farms—a major portion of district land dedicated to these crops. *Strawberry* and *Water Fruit* were recorded as ’area under trees’, a category that includes areas within and outside garden boundaries, providing a broader view of their spatial distribution.

Furthermore, the cultivation of *Date-palm* and *Palmyra palm*—crops grown for both their juice (or Tal) and fruits—was analyzed. These sugar crops were quantified in terms of ’area under gardens’ in acres, indicating their cultivation within designated garden spaces.

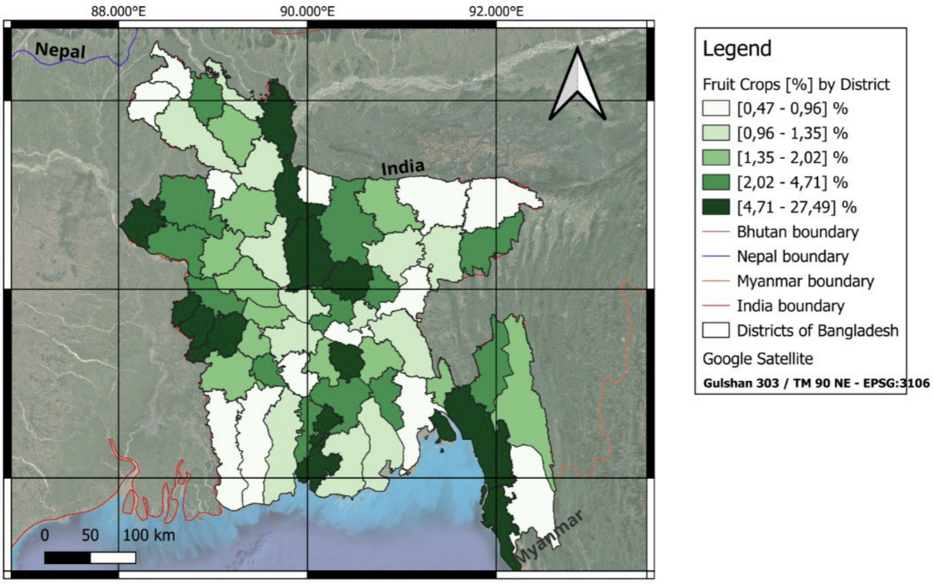


Figure 4.4: Spatial distribution of fruit crops in Bangladesh, with darker green shades indicating districts with higher percentages of land area dedicated to fruit cultivation relevant for bat diets, showcasing regions where these resources are most abundant.

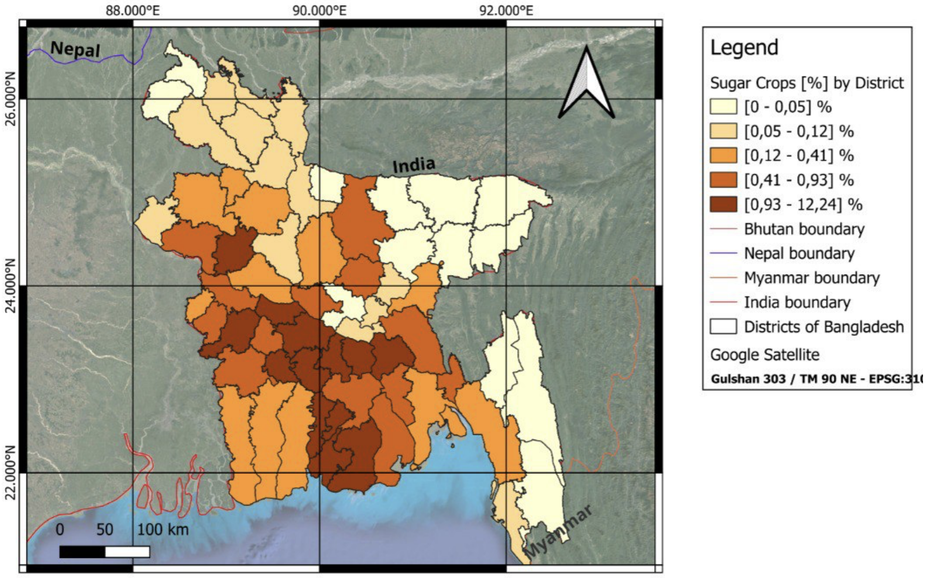


Figure 4.5: Spatial distribution of sugar crops in Bangladesh including Date Palms, with darker orange shades indicating districts with higher percentages of land area dedicated to sugar cultivation relevant for bat diets, showcasing regions where these resources are most abundant.

All data, including district-level crop areal values and the re-projected administrative areas in square kilometers calculated in QGIS, were systematically organized and processed in Excel. Calculations included converting areas from acres to square kilometers to compute the percentages of district lands devoted to various crops, and aggregating the percentages of all fruit and sugar crops by district.

Before proceeding, the reliability and the accuracy of these calculations were validated by comparing the aggregated district percentages nationally with the statistics reported in the 2022 Yearbook. The results were reasonably consistent with national benchmarks;

specifically, 2.72% for fruit crops aligns with the reference value of over 2.48%, and 0.53% for sugar crops is considered acceptable given the reference of over 0.97%. Notably, the result for sugar crops, despite being less precise, is deemed acceptable considering the exclusion of sugarcane from the analysis and the fact that only areas within household gardens were considered.

As a result, a dataset detailing areal percentages by crop type and district was compiled. This non-geometric dataset was then exported as a `.CSV` file and imported into QGIS, where it was `joined` with a shapefile of the district boundaries. This integration allowed the creation of spatially referenced data, enabling further geo-spatial analyses. In particular, this comprehensive analysis offers a detailed overview of the agricultural landscape across Bangladesh's districts, highlighting the varying abundance of specific crops relevant to the diet of bats (refer to Figures 4.4 and 4.5).

4.2 Spatial clustering of districts

To cluster districts into contiguous areas sharing similar land use, land cover, and other NiV ecological drivers, the following district-level features were analyzed:

- Average monthly climatic features, including temperature and precipitation, were calculated over the 18-year reference period. Specifically, for each district, monthly averages were first computed annually, and then averaged across the entire period.
- Percentage of relevant land-use categories, including trees, water bodies, crops, built-up areas, rangeland, and bare ground, calculated at the district level. Special focus was placed on the percentage of land dedicated to fruit and sugar crops, which are key to bat diets.
- Livestock and human population density, averaged annually over the 18-year reference period.
- Percentage of flood-prone areas, calculated at the district level.

The *K-means clustering* algorithm was employed for grouping, with the data standardized using the `Standard Scaler` from `scikit-learn`. This pre-processing step normalized the spatial data by setting each variable's mean to zero and standard deviation to one, ensuring that all features contributed equally to the analysis, eliminating bias caused by differences in units and scales.

4.2.1 Methods

K-means clustering

The K-means algorithm is an iterative technique employed to partition a dataset into a predefined number of non-overlapping clusters, ensuring each data point belongs exclusively to one cluster. Specifically, it organizes N observations $X = \{x_1, x_2 \ldots, x_N\}$ into K clusters $C = \{C_1, C_2, \ldots, C_i, \ldots, C_K\}$, where each observation is a D-dimensional variable x, and K is set to be less than N. In this study, the dataset comprises 64 districts of Bangladesh, each characterized by a 35-dimensional feature vector that reflects key factors for NiV spillover risk analysis such as climate data, land-use patterns, livestock and population densities, and flood-prone area percentages.

Each cluster C_i is defined by a *centroid* μ_i, which is the mean of all data points in C_i and acts as the cluster's prototype [108], [109]. The centroids are 35-dimensional vectors representing the average characteristics of the districts within each cluster.

The algorithm's goal is to minimize the *within-cluster sum-of-squares* (WCSS) (also referred to as *inertia*), which is the total squared Euclidean distance between each district's feature vector and its associated centroid [108], [109]. This minimization ensures that districts within the same cluster are as similar as possible in terms of NiV risk factors, while maximizing the differences between clusters. The iterative process of assigning districts to the nearest centroid and updating centroids continues until the cluster assignments stabilize, indicating an optimal clustering solution has been reached [108], [109]. This results in a detailed grouping of districts based on ecological similarities that may influence NiV transmission dynamics. For a detailed explanation of the K-means clustering process, including specific formulas and step-by-step methodology, please refer to Section A.4.1 in Appendix A.

Determination of the optimal number of clusters

To determine the optimal number of clusters, *Silhouette analysis*, *Elbow method* and *Principal Component Analysis* (PCA) were employed.

Silhouette analysis Silhouette analysis evaluates the cohesion within and separation between clusters by calculating the *Silhouette score* or *coefficient* s for each data point i, which is a standardized measure of the separation distance between an observation and the decision boundary or the nearest cluster [110]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4.1}$$

Where:

- $a(i)$ is the intra-cluster distance, thus the average distance between the data point i and all other data points within the same cluster.
- $b(i)$ is the nearest-cluster distance, thus the distance between the data point i and the nearest cluster that the data point doesn't belong to.

This score ranges from -1 to $+1$, where a score close to $+1$ suggests a well-fitted data point in its cluster, and scores near -1 indicate potential misplacements [111].

Silhouette plots for cluster counts ranging from 2 to 7 were examined to determine the optimal number of clusters (K). The selection criteria included maximizing the average silhouette score ($\bar{s}$), minimizing negative silhouette values (s), and reducing variations in the thickness of the silhouette plot, which reflects cluster uniformity. In addition to silhouette analysis, the *Elbow method* was employed to comprehensively assess the clustering results. In Figure4.7-a is reported the Silhouette plot for $K = 5$.

Elbow Method The Elbow method plots the within-cluster sum of squares (WCSS), or inertia, against different K values. A visible 'elbow' (i.e. point at which the decrease in inertia begins to slow down) typically indicates an optimal number of clusters.

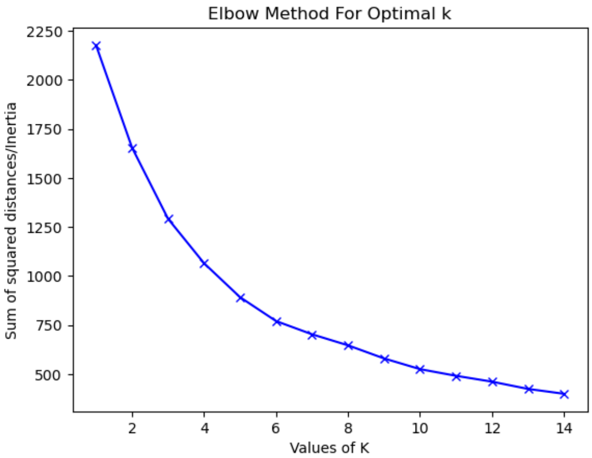


Figure 4.6: Elbow method's plot for determining the optimal number of clusters (K). The plot shows the sum of squared distances (inertia) against the number of clusters, with the "elbow" indicating the optimal K.

Principal Component Analysis (PCA) *PCA* was conducted to reduce the dimensionality of the dataset, enabling clearer visualization of cluster separation in a two-dimensional space. The analysis involved transforming the 35 standardized spatial variables into two principal components that captured the most significant variances within the dataset. Specifically, the first principal component accounted for the largest variance, while the second one represented the most variance remaining after accounting for the first component's effects [112]. The PCA was implemented using the `covariance_eigh solver` from `scikit-learn`. In Figure 4.7, the biplot displays each district of Bangladesh. Colors denote clusters assigned by the K-means algorithm, and the spatial arrangement of points reflects the principal components. This visualization aids in determining the optimal number of clusters by illustrating the separation and cohesion among them, facilitating

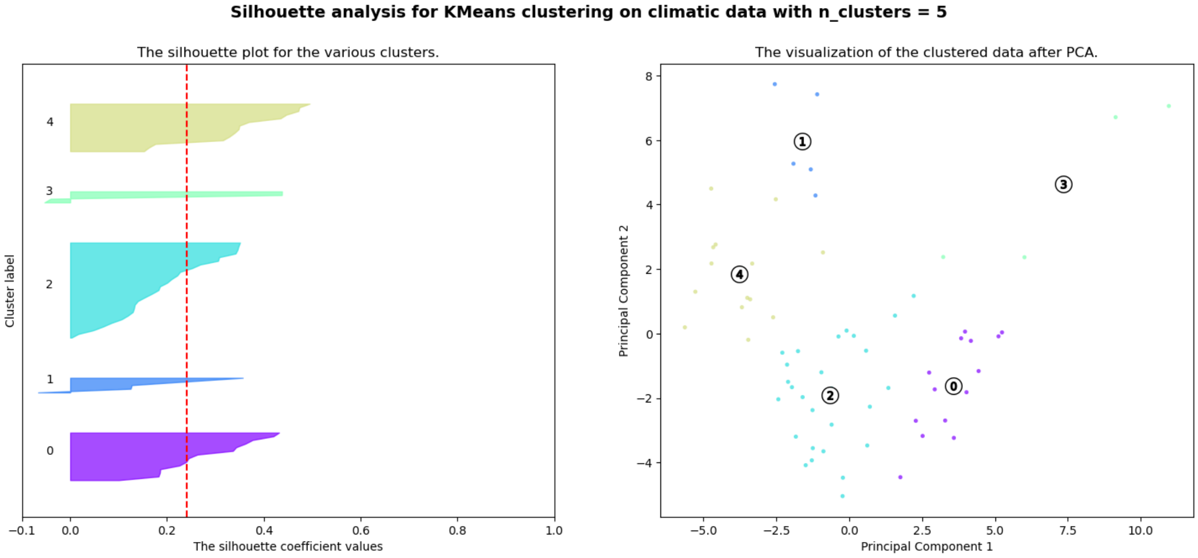


Figure 4.7: (**Left**) Silhouette plot for $K = 5$ showing the s coefficient values for each cluster, with the average silhouette score $\bar{s}$ indicated by the vertical red dashed line; (**Right**) Visualization of the clustered data after PCA, with each point representing a district colored according to its cluster assignment.

the visual assessment of clustering efficacy.

Therefore, in choosing the optimal cluster configuration, several metrics were considered:

1. **Average silhouette score $\bar{s}$:** High scores indicate distinct and well-defined clusters, suggesting effective grouping;
2. **Distribution of silhouette scores s:** Minimal negative scores suggest correct cluster assignments and good fit within clusters;
3. **Silhouette scores visualization:** Evaluating the spatial distribution of scores on the map (Figure 4.8-b) helps confirm the spatial coherence of clusters;
4. **Elbow plot analysis:** This plot guides the selection of the number of clusters by showing the point where increases in cluster count yield diminishing returns in model improvement;
5. **PCA visualization:** This analysis aids in visually confirming the distinctness and appropriate separation between clusters.

These metrics collectively inform the selection of the most appropriate clustering configuration, grounded in the ecological and climatic variables relevant to NiV transmission dynamics.

4.2.2 Results and Discussion

K-means clustering results

Considering the comprehensive evaluation metrics, the optimal number of clusters is determined to be five. This choice is substantiated by the high average Silhouette score of 0.342, representing one of the highest observed, suggesting significant internal cohesion and distinct separation among clusters (refer to the left-hand side panel in Figure 4.8). Additionally, the PCA visualization supports this selection by demonstrating clear demarcation and minimal overlap between clusters, validating the distinctiveness of each cluster. Notably, using the `.explained_variance_ratio_` attribute of `scikit-learn`, it has been established that the first two principal components account for 62% of the data's variance, thereby confirming the PCA's effectiveness in capturing the primary trends in the dataset.

The Elbow method reinforces this decision, showing a balance between reducing the within-cluster sum-of-squares and avoiding overfitting by not excessively increasing the number of clusters (refer to Figure4.6).

While some Silhouette coefficients are negative, indicating potential misclassification, these primarily concern a few districts in clusters 1 and 3. This is clearly depictable in both the map of Silhouette coefficients (Figure 4.8) and the Silhouette plot (Figure 4.7). However, the geographical positioning of these districts—located at the extreme northeast and southeast—is of minimal interest for subsequent analyses focused on the Nipah belt region. This observation reinforces that the clustering is robust for districts impacted by spillover events, which are central to this thesis.

Notable result: Furthermore, transitioning from a clustering model based solely on climatic variables to this more comprehensive approach, which includes additional ecological drivers, has significantly improved the clustering. This refined model more accurately reflects the geographical layout of the Nipah Belt, dividing it into distinct northern and southern regions, highlighting the importance of integrating these factors to better understand spillover dynamics.

These findings suggest that a five-cluster configuration offers a robust and meaningful segmentation of districts based on ecological and climatic data relevant to NiV spillover risks.

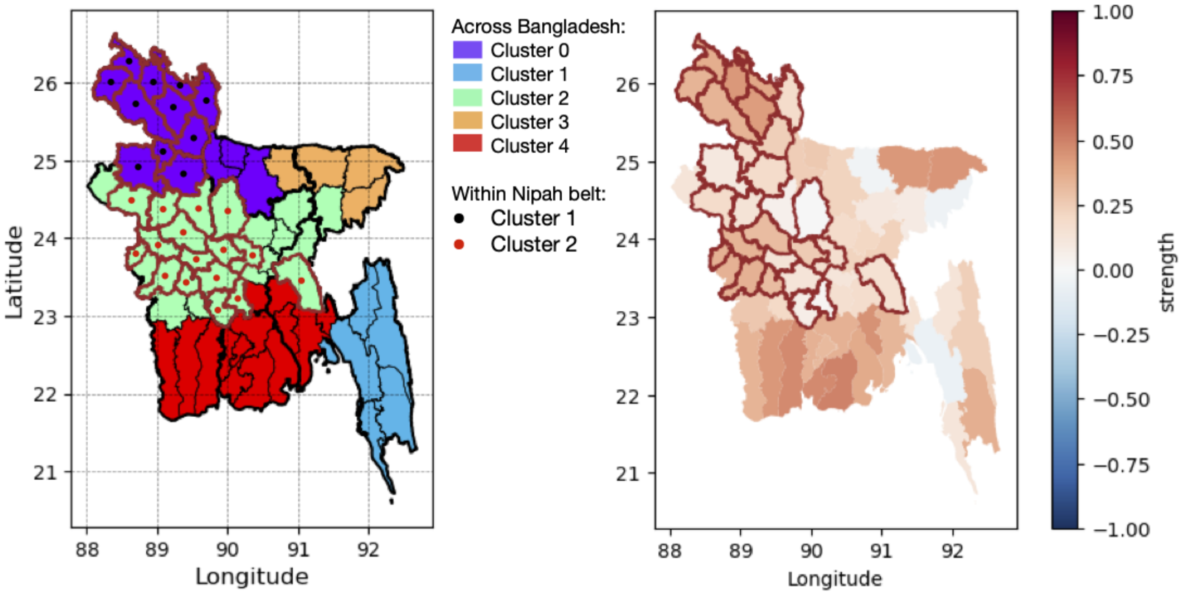


Figure 4.8: **(Left)** K-means clustering of Bangladesh districts based on land use, land cover and monthly climatic features, showing the optimal configuration with $K = 5$ clusters. Different colors represent the clusters across Bangladesh. Within the Nipah belt, clusters 1 and 2 are specifically highlighted with black and red dots, respectively. **(Right)** Map of Silhouette coefficients illustrating cohesion and separation strength of each district's cluster, represented with a red-to-blue color gradient. The red boundaries indicate the NiV-affected districts.

In this configuration, the affected districts within the Nipah Belt are divided into two main groups:

- **Cluster 1** includes ten districts, primarily from the Rangpur and Rajshahi divisions in the northwest of Bangladesh (see districts marked with black centroids in Figure 4.8).
- **Cluster 2** encompasses fifteen districts in the central Nipah Belt, spanning the Rajshahi, Khulna, and Dhaka divisions (see districts marked with red dots in Figure 4.8). The Comilla district, although affected by spillover events, is the only non-contiguous district in this group. It is excluded from subsequent analyses due to incomplete data on spillover events in this area.

Geospatial analysis findings

Unique cluster: Nipah belt The spatial analysis corroborates existing literature, identifying the Nipah Belt as a region characterized by high human population density, extensive built-up areas, widespread agricultural land use, and reduced tree cover (refer to Figure 4.9) [33], [55]. Fruit crops essential to bat diets are primarily located along the periphery of the Nipah Belt, while sugar crops, including date palms, are concentrated in the central and southern regions. These agricultural patterns show no significant proportional differences between areas inside and outside the spillover zones. Moreover, while flood-prone areas are predominantly outside the Nipah Belt (see bar plot in Figure 4.9), significant at-risk regions stretch across it from north to south (Figure 4.10b). Rajbari and Faridpur districts stand out as the most flood-susceptible, coincidentally reporting higher densities of date palms (Figure 4.10c) and being among the districts most affected by NiV spillover events over the 18-year reference period (Figure 2.6b). This overlap suggests a potential link between flooding, storm events, and defoliation—conditions that increase bats' exposure to predation, hunting, and adverse weather as they roost in exposed branches (refer to Section 1.6.3 for further details). These environmental disturbances could intensify human-bat interactions, particularly in areas with dense date palm cultivation, amplifying the risk of NiV transmission.

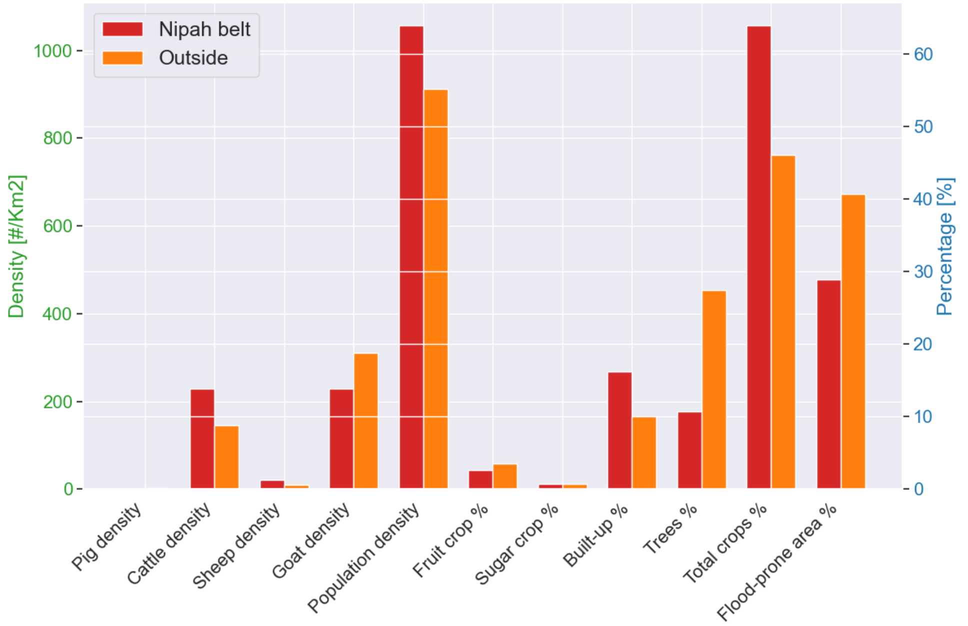


Figure 4.9: Nipah belt as a unique cluster Comparison of established ecological drivers for NiV spillover events between the Nipah belt (shown in red) and region outside the belt (shown in orange). District-level values and percentages are averaged within the two clusters: the Nipah belt and the remaining districts of Bangladesh.

Two clusters within the Nipah belt The analysis of the two clusters within the Nipah belt reveals distinct ecological and climatic characteristics. Cluster 1 is marked by extensive agricultural land use, higher urbanization, fewer tree cover, and greater cattle density compared to Cluster 2 (refer to Figure 4.11). Furthermore, climate analysis over the 18-year reference period (2000-2018) shows that Cluster 1 experiences colder and wetter conditions almost year-round (see Figure 4.13), with particularly lower temperatures during the December-February (DJF) months compared to Cluster 2 (refer to Figure 2.3). In contrast, Cluster 2 is generally warmer and drier (see Figure 4.13), especially during the monsoon period, characterized by higher temperatures and significantly lower rainfall than Cluster 1 (see Figure 2.3).

These observations suggest that diverse environmental and climatic patterns differentiate the two clusters:

- **Cluster 1** is distinguished by harsh winters, a landscape dominated by urban and agricultural areas in place of forests, and a significant presence of cattle.

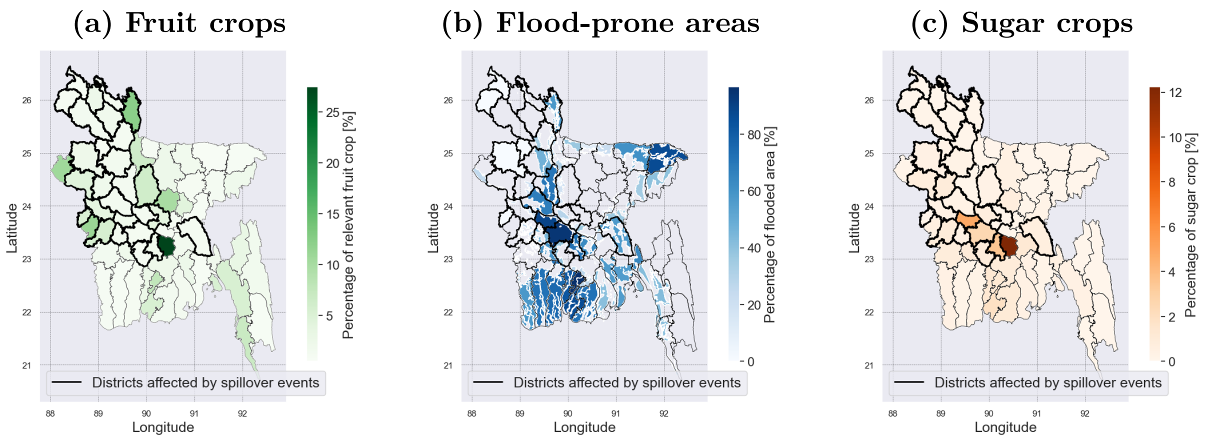

(a) Fruit crops (b) Flood-prone areas (c) Sugar crops

Figure 4.10: Spatial distribution of district-level percentages for land devoted to (a) fruit and (c) sugar crops, as well as (b) flood-prone areas. Black-bordered districts indicate those belonging to the Nipah belt, which were affected by spillover events during the 2000-2018 period.

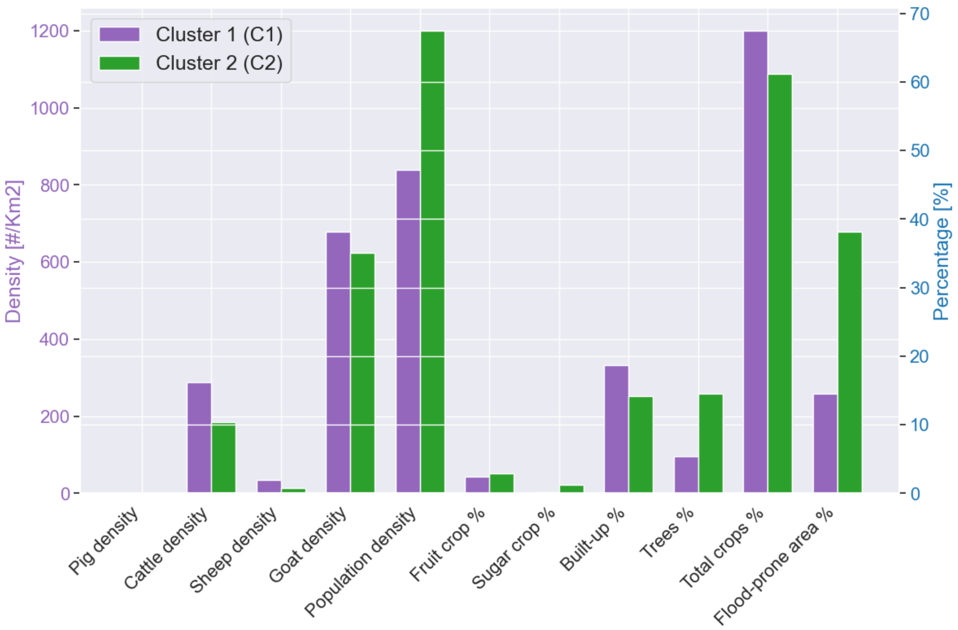
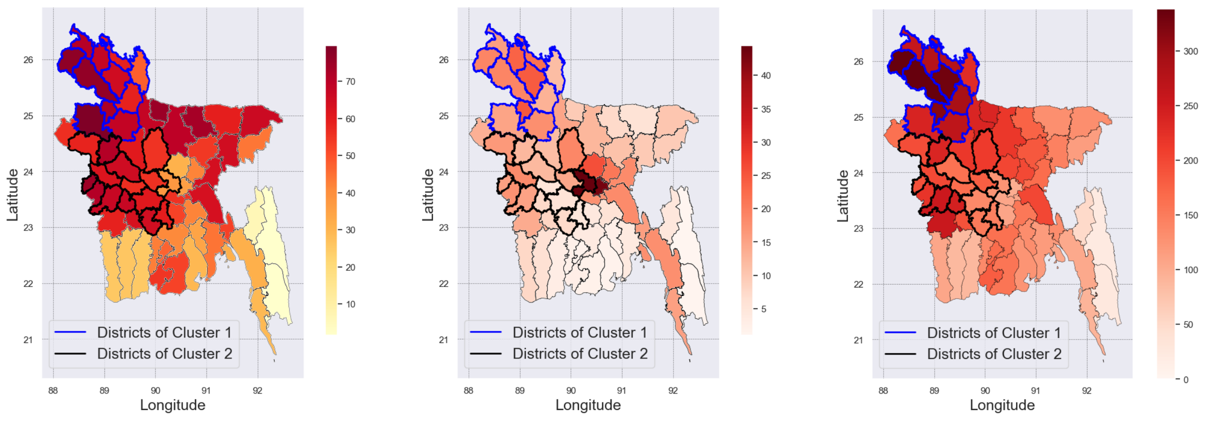


Figure 4.11: Two clusters within the Niaph belt Comparison of established ecological drivers for NiV spillover events between the two clusters within the Nipah belt: cluster 1 (shown in purple) and cluster 2 (shown in green). District-level values and percentages are averaged within the two clusters.

- **Cluster 2**, on the other hand, is noted for its hot and dry monsoon periods, higher population density, and more prevalent flood-prone areas.

This differentiation suggests that diverse factors may distinctly influence the NiV transmission dynamics within each cluster. Variations in climatic conditions and land use characteristics across clusters may specifically alter bat roosting and foraging behaviors, influencing their interactions with humans. As a consequence, this spatial agglomeration of unique environmental stressors may trigger a spatial heterogeneity in NiV transmission dynamics and spillover patterns, altering the risk profile across the clusters. The subsequent Chapter will employ regression analysis to better understand which variables most



(a) Crops **(b) Built-up areas** **(c) Cattle density**

Figure 4.12: Spatial distribution of district-level (c) cattle density and percentages for land devoted to (a) agricultural crops and (b) human settlements. Black-bordered districts indicate those of Cluster 2, while blue-bordered districts indicate those of Cluster 1.

effectively explain the temporal variability of spillover counts within each cluster.

This integrative approach sets the stage for applying robust statistical methods to identify and quantify the specific environmental and climatic drivers that critically influence NiV spillover within these spatially and ecologically distinct clusters.

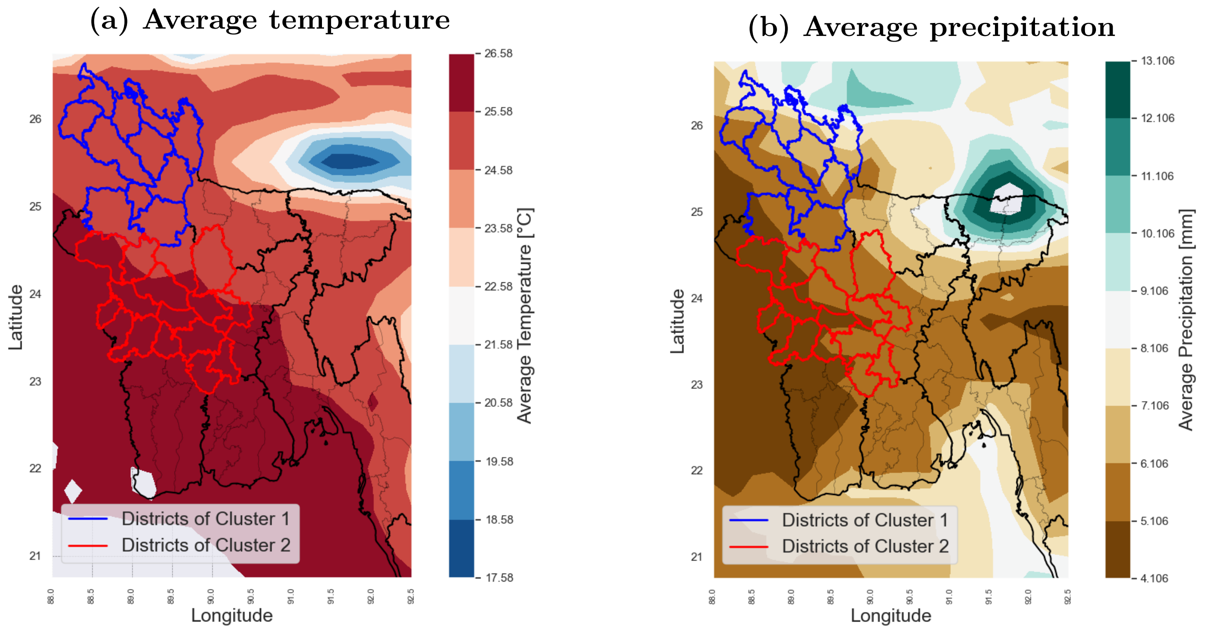

(a) Average temperature **(b) Average precipitation**

Figure 4.13: Contour plot of average annual temperature and precipitation. Red-bordered districts represent those in Cluster 2, while blue-bordered districts indicate those in Cluster 1.

Chapter 5

Regression analysis of climate variables, extremes and spillover events within spatial clusters

5.1 Regression analysis: Generalized Linear Models

Building on the findings from Chapter 3, which identified climatic drivers potentially linked to increased spillover events during high-risk years (Y_H), and the spatial analysis in Chapter 4, which revealed significant heterogeneity in environmental stressors both between the Nipah belt and surrounding regions, and within the Nipah belt itself-divided into two distinct clusters-this Chapter integrates these insights through a comprehensive regression analysis using Generalised Linear Models (GLMs).

The goal is to test and quantify the hypothesized relationships between the selected climate variables and the annual variability in spillover events, while also accounting for spatial variability at different resolutions.

Following the approach of McKee et al. [33], the association between yearly independent NiV spillover events and climate statistics of selected variables was examined using separate univariate generalized linear models. Model selection was subsequently conducted to identify the best-fitting combination of climate variables using Akaike's information criterion corrected for small sample sizes (AICc). However, this analysis improve upon the original study by utilizing ERA5 reanalysis data (see Section 3.1.1), extending the temporal scope to include variables beyond the spillover period, exploring potential interactions between selected variables to enhance model accuracy, and adopting a negative binomial distribution to better address the observed data dispersion.

5.1.1 Data and data pre-processing

Building on the statistically significant results from the climate analyses in Chapter 3, this section investigates additional variables and extreme event indices, summarized in Table 5.1. To replicate the methodology from the reference article by McKee et al. [33], the variables $Pr_{cum,winter}$ (total winter precipitation) and CD_{winter} (number of winter cold days) were added. For greater accuracy, the study also includes DD_{winter} (number of winter dry days) and DCD_{winter} (number of winter dry days followed by cold nights). These additional variables are highlighted in blue in the table, while the temporal resolutions explored in Chapter 3 but excluded from the final regression analysis are marked in red. These were omitted to prevent data fragmentation and to maintain focus on the most impactful periods, as detailed below.

Temporal aggregation considerations In the regression analysis, both seasonal and monthly data were considered, depending on the period:

- For the **spillover period** (winter and pre-monsoon), seasonal aggregation of climate variables was prioritized to avoid data fragmentation. In the **winter period** (DJF), where 74 of the 101 spillover events occurred, both December and February were statistically significant in the monthly analysis. Aggregating these months into a single winter season was more effective, as it captured the overall winter dynamics influencing spillover events, providing a cohesive and coherent approach while preventing unnecessary fragmentation.
  Similarly, for the **pre-monsoon period** (MAM), March showed significant anomalies, but the entire season was analyzed to capture spillover events extending into March and April. However, due to the lack of significant anomalies in the other months, pre-monsoon variables were excluded from the final analysis to maintain focus on the most impactful periods and prevent unnecessary fragmentation.
- For the **pre-winter periods**, including the **monsoon** (JJAS) and **post-monsoon** seasons (ON), both seasonal and monthly data were explored. Specifically, July and November were analyzed alongside their respective seasons to determine which temporal aggregation better explains the annual variability in spillover events.

This approach, by considering both seasonal and monthly temporal aggregations and focusing on the most relevant periods, ensures a thorough and structured examination of the climatic factors influencing spillover events. It also provides a solid foundation for comparison with previous research findings.

Seasonal results		**Monthly results**		**Indexes**	
Monsoon	$Tavg_{monsoon}$ $Pr_{monsoon}$	**July**	$Tavg_{july}$, $Tmin_{july}$, $Tmax_{july}$, $Trange_{july}$, Pr_{july}	**Yearly**	DD_y, CD_y, DCD_y
Post-monsoon	Pr_{ON}	**November**	Pr_{nov}		
Winter	$Tavg_{winter}$, $Tmin_{winter}$, $Pr_{cum,winter}$	**December** **February**	$Tavg_{dec}$, $Tmin_{dec}$, $Tmax_{dec}$ $Tavg_{feb}$, $Tmin_{feb}$, $Trange_{feb}$	**Winter**	DD_{winter}, CD_{winter}, DCD_{winter}
Pre-monsoon	-	**March**	$Trange_{march}$, Pr_{march}		

Table 5.1: Summary of climate variables and indices analyzed. Variables shown in black and red were statistically significant in the climate analysis (see Chapter 3). Variables in red were ultimately not included in the regression analysis (see considerations in paragraph 5.1.1). Variables in blue were added to replicate the methodology from the reference article [33], allowing for a direct comparison of results.

5.1.1.1 Pre-processing: temporal resolution

Following these temporal aggregation considerations, a data pre-processing phase was conducted to ensure that the temporal resolution was appropriate for the analysis:

- For temperature variables ($Tavg, Tmin, Tmax, Trange$), daily data aggregated by district was resampled to compute monthly and seasonal averages, with July month extracted for focused analysis.
- Precipitation data was similarly aggregated to both monthly (July and November) and seasonal levels, with additional calculations for total winter precipitation to align with previous studies [33].
- Extreme event indices ($DD_y, CD_y, DCD_y, DD_{winter}, CD_{winter}, DCD_{winter}$) were converted into percentages to allow for a consistent comparison across years and seasons.

5.1.1.2 Pre-processing: spatial resolution

Building on the results of Chapter 4, the analysis is conducted at two different spatial scales:

Case A: Single spatial cluster All districts ($\hat{d} \in \hat{D} \subset D$), within the Nipah belt, that experienced at least one spillover event over the 18-year period (May 1, 2000, to April 30, 2018) are analyzed as a single cluster, with the exception of Comilla district (see Section 4.2.2 for details). This approach aligns with the methodology used in McKee et al. [33], facilitating comparisons. The cluster, referred to as $\text{cluster}_{\text{unique}}$, includes the following districts:

$$\begin{aligned}\text{cluster}_{\text{unique}} = \{\hat{d} \mid \hat{d} \in [&\text{'Panchagarh', 'Thakurgaon', 'Dinajpur', 'Rangpur', 'Kurigram',}\\ &\text{'Bogra', 'Joypurhat', 'Gaibandha', 'Lalmonirhat', 'Nilphamari',}\\ &\text{'Meherpur', 'Dhaka', 'Tangail', 'Gopalganj', 'Jhenaidah', 'Mada-',}\\ &\text{-ripur', 'Magura', 'Naogaon', 'Pabna', 'Natore', 'Kushtia', 'Raj-}\\ &\text{-bari', 'Manikganj', 'Faridpur', 'Rajshahi']}\}\end{aligned}$$

For each year, weather summary statistics are calculated for this cluster, with area-based weighting applied to each district to ensure that larger districts have a proportionally greater influence on the overall results. The weights were determined using the areas of the administrative units ($A_{\hat{d}}$), which were calculated by reprojecting in the `Gulshan 303 / TM 90 NE - EPSG:3106` coordinate system the shapefile of the Bangladesh districts in QGIS. The weight $w_{\hat{d}}$ for each district was calculated as follows using the `Field calculator` tool:

$$\forall \hat{d} \in \hat{D}: \quad w_{\hat{d}} = \frac{A_{\hat{d}}}{A_{\text{cluster}_{\text{unique}}}}; \quad A_{\text{cluster}_{\text{unique}}} = \sum_{\hat{d} \in \text{cluster}_{\text{unique}}} A_{\hat{d}} \tag{5.1}$$

For each variable X, the spatial average ($< \cdot >$) was computed for each year y as:

$$\forall y: \quad < \mathrm{X}(y) > = \sum_{\hat{d} \in \text{cluster}_{\text{unique}}} \mathrm{X}_{\hat{d}}(y) \cdot w_{\hat{d}} \tag{5.2}$$

Accordingly to this spatial resolution, the annual spillover counts were calculated as the sum of events across all affected districts within the unified spatial cluster:

$$\forall y: \quad \mathrm{Spill}_{tot}(y) = \sum_{\hat{d} \in \text{cluster}_{\text{unique}}} \mathrm{Spill}_{\hat{d}}(y) \tag{5.3}$$

Case B: Two spatial clusters Based on the spatial analysis from Chapter 4, districts within the Nipah belt are divided into two clusters (Cluster 1 and Cluster 2, representing the northern and southern regions, respectively) according to climatic and spatial characteristics. Specifically, the clusters were defined as:

- $\text{cluster}_1 = \{\hat{d} \mid \hat{d} \in$ ['Panchagarh', 'Thakurgaon', 'Dinajpur', 'Rangpur','Kurigram', 'Bogra', 'Joypurhat', 'Gaibandha', 'Lalmonirhat', 'Nilphamari']$\}$;
- $\text{cluster}_2 = \{\hat{d} \mid \hat{d} \in$ ['Meherpur', 'Dhaka', 'Tangail', 'Gopalganj', 'Jhenaidah', 'Ma--daripur', 'Magura', 'Naogaon', 'Pabna','Natore', 'Kushtia', 'Rajbari', 'Manikganj', 'Faridpur', 'Rajshahi']$\}$.

For each year, weather summary statistics are calculated separately for these two clusters, with area-based weighting applied to each district similarly to Case A, ensuring that larger districts contribute proportionally more to the overall results.
In this case, the annual spillover counts for each spatial cluster are calculated separately as follows:

$$\forall y : \quad \text{Spill}_{C1}(y) = \sum_{\hat{d} \in \text{cluster}_1} \text{Spill}_{\hat{d}}(y)\,; \quad \text{Spill}_{C2}(y) = \sum_{\hat{d} \in \text{cluster}_2} \text{Spill}_{\hat{d}}(y) \tag{5.4}$$

For each cluster, data distribution was analyzed, revealing overdispersion, where the variance exceeded the mean. Consequently, a negative binomial model was fitted to the data, with parameters estimated using the Maximum Likelihood Estimation (MLE) method.

5.1.2 Methods

To explore the relationship between the response variable (i.e., annual spillover events at each cluster resolution) and a set of independent explanatory variables (i.e., climate variables and extreme indices), a regression analysis using Generalized Linear Models (GLMs) was conducted.

Generalized Linear Models (GLMs), introduced by Nelder and Wedderburn in 1972 [113], extend traditional linear regression by offering a flexible framework suitable for various types of response variables, including count data. This flexibility makes GLMs particularly appropriate for analyzing the annual number of Nipah virus spillover events, which range from 0 to 8, 0 to 10, or 0 to 17, depending on the spatial cluster analyzed (Cluster 1, Cluster 2, or the Unique cluster, respectively).

While the *Poisson regression model* is commonly used for count data, the presence of overdispersion in the dataset—where the variance exceeds the mean—requires the use of the *negative binomial regression model.* This model introduces a dispersion parameter (α) to handle the excess variability that the Poisson model cannot account for.

Following the confirmation of overdispersion in the spillover data across clusters, negative binomial GLMs were employed to model the expected annual count of spillover events as a function of the selected climate variables. A log link function was used to relate the linear predictors to the expected response, with the models implemented using the `statsmodels.api` module from the `statsmodels` library in Python.

The negative binomial regression model, with a log link function, is expressed as [114]:

$$log(\mu) = \eta = X\beta = \beta_0 + \beta_1 x_1 + .. + \beta_p x_p \tag{5.5}$$

where μ represents the expected (mean) count of yearly spillover events, X is the matrix of climate variables, and β is the vector of regression coefficients to be estimated. The log link function ensures that μ is related to the linear predictor η by:

$$\mu = e^{\eta} = e^{(X\beta)} = e^{(\beta_0 + \beta_1 x_1 + .. + \beta_p x_p)} \tag{5.6}$$

Model parameters, including the dispersion parameter (α) and regression coefficients (β), were estimated using the *Maximum Likelihood Estimation* (MLE) method. The negative log-likelihood function was optimized using the `L-BFGS-B` algorithm to ensure the best fit to the data, while addressing overdispersion in the spillover events and their relationship to climate variables.

For further theoretical details on the negative binomial model and parameter estimation, please refer to Section A.5 in Appendix A.

Univariate GLM

Separate univariate Generalized Linear Models (GLMs) were developed for each climate variable and extreme index, following the general methodology from McKee et al. [33], with adjustments specific to this study. The primary goal was to identify the best-fitting covariates—independent variables or predictors—that most effectively explain spillover events. Negative binomial GLMs were employed to account for overdispersion, and model fit was evaluated using criteria like *pseudo-*R^2 and the statistical significance of coefficients.

The process involved estimating the dispersion parameter α using the `NegativeBinomial` function from the `statsmodels` library and fitting the final model based on the Maximum Likelihood Estimation (MLE) method. Model parameters, including regression coefficients, were evaluated using *p-values* and *Wald tests*, with statistical significance denoted by: "$***$" for p-value< 0.001, "$**$" for p-value< 0.01, and "$*$" for p-value< 0.05. Coefficients were interpreted through the log link function by calculating change factors (i.e., exponentiating the coefficient). Predictions, along with 95% confidence intervals for the predicted mean counts, were generated to visualize the model's behavior.

Full details of the steps, code (see Listing A.1), and model evaluation criteria are provided in Appendix, Section A.5.1. From this point onward, it is assumed that all analyses are conducted for each spatial cluster, and this will not be explicitly stated in each step.

Pairwise Pearson's correlation

After the univariate GLMs analysis, variables with non-significant coefficients (p-value $>$ 0.05 and pseudo-$R^2 < 0.16$) were excluded. For the remaining variables, *pairwise Pearson's correlation* coefficients ρ were calculated using `pearsonr` from the `scipy` library to assess multicollinearity. Variables with $|\rho| > 0.90$ were further screened, and the one with a lower pseudo-R^2 or less significant coefficient in the univariate analysis was excluded. If these selection criteria produced similar results, the variable most relevant to the study was retained.

Scaling of the explanatory variables

After selecting the non-correlated explanatory variables, they were scaled using `MinMaxScaler` from `sklearn.preprocessing`. This function scales each feature to a range of 0 to 1 using the formula:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{5.7}$$

where X_{min} and X_{max} are the minimum and maximum values of the feature X. Scaling ensures that all features contribute equally, preventing larger-scaled features from dominating, thus improving model selection and performance.

GLM: Exhaustive feature selection

The next step involved a rigorous model selection process using *Akaike's Information Criterion, corrected for small sample sizes* (AICc) [115]. AIC is a widely used metric in model selection, balancing model fit and complexity by penalizing the number of parameters in the model. It is calculated as:

$$AIC = 2p - 2\,ln(\ell) \tag{5.8}$$

where p is the number of parameters and ℓ is the maximized log-likelihood of the model. However, for small sample sizes, AIC can lead to overfitting, so AICc introduces an additional penalty based on the number of observations n [116]:

$$AICc = AIC + \frac{2p(p+1)}{n-p-1} \tag{5.9}$$

This adjustment ensures a more accurate model selection process when the sample size is limited, preventing models from becoming overly complex without a corresponding improvement in fit.

The variable selection process began with a baseline model containing only the intercept. From there, the *exhaustive feature selection* algorithm systematically evaluated all possible combinations of scaled climate variables. It started with single-variable models, adding one variable at a time and assessing the collective impact of additional variables at each step. This incremental approach allows for testing interactions and capturing the complex relationships between climate variables and spillover events.

For each combination of variables, the model's log-likelihood was optimized using the `L-BFGS-B` algorithm, and with the optimized α_{opt}, a negative binomial GLM was fitted using `statsmodel.api` in Python. The AICc was computed at each iteration, and the model with the lowest AICc value was selected as the best-fitting model. Additionally, models with a $\Delta AICc$ of less than 2 from the top model were considered statistically indistinguishable, thus equally plausible, and part of the set of best models, as they offer a similar balance between simplicity and explanatory power.

Full details on the selection process and code implementation are provided in Appendix A (Section A.5.1, see Listing A.2).

Inclusion of interaction factors

To explore potential improvements, interaction terms (i.e., products of standardized variables) were introduced to examine how the effect of one variable on the response might depend on another, aiming to enhance the model's explanatory power.

For the top models (those with ΔAICc < 2 from the top one), interaction terms were systematically added. Specifically, for models with with one variable ($k = 1$), no interaction terms were added and the original model was retained. For models with two variables ($k = 2$), the interaction term $X_{1,st} \cdot X_{2,st}$ was computed and added after standardizing X_1 and X_2 original variables using `MinMaxScaler`, generating three candidate models. For models with three variables ($k = 3$), all pairwise interactions ($X_1 \cdot X_2$, $X_1 \cdot X_3$ and $X_2 \cdot X_3$) were tested, yielding nine candidate models. For further details regarding the procedure for incorporating these terms and the corresponding candidate models tested, please refer to Sections A.5.3 and A.5.4, as well as Tables A.6 and A.8.

Each interaction-enhanced model was then evaluated using AICc, *pseudo*-R^2, and the statistical significance of coefficients. This process refined the final selection of models, identifying those with $\Delta AICc < 2$ that most accurately explained the yearly variation in spillover events.

Model evaluation

To evaluate the model's validity and robustness, several diagnostic metrics were employed, focusing on residual analysis and influence diagnostics. These are essential for assessing model assumptions, detecting outliers, and identifying influential data points.

The types of residuals used in model diagnostics are as follows [114], [117]:

- **Raw residuals** These are the simple differences between observed values of annual spillover counts (y_i) and predicted values ($\hat{\mu}_i$) and are useful for identifying outliers or inconsistencies in the model. They are computed as:

$$r_i = y_i - \hat{\mu}_i, \quad i = 1, \ldots, n \tag{5.10}$$

- **Pearson residuals** Standardized residuals calculated by dividing raw residuals by the estimated standard deviation of the predicted values. They help assess whether residual variance is constant (homoscedasticity). They are computed as:

$$r_i^P = \frac{r_i}{\sqrt{\sigma^2(\hat{\mu}_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \alpha\hat{\mu}_i^2}} \tag{5.11}$$

- **Deviance residuals** Specific to GLMs, deviance residuals quantify the contribution of each observation to the model's deviance, providing insight into model fit. They are computed as:

$$r_i^D = \mathrm{sgn}(y_i - \hat{\mu}_i)\sqrt{D_i}, \quad i = 1, \ldots, n \tag{5.12}$$

  where D_i represents the contribution of a single observation to the deviance (for a detailed discussion on deviance, refer to Section A.5.2).

- **Standardised Pearson residuals** These are Pearson residuals adjusted by their leverage values, making them more suitable for identifying outliers and influential points. They are calculated as:

$$r_{i,st}^P = \frac{r_i^P}{\sqrt{1 - h_i}} \tag{5.13}$$

  where h_i is the leverage of the i^{th} observation.

In addition to residuals, other key diagnostic tools employed for linear models extend straightforwardly to GLMs and are employed in this study, including:

- **Hat-values (or Leverages, h_i):** These quantify how much an observation influences its predicted value. Higher leverage indicates that a point is far from the mean of predictor values and may significantly influence the model's fit. In GLMs, hat-values are obtained from the final iteration of the Iteratively Weighted Least Squares (IWLS) algorithm.

- **Cook's Distance (D_i):** This metric assesses the influence of each data point on the model's estimates by evaluating how much the predicted values change when a specific observation is removed. Points with large Cook's distances are flagged as influential, meaning they have a considerable impact on the model.

For further details on the calculation and theoretical background of these diagnostics, see Section A.5.2 in Appendix A.

Model diagnostics and assumptions verification To ensure that the optimal models meet key assumptions and to detect potential issues, the following diagnostics were performed:

1. **Linearity check:** A plot of predicted vs. observed values was used to confirm linearity, ensuring that points lie close to the diagonal. Deviations from this line may suggest non-linearity, requiring further model adjustments or transformations of the variables.

2. **Independence of observations:** To assess independence, plotting the deviance residuals in sequential order is an effective method. If the observations are truly independent, this plot should not reveal any patterns or systematic structures, such as trends, cycles or clusters. This visual inspection was complemented by the *Durbin-Watson test*, applied to the deviance residuals to detect serial correlation. Although typically associated with linear regression models, this test can also be applied to GLMs, including those with a negative binomial distribution. Values between 1.5 and 2.5 suggest no autocorrelation.

3. **Heteroskedasticity check:** Pearson residuals were plotted against fitted values and observed values to assess homoscedasticity. Although GLMs do not strictly require normally distributed errors, the expectation that 95% of the standardized residuals fall within ±2 standard deviations serves as a rough guide for detecting significant deviations from homoskedasticity [118].

4. **Deviance residual analysis:** Although deviance residuals in GLMs like negative binomial models do not require normality, they often behave similarly to ordinary residuals in standard linear regression models [119], providing useful insights. Q-Q plots are used to visually assess how closely residuals align with a normal distribution. This is particularly helpful for identifying outliers, skewness, or model misspecification.

   To further enhance diagnostics, future efforts could incorporate *half-normal plots*, which compare absolute residual values to the truncated normal distribution, along with simulation envelopes to capture prediction intervals more accurately and identify areas where the model may require adjustments [120].

The *Shapiro-Wilk test* is also applied to assess the normality of deviance residuals, with W statistic values closer to 1 suggesting normality. Although normality is less critical in GLMs, this test can still highlight potential fit issues or the need for data transformation [119].

5. **Leverage and influence analysis:** As recommended by Hilbe [114], standardized Pearson residuals were plotted against leverage values, in the so called *influential plot*, where poor model fit is indicated by residuals falling outside the interval ±2 when leverage is high (i.e., $h_i > h_{cut-off}$) [114]. Notably, to enhance the traditional plot, the size of each point was scaled according to its corresponding *Cook's distance*, providing a quick visual cue of the influence each observation has on the model's coefficients. Additionally, an *index plot of Cook's distance* was generated for further analysis.

5.1.3 Results - Case A: single spatial cluster

The distribution of aggregated spillover data across the unified cluster, encompassing all districts within the Nipah belt (excluding the Comilla district), aligns well with a negative binomial model. This distribution effectively captures the observed overdispersion, with a mean of 5.62, variance of 30.25, and a dispersion coefficient (α_{opt}) of 1.035, confirming the appropriateness of this model for the 18-year period under study (refer to Figure 2.9).

Univariate GLM results

Using the identified negative binomial GLM and the optimal dispersion parameter α_{opt} obtained via Maximum Likelihood Estimation (MLE), separate univariate models were developed for each climate variable and extreme index. The results are summarized in Table 5.2.

As outlined in Section 5.1.1, both seasonal and monthly variables in the pre-winter periods were evaluated to determine the best temporal aggregation for explaining annual spillover variability. July's variables showed stronger explanatory power and statistical significance compared to the entire monsoon season. Indeed, moving from seasonal to monthly data improved both the pseudo-R^2 values and the significance of the β coefficient. Moreover, while monsoon precipitation was not statistically significant (p-value > 0.10), July precipitation was, justifying the focus on July's variables for further analyses as they offer a more precise and robust explanation of spillover variability. Focusing on July allows for the capture of key climatic influences that might otherwise be diluted when analyzing the entire monsoon season.

In contrast, neither November nor the post-monsoon season produced significant results, leading to their exclusion from further analysis.

Interestingly, while the variable "Percentage of winter cold days" was identified as the best-fitting variable in the reference article [33], it does not hold the same significance in the present study. Instead, the best covariates identified here are $T_{\text{avg,winter}}$, followed by $T_{\text{avg,july}}$, as shown in Figures 5.1a and 5.1b respectively.

Variables such as $T_{\text{max,july}}$, $T_{\text{range,july}}$, $Pr_{\text{cum,winter}}$, DD_{winter}, and DD_y, highlighted in red in Table 5.2, can be excluded from further analysis, as they show no statistically significant associations with yearly spillover events and do not meaningfully contribute to explaining

Period		**2000-2018**	
Season or month	Statistics Variable	**Pseudo-R^2**	**β,p-value**
Monsoon	$T_{avg,monsoon}$	0.15	1.339
	$Pr_{monsoon}$	0.10	-0.002
July	$T_{avg,july}$	0.40	1.930^{**}
	$T_{min,july}$	0.26	2.242^{*}
	$T_{max,july}$	0.10	1.061
	$T_{range,july}$	0.00	-0.240
	Pr_{july}	0.24	-0.171^{*}
Post-monsoon	Pr_{ON}	0.03	-0.176
November	Pr_{nov}	0.00	-0.098
Winter	$T_{avg,winter}$	0.47	-1.116^{**}
	$T_{min,winter}$	0.25	-0.851^{*}
	$Pr_{cum,winter}$	0.03	0.013
Indexes	DD_{winter}	0.00	0.025
	CD_{winter}	0.21	0.096^{*}
	DCD_{winter}	0.20	0.091^{*}
	DD_y	0.16	0.154
	CD_y	0.23	0.174^{*}
	DCD_y	0.27	0.206^{*}

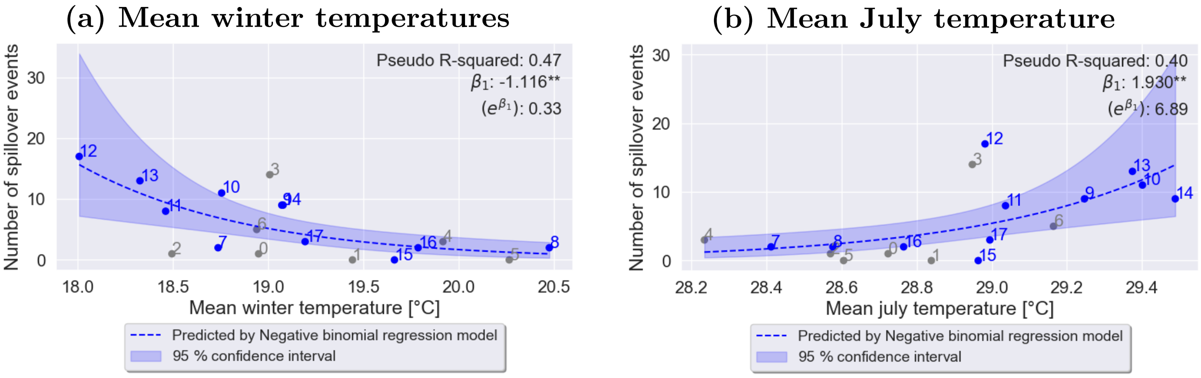
Table 5.2: Summary of univariate GLM analyses. Coefficients highlighted in red indicate p-values greater than 0.05, suggesting non-significance. Asterisks denote significance levels: * for p-values less than 0.05, ** for p-values less than 0.01, and *** for p-values less than 0.001. Coefficients highlighted in blue represent variables excluded due to multicollinearity issues. Periods and variables included in further analysis are shown in black.

the observed variability, making them irrelevant for further modeling.
In addition, two more variables (highlighted in blue in Table 5.2) were excluded after evaluating multicollinearity issues. Pairwise Pearson's correlation ρ revealed high correlations ($|\rho| > 0.9$) between $T_{avg,winter}$ and $T_{min,winter}$ ($\rho = 0.96^{***}$), and between CD_y and DCD_y with $\rho = 0.91^{***}$ (see Figure A.5). Based on the selection criteria discussed in Section 5.1.2, $T_{min,winter}$ and CD_y were excluded, as their univariate GLMs yielded lower *pseudo-*R^2 values compared to their counterparts.

To ensure comparability among the remaining eight explanatory variables, they were standardized to a common scale, ranging between 0 and 1. This standardization eliminates the influence of differing magnitudes, allowing for meaningful comparisons and enhancing the interpretability of their relative impacts on the response variable in subsequent analyses.

Model selection results

The *exhaustive feature selection* process produced a ranked list of GLMs, with the top five exhibiting $\Delta AICc < 2$ (see Table A.5). In addition to these models, others were tested


(a) Mean winter temperatures

(b) Mean July temperature

Figure 5.1: Variation in the yearly number of Nipah spillover events explained by: (a) mean winter temperatures, and (b) mean July temperature. Observed values are represented as dots, predicted values by the negative binomial regression model are shown as dashed lines, and the 95% confidence intervals of the predicted values are indicated by the shaded blue areas.

by introducing interaction terms, following the procedure outlined in Section A.5.3 and detailed in Table A.6. These interaction terms were added to assess whether combining variables could reveal synergies that enhance the models' explanatory power. The final full set of models, including those with and without interaction terms, is summarized in Table 5.3, which reports key statistics such as $\Delta AICc$ values (up to 4), Pseudo-R^2, optimal dispersion parameters α_{opt}, and regression coefficients β, with clear indications of significant and non-significant results.

5.1.4 Discussion - Case A

Comparison between reference article and models without interaction terms
When comparing the results of the current analysis to those presented in the reference article [33], several key improvements are evident (see Table 5.4, and for further details refer to Table A.5). The Akaike Information Criterion corrected for small sample sizes (AICc) shows a substantial reduction of 34%, decreasing from 100.3 in the reference model (the only one with $\Delta AICc < 2$) to a range of 95.6 to 97.4 in the five models identified in this study with $\Delta AICc < 2$. This indicates a better balance between model fit and complexity, suggesting that the current models offer a more efficient explanation of the spillover data.

Additionally, the *pseudo*-R^2 value, representing the proportion of variance explained by the model, significantly improved from 0.53 in the reference model to a range of 0.65 to 0.84 in four out of the top five models. This demonstrates stronger explanatory power and shows that the selected variables in this study more effectively capture patterns associated with spillover events.

Furthermore, a key distinction lies in the contributing predictors: while the reference article focused solely on winter covariates, the present models incorporate both winter and July (monsoonal) variables, highlighting their combined importance in explaining annual spillover variability. The variable DCD_y, though calculated on an annual basis, mainly reflects extended winter conditions (November-March, refer to Figure 3.7a), further emphasizing the broader temporal scope of the current models in capturing climatic drivers of spillover events.

Model name	**Model expression**	**AICc**	$\boldsymbol{Pseudo\text{-}R^2}$	α_{opt}	β	$\Delta AICc$
M_1^*	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,july} + \beta_3 (T_{avg,winter} \cdot T_{avg,july})$	92.85	0.79	0.211	$\beta_0 = 0.446$ $\beta_1 = 3.568^{***}$ $\beta_3 = -4.535^{***}$	0.0
M_2^*	$log(\mu) \sim \beta_0 + \beta_1 DCD_y + \beta_2 T_{avg,july} + \beta_3 (Pr_{july} \cdot DCD_y)$	93.46	0.90	0.1	$\beta_0 = -2.457^{**}$ $\beta_1 = 1.451^{*}$ $\beta_2 = 3.848^{***}$ $\beta_3 = 4.960^{**}$	0.61
$\boldsymbol{M_{ref}}$	$\boldsymbol{log(\mu) \sim \beta_0 + \beta_1 T_{avg,july} + \beta_2 T_{avg,winter}}$	**95.60**	**0.75**	**0.225**	$\boldsymbol{\beta_0 = 1.423^{*}}$ $\boldsymbol{\beta_1 = 1.690^{**}}$ $\boldsymbol{\beta_2 = -2.056^{**}}$	**2.75**
$M_5^{(2)*}$	$log(\mu) \sim \beta_0 + \beta_2 T_{avg,july} + \beta_3 (T_{avg,july} \cdot CD_{winter})$	95.64	0.68	0.31	$\beta_0 = 0.107$ $\beta_2 = 0.685$ $\beta_3 = 2.890^{**}$	2.79
$M_5^{(1)*}$	$log(\mu) \sim \beta_0 + \beta_1 CD_{winter} + \beta_3 (T_{avg,july} \cdot CD_{winter})$	95.69	0.66	0.34	$\beta_0 = 0.466$ $\beta_1 = -0.635$ $\beta_3 = 4.065^{***}$	2.84
$M_1^{(3)}$	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,winter} + \beta_2 T_{avg,july} + \beta_3 (T_{avg,winter} \cdot T_{avg,july})$	96.15	0.79	0.22	$\beta_0 = 0.218$ $\beta_1 = 0.437$ $\beta_2 = 3.932^{**}$ $\beta_3 = -5.304$	3.30
M_2	$log(\mu) \sim \beta_0 + \beta_1 DCD_y + \beta_2 T_{avg,july} + \beta_3 Pr_{july}$	96.27	0.84	0.15	$\beta_0 = -3.698^{*}$ $\beta_1 = 3.072^{***}$ $\beta_2 = 4.061^{***}$ $\beta_3 = 2.944^{***}$	3.42
M_3	$log(\mu) \sim \beta_0 + \beta_1 DCD_y + \beta_2 T_{avg,july}$	96.56	0.67	0.31	$\beta_0 = -0.657$ $\beta_1 = 1.981^{**}$ $\beta_2 = 2.113^{**}$	3.71

Table 5.3: Summary of the final full set of models, including those with interaction terms, obtained by testing selected candidate negative binomial GLMs with specific interaction terms. For each model, the table provides: model expressions, $\Delta AICc$ values (in ascending order and relative to the reference model), Pseudo-R^2, the optimal dispersion parameter obtained by minimizing the negative log-likelihood function, and regression coefficients (β). Coefficients in red indicate non-significant results (p-value $>$ 0.05), while black coefficients with asterisks denote significant levels: * (p-value $<$ 0.05), ** (p-value $<$ 0.01), and *** (p-value $<$ 0.001).

The inclusion of July variables confirms the previously hypothesized link between monsoon climate conditions and the frequency of spillover events in the following winter season, indicating a delayed effect on bat behavior and/or their immune system. These findings underscore the effectiveness of the exhaustive feature selection process and the necessity of considering variables beyond the winter season. **Notably, to the best of the author's knowledge, the observed association between monsoon climate conditions and Nipah virus spillover events in Bangladesh had not been documented before this study**. This discovery parallels a similar pattern identified for the Hendra virus in Australia, another *Paramyxovirus*, marking a novel contribution to the understanding of Nipah virus dynamics.

Comparison among models obtained in the current study The top two models reported in Table 5.3 have a $\Delta AICc < 2$, making them statistically indistinguishable and equally plausible as the best-fitting models in this analysis. Both include interaction terms, highlighting the importance of interactions between monsoonal and winter variables in explaining spillover variability.

The third model, with a slightly higher $\Delta AICc$ of 2.75, remains noteworthy due to its second-highest *pseudo-*R^2 value, indicating strong explanatory power despite lacking interaction terms. As the best model without interactions, it provides a useful benchmark

Statistics	**Best models** current analysis	**Best model** reference article [33]
Models with $\Delta AICc < 2$	Five models $M_j^{(0)}$ with $AICc \in [95.6 - 97.4]$	Only one model with $AICc = 100.3$
$\boldsymbol{Pseudo\text{-}R^2}$	$0.47 - 0.84$	0.53
p-values	$0.001\,(^{***}), 0.01\,(^{**}), 0.05\,(^{*})$	0.001
Explanatory variables	1) Winter variables: $T_{avg,winter}, DCD_y, CD_{winter}$ 2) July (monsoon) variables: $T_{avg,july}, Pr_{july}$	Percentage of winter cold days

Table 5.4: Comparison between the best-fitting models derived from the current analysis and the reference article [33].

for evaluating the added value of interaction terms.

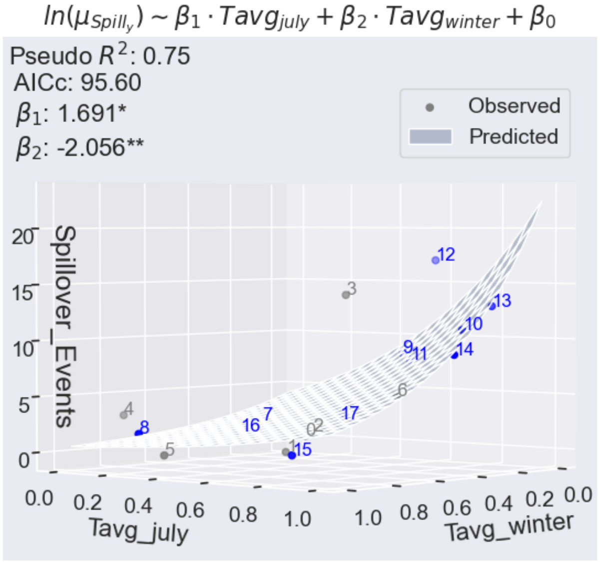
In this section, these three models will be examined in greater detail to evaluate their performance and the influence of interaction terms on model fit and accuracy.

Top model without interaction terms The top model without interaction terms, M_{ref}, (Figure 5.2), shows that lower winter temperatures, combined with higher July temperatures, are associated with an increase in spillover events from 2001 to 2018 [33]. Specifically, colder winters from 2009/2010 to 2014/2015 (years labelled as 9 to 14) correlate with more frequent spillovers, while fewer events occurred in the warmer winters of 2008/2009 (i.e., year 8) and 2015/2016 to 2017/2018 (i.e., year 15-17) (refer to Figure 5.2b) [33]. After 2007, higher July temperatures seem to have a stronger influence on winter spillover, suggesting a delayed effect of monsoonal conditions on NiV transmission. Furthermore, an analysis of the model coefficients (reported in Figure 5.2) shows that July temperatures have a greater predictive weight than winter temperatures.
As discussed throughout this thesis, the relationship between monsoonal, winter climate conditions, and the incidence of spillover events, can be explained by a combination of ecological and behavioral dynamics in both humans and bats. Climate, in particular, plays a key role in shaping these interactions as previously discussed in Section 3.1.5.

Figure 5.2a illustrates the observed versus predicted spillover events from the M_{ref} model, which includes mean winter and July temperatures. The predicted values align closely with most observations, indicating a strong overall fit. However, certain years, such as 2003/2004 and 2012/2013 (labeled as 3 and 12, respectively), show higher observed spillovers than predicted, suggesting the presence of unmodeled factors. To assess whether these deviations arise from influential data points or model limitations, diagnostic tools such as *Cook's distance* and *leverage plots* were employed. Overall, the model meets the key assumptions of linearity, independence of observations, and homoscedasticity of residuals (see Appendix for detailed diagnostic plots in Figure A.6 and related comments). Nevertheless, specific years—namely 3, 4, and 12—display diagnostic deviations that warrant further attention.

Referring to Figure A.7, years 3 and 4 exhibit high standardized residuals, indicating underpredictions of spillover events, but their low leverage suggests they do not distort the model. Despite this, their Cook's distance values indicate a significant impact on the regression coefficients. Year 12, while not showing problematic residuals, has high leverage, meaning it heavily influences the model due to extreme predictor values, likely driven by

(a) Top model without interaction terms



(b) Association between explanatory variables and yearly spillover events

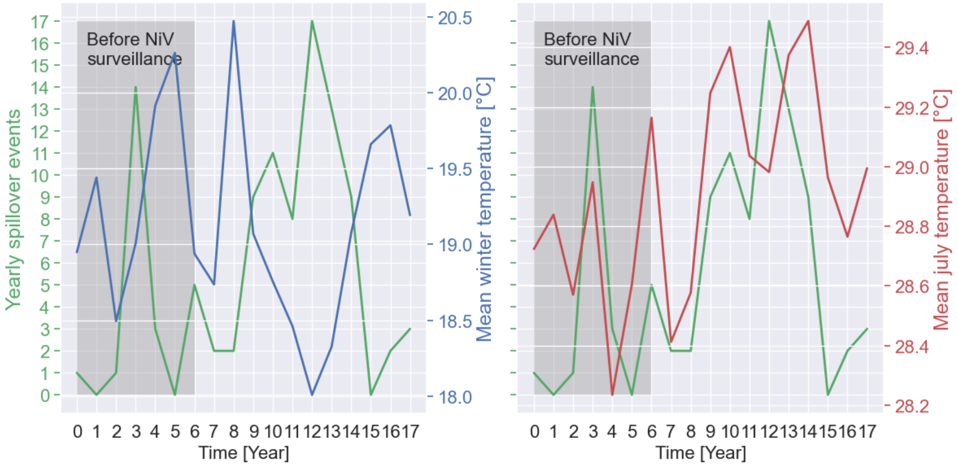


Figure 5.2: Top GLM without interaction terms and temporal trends: (a) The model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo-*R^2 and coefficient estimates along with their statistical significance. Each data point corresponds to a specific year, labeled from 0 to 17, corresponding to the annual periods from 2000/2001 to 2017/2018 (where the year spans from May to April). The model's predicted values are depicted by the smooth surface. (b) Annual time series of the considered explanatory variables (i.e. mean July and winter temperatures), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh.

a significant winter temperature anomaly (refers to Figure 3.4). For further details, refer to Section A.5.3 in Appendix A; however, these findings suggest that while the model is robust, it could benefit from refinements, such as incorporating interaction terms or adjustments to better capture the impact of rare or severe weather events. Therefore, the following analysis focuses on the two top models that integrate these interactions.

Two top models with interaction terms The most striking comparison in the current analysis is between the top model without interaction terms M_{ref} and the improved model with interaction terms M_1^* (refer to Table 5.3). M_1^* demonstrates a substantial improvement, with a reduction in AICc by 2.75 and an increase in *pseudo*-R^2 by 0.04, signaling a stronger fit and explanatory power. This improvement is particularly evident when considering the reduction in outliers: progressing from M_{ref} to M_1^*, via M_2^*, the number of outliers in terms of the response variables drops from two to one, and the number of influential points decreases from three to one. These changes are highlighted in the *Influence* and *Cook's distance* plots (see Figure A.7 for M_{ref} and Figure 5.3 for M_1^*). Moreover, the improvement in the model statistics highlights the significant boost in predictive power provided by the interaction terms, particularly in M_1^*. As shown in Figure 5.5, which compares the observed spillover events with predictions from the top three models, M_1^* (blue line) better captures the variability in spillover events, especially in high-incident years like 2012/2013. While M_{ref}(red line) underestimates these events, M_1^* corrects these discrepancies, offering a more accurate representation. The difference in how the M_1^* predictors track spillover events over time is further demonstrated by comparing Figures 5.2b and 5.4b. The combined effects of July and winter temperatures in M_1^* offer deeper insights into spillover patterns, especially during years with significant fluctuations, such as 12 and 15. The interaction between rising July temperatures and decreasing winter temperatures is linked to a higher incidence of spillover events, underscoring the complex interplay between these climatic variables. Climate anomaly analyses (Figure 3.5) suggest that significant negative winter anomalies in year 12 likely contributed to the surge in events, while the combination of positive winter and negative July temperature anomalies in year 15 led to a decline. Overall, M_1^* captures these dynamics more effectively than the simpler M_{ref}, demonstrating a more nuanced understanding of the drivers behind spillover events.

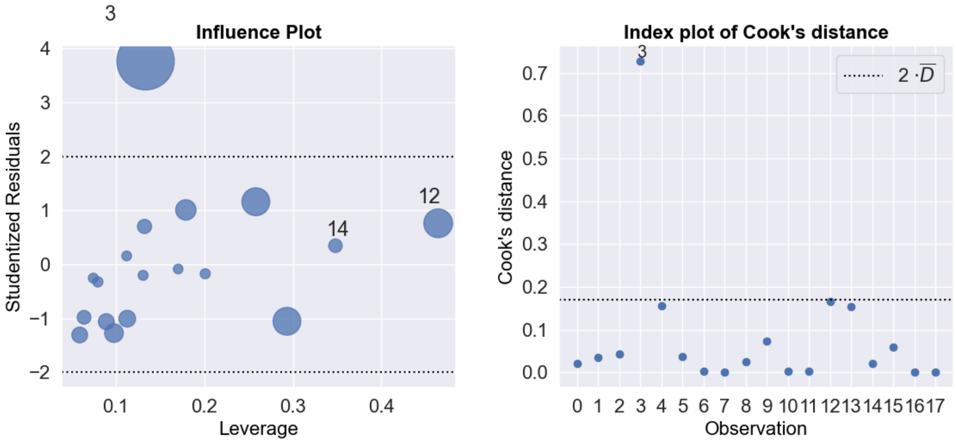
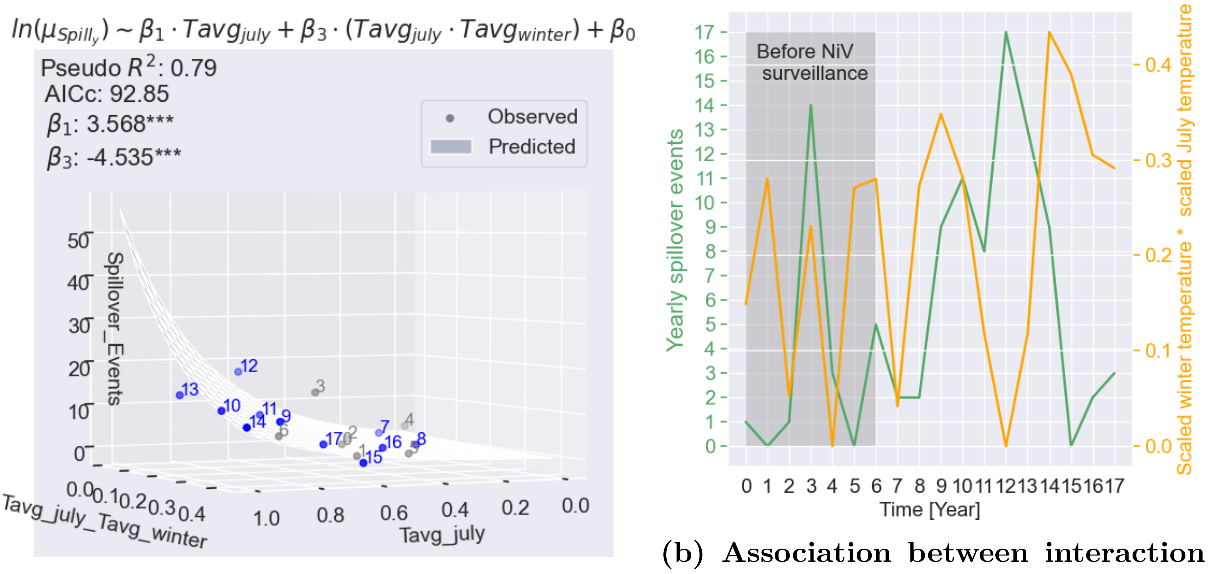


Figure 5.3: Diagnostic analysis of the improved M_1^* negative binomial GLM: (Left) enhanced influence plot with scatter point sizes proportional to Cook's distance, and (Right) index plot of Cook's distance, both with indicated thresholds for identifying outliers and influential points, respectively.

Despite the improvements in M_1^*, some challenges remain. Given that multicollinearity is not an issue and the variables have been standardised to address differences in scale, the high sensitivity of the model to changes in $T_{avg,july}$, as seen in the large value of exponentiated β_1 and the small value of the interaction term β_3, suggests that more complex nonlinear relationships may exist. This is further supported by residual and



(a) Best improved model

(b) Association between interaction term and yearly spillover events

Figure 5.4: Top GLM with interaction terms and temporal trends: (a) The model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo*-R^2 and coefficient estimates along with their statistical significance. Each data point corresponds to a specific year, labeled from 0 to 17, corresponding to the annual periods from 2000/2001 to 2017/2018 (where the year spans from May to April). The model's predicted values are depicted by the smooth surface. (b) Annual time series of the interaction term (i.e. $T_{avg,july} \cdot T_{avg,winter}$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh.

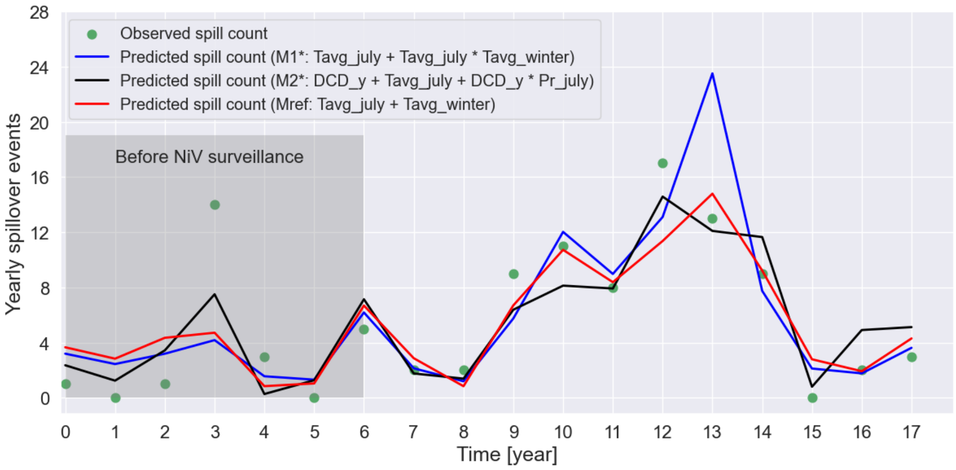


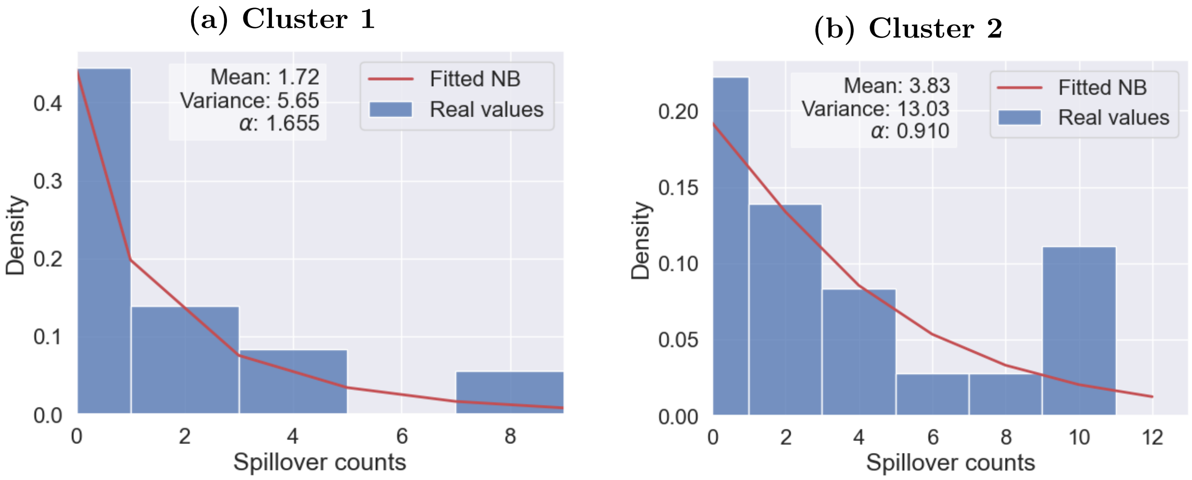
Figure 5.5: Comparison between observed annual spillover counts (green dots) and predictions from the top three models.

influence plot analyses, which indicate that the model does not fully explain all data points, particularly for the third year.

To address these limitations, future analyses could explore more flexible models, such as Generalized Additive Models (GAMs), to better capture the nonlinear relationships between predictors and spillover events.

5.1.5 Results - Case B: two spatial clusters

For each cluster, spatially aggregated spillover data were analyzed, and the statistically significant dispersion parameter α, estimated through Maximum Likelihood Estimation (MLE), confirmed that the variance exceeded the mean, justifying the use of a negative binomial (NB2) distribution. Figures 5.6a and 5.6b display the histograms and fitted NB2 distributions for the two clusters, along with key statistics such as the mean, variance, and dispersion parameter α.


(a) Cluster 1

(b) Cluster 2

Figure 5.6: Histogram of annual (May to April) spillover events with fitted Negative Binomial distribution (red line). Key statistics, including mean, variance, and dispersion parameter α are reported for: (a) Cluster 1, and (b) Cluster 2.

Univariate GLM results

The results of the univariate GLM analysis for both clusters are summarized in Table 5.5. As in Case A, focusing on July's climatic variables instead of the entire monsoon season led to stronger results in Cluster 2, with higher pseudo-R^2 values and improved statistical significance of the β coefficients. In particular, July's temperatures ($T_{min,july}, T_{avg,july}$) and precipitation ($Pr_{july}$) showed significant correlations with spillover events, reinforcing the decision to prioritize July over the broader monsoon period.

In contrast, in Cluster 1, neither monsoonal nor July variables were statistically significant, whereas several winter-related variables were closely associated with spillover events. This distinction highlights the differing climatic drivers in each region. Cluster 1 appears more influenced by winter conditions, while Cluster 2 is shaped by both July's monsoonal and winter variables. These results align with those of Chapter 4, which emphasized harsh winters as critical in Cluster 1, while hot, dry monsoon conditions played a larger role in Cluster 2. This confirms the varying climatic patterns driving NiV transmission dynamics across the two clusters.

Additionally, variables marked in red in Table 5.5 were excluded due to their lack of statistical significance. In Cluster 2, the CD_y variable (marked in blue) was removed to address multicollinearity issues with DCD_y (refer to Figures A.8 and A.9, to see the correlation matrices). Although both variables yielded similar statistics, DCD_y was

Period 2000-2018		**Cluster 1**		**Cluster 2**	
Season or month	Statistics / Variable	**Pseudo-R^2**	**β,p-value**	**Pseudo-R^2**	**β,p-value**
Monsoon	$T_{avg,monsoon}$	0.05	1.140	0.17	1.307
	$Pr_{monsoon}$	0.02	-0.001	0.06	-0.002
July	$T_{avg,july}$	0.13	1.555	0.58	2.390^{***}
	$T_{min,july}$	0.12	2.127	0.31	2.546^{**}
	$T_{max,july}$	0.08	1.178	0.09	1.163
	$T_{range,july}$	0.01	0.729	0.03	-0.999
	Pr_{july}	0.03	-0.085	0.57	-0.233^{***}
Post-monsoon	Pr_{ON}	0.03	-0.182	0.06	-0.246
November	Pr_{nov}	0.00	-0.098	0.00	-0.015
Winter	$T_{avg,winter}$	0.58	-1.792^{***}	0.32	-0.886^{**}
	$T_{min,winter}$	0.44	-1.547^{**}	0.11	-0.565
	$Pr_{cum,winter}$	0.01	0.010	0.04	0.011
Indexes	DD_{winter}	0.01	0.052	0.01	-0.038
	CD_{winter}	0.19	0.167	0.12	0.054
	DCD_{winter}	0.22	0.176^{*}	0.09	0.048
	DD_y	0.03	0.084	0.24	0.192^{*}
	CD_y	0.12	0.185	0.26	0.163^{*}
	DCD_y	0.16	0.218^{*}	0.26	0.176^{*}

Table 5.5: Summary of univariate GLM analyses for different clusters over the reference period 2000-2018. Coefficients highlighted in red indicate p-values greater than 0.05, suggesting non-significance. Asterisks denote significance levels: * for p-values less than 0.05, ** for p-values less than 0.01, and *** for p-values less than 0.001. Coefficients highlighted in blue represent variables excluded due to multicollinearity issues. Periods and variables included in further analysis are shown in black.

retained because it better represents the xylem refilling phenomenon, which is closely associated with increased DPS production (refer to Section 1.6.3).

To ensure comparability and eliminate the effects of differing units and magnitudes, the remaining variables in each cluster were standardized to a 0-1 scale, enhancing the interpretability of their effects on the response variable.

Model selection results The *exhaustive feature selection* process for Case B produced a ranked list of GLMs for each cluster based on their AICc values (see Table A.7 for Cluster 1 and 2).

For Cluster 1, only one model with a $\Delta AICc < 2$ was identified ($M_{ref,C1}$), comprising a single explanatory variable. As a result, no interaction terms were tested, confirming $M_{ref,C1}$ as the reference model. In contrast, for Cluster 2, additional models incorporating interaction terms were tested (see Table A.8 in Section A.5.4) to explore potential synergies that might enhance the models' explanatory power.

The final full set of models for both clusters, including those with and without interaction terms, is summarized in Table 5.6. This table report key statistics such as $\Delta AICc$ values (up to 2), Pseudo-R^2, optimal dispersion parameters α_{opt}, and regression coefficients β, with clear indications of significant and non-significant results.

Cluster C1						
Model name	**Model expression**	**AICc**	***Pseudo-R^2***	α_{opt}	β	$\Delta AICc$
$\boldsymbol{M_{ref,C1}}$	$\boldsymbol{log(\mu) \sim \beta_0 + \beta_1 T_{avg,winter}}$	**59.07**	**0.58**	**0.325**	$\boldsymbol{\beta_0 = 1.765^{***}}$ $\boldsymbol{\beta_1 = -4.290^{***}}$	**0.0**
Cluster C2						
Model name	**Model expression**	**AICc**	***Pseudo-R^2***	α_{opt}	β	$\Delta AICc$
$\boldsymbol{M_{ref,C2}}$	$\boldsymbol{log(\mu) \sim \beta_0 + \beta_1 T_{avg,july} + \beta_2 DD_y}$	**80.57**	**0.77**	**0.132**	$\boldsymbol{\beta_0 = -0.862}$ $\boldsymbol{\beta_1 = 2.174^{***}}$ $\boldsymbol{\beta_2 = 1.654^{*}}$	**0.0**
$M^{*}_{3,C2}$	$log(\mu) \sim \beta_0 + \beta_2 DD_y + \beta_3(DD_y \cdot Pr_{july})$	80.77	0.80	0.01	$\beta_0 = 0.329$ $\beta_2 = 2.934^{***}$ $\beta_3 = -4.250^{***}$	0.2
$M_{2,C2}$	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,july} + \beta_2 DD_y + \beta_4(DD_y \cdot DCD_y)$	81.33	0.83	0.1	$\beta_0 = -1.259^{*}$ $\beta_1 = 1.880^{***}$ $\beta_2 = 1.495^{*}$ $\beta_4 = 1.178$	0.76
$M^{*}_{4,C2}$	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,winter} + \beta_3 DD_y + \beta_4(T_{avg,july} \cdot T_{avg,winter})$	81.94	0.82	0.1	$\beta_0 = 0.520$ $\beta_1 = -2.216^{**}$ $\beta_3 = 1.693^{*}$ $\beta_4 = 3.549^{***}$	1.37
$M^{**}_{4,C2}$	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,july} + \beta_3 DD_y + \beta_4(DD_y \cdot T_{avg,winter})$	82.07	0.82	0.1	$\beta_0 = -0.511$ $\beta_1 = 1.865^{***}$ $\beta_3 = 1.875^{**}$ $\beta_4 = -1.178$	1.5
$M_{5,C2}$	$log(\mu) \sim \beta_0 + \beta_1 T_{avg,july}$	82.49	0.58	0.297	$\beta_0 = -68.07^{***}$ $\beta_1 = 2.390^{***}$	1.92

Table 5.6: Summary of the final full set of models, including those with interaction terms, obtained by testing selected candidate negative binomial GLMs with specific interaction terms for both clusters. For each model, the table provides: model expressions, $\Delta AICc$ values (in ascending order and relative to the reference model), Pseudo-R^2, the optimal dispersion parameter obtained by minimizing the negative log-likelihood function, and regression coefficients (β). Coefficients in red indicate non-significant results (p-value > 0.10), while black coefficients with asterisks denote significant levels: * (p-value < 0.05), ** (p-value < 0.01), and *** (p-value < 0.001).

5.1.6 Discussion - Case B

Comparison among models obtained for different clusters As anticipated by the results in Chapter 4 and the univariate GLM analysis, the model selection process confirmed that different climatic factors drive the annual variability of spillover events across Cluster 1 and Cluster 2. Thus, the comparison of these models (see Table 5.7) underscores the significant role that varying climate conditions may play in shaping spillover dynamics across Bangladesh.

In Cluster 1, located in the northwestern region, lower mean winter temperatures are strongly linked to higher spillover occurrences, suggesting that harsh winter conditions in this already colder area (compared to Cluster 2) may lead to ecological or physiological changes in reservoir or recipient hosts, increasing the likelihood of spillover events. Notably, the GLM for Cluster 1 ($M_{ref,C1}$), which focuses solely on mean winter temperature, explains 83% of the annual variability in spillover events, underscoring the central role of winter temperatures as a key predictor in this region. The $\Delta AICc$ gap of 2.75 between the top two models confirms the superiority of $M_{ref,C1}$, while the second model, which includes only minimum winter temperatures, reinforces the critical role of winter-related variables in driving spillover events in this area.

In Cluster 2, which includes districts from the Rajshahi, Dhaka, and Khulna divisions in central Bangladesh, the top models identify a combination of higher July temperatures, lower July precipitation, an extended dry period from October to April, as well as lower winter temperatures as key drivers of annual spillover variation. This July-April period notably overlaps with the pregnancy and lactation phases of the NiV reservoir hosts in Bangladesh. The findings suggest that in this region, the warm monsoon season, followed by a dry pre-winter period, may increase nutritional stress or intensify human-bat interactions, leading to a higher incidence of spillover events. Indeed, these climatic factors likely influence food availability and trigger changes in host behavior and immune function during crucial reproductive phases, further elevating the risk of spillover, particularly in colder winters (refer to Discussion in Section 3.1.5).

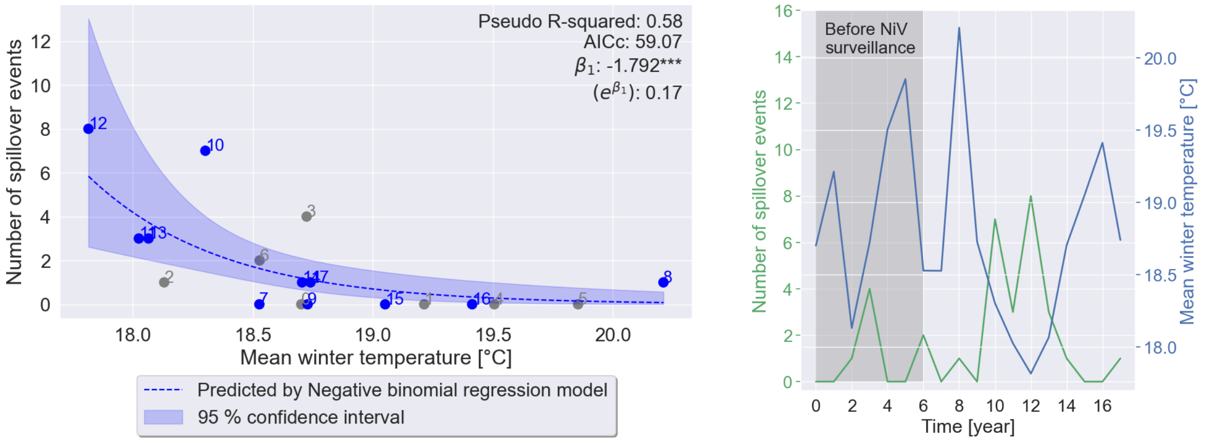
Statistics	**Cluster 1** Top model $M_{ref,C1}$	**Cluster 2** Top models
Models with $\Delta AICc < 2$	One model $M_j^{(0)}$ with $AICc = 59.07$	Five models $M_j^{(0)}$ with $AICc \in [80.57 - 82.49]$
Pseudo-R^2 **p-values**	0.58 $0.001\,(^{***})$	[0.58-0.83] $0.001\,(^{***}), 0.05\,(^{*})$
Explanatory variables	Winter average temperature	1) July (monsoonal) variables: $T_{avg,july}, Pr_{july}$ 2) Winter variables: $DCD_y, T_{avg,winter}, DD_y$

Table 5.7: Comparison between the best-fitting models derived for Cluster 1 and Cluster 2.

Focus: Best model for Cluster 1 The top-performing model for Cluster 1 shows that colder winter conditions during years 10 to 13, with temperatures approximately $2°C$ lower than in years 14 to 17 and 7 to 8, correlate with a higher incidence of spillover events. This trend is clearly illustrated in Figure 5.7, where the model's predictions align closely with observed data, though discrepancies arise for the years 2003/2004, 2008/2009, and 2010/2011 (years 3, 8, and 10). The spillover counts are underestimated in these years, prompting further investigation using diagnostic plots. Figures A.10 confirm that the model generally meets assumptions of linearity, independence, and homoscedasticity. However, years 3, 8, and 10 stand out as outliers, with standardized Pearson residuals exceeding +2, indicating inaccurate predictions for these years (see Figure 5.8). Despite being outliers, they show low leverage, meaning their predictor values are typical for the dataset and don't distort the overall model fit. However, removing these points would significantly impact the regression coefficients, as indicated by their high Cook's distance.

In summary, while these residual outliers don't undermine the model's validity, there is room for refinement. Incorporating interaction terms could potentially address these discrepancies and improve the model's accuracy.

Focus: Best models for Cluster 2 The top six models for Cluster 2 (see Table 5.6) have a $\Delta AICc < 2$, with four out of six including interaction terms between winter and July variables. This confirm that both periods significantly contribute to explaining annual spillover variability in this area.



(a) Best model for Cluster 1

(b) Cluster 1: Explanatory variable vs. yearly spillover events

Figure 5.7: Best negative binomial GLM and temporal trends for Cluster 1: (a) The best model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo*-R^2 and coefficient estimates along with their statistical significance. (b) Annual time series of the explanatory variable (i.e. $T_{avg,winter}$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh.

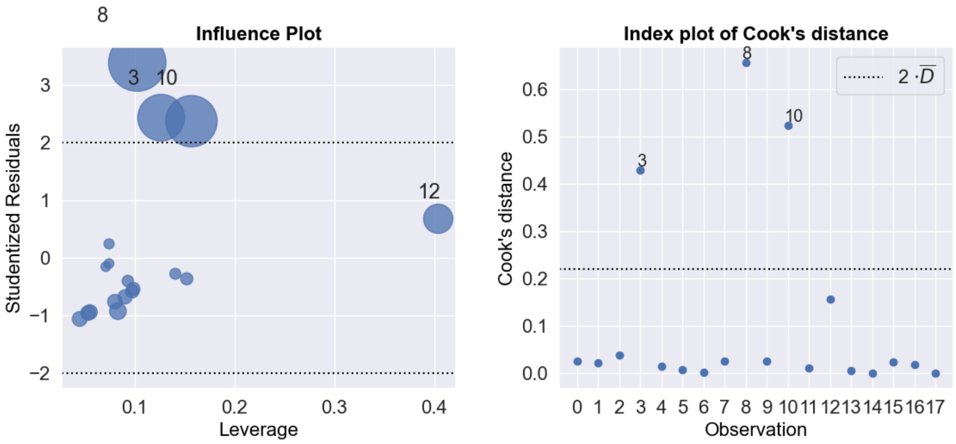
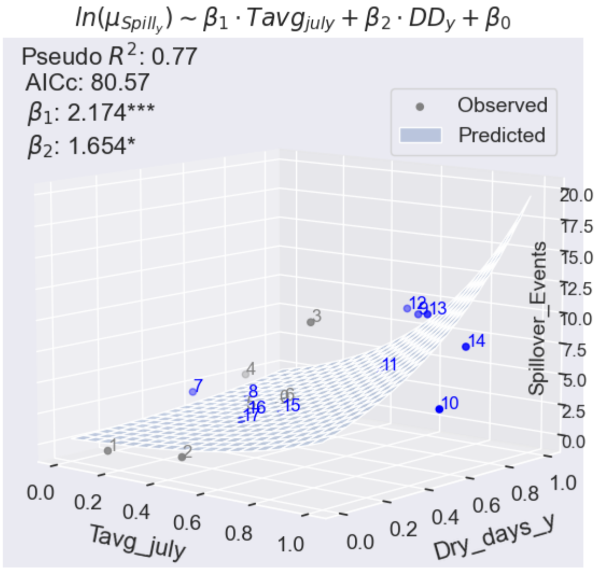


Figure 5.8: Cluster 1: Diagnostic analysis of the best-fit negative binomial GLM: enhanced influence plot (left) with scatter point sizes proportional to Cook's distance, and index plot of Cook's distance (right), both with indicated thresholds for identifying outliers and influential points, respectively.

Notably, the top model, which does not include interaction terms, highlights July temperatures and dry days on an annual scale as key predictors of spillover events. Peaks in spillover events (years 9, 12, 13, and 14) coincide with increased July temperatures and higher numbers of dry days (see Figure 5.9b), while lower July temperatures and fewer dry days correspond with reduced NiV occurrences in years 8, 10, 11, and 15. July temperatures, in particular, demonstrate greater predictive power, as reflected in the model coefficients (see Figure 5.9a). Despite the overall fit being strong, discrepancies arise in years 3 and 10, where the model underestimates and overestimates spillover events, respectively. Diagnostic plots (Figure A.11) confirm the model's assumptions of linearity and independence, but Pearson residuals for years 3 and 10 indicate potential outliers (see Figure 5.10). Year 3 shows significant underestimation with low leverage but

(a) Top model for cluster 2



(b) Cluster 2: Explanatory variables vs. yearly spillover events

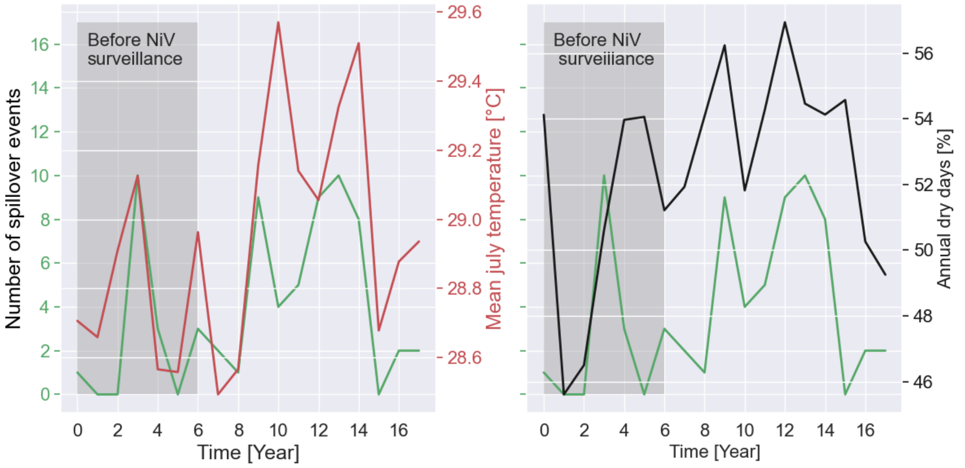


Figure 5.9: Best reference negative binomial GLM and temporal time series for Cluster 2: (a) The best model fitting the observed data, displaying its equation and key model statistics, including AICc, *Pseudo*-R^2 and coefficient estimates along with their statistical significance. (b) Annual time series of the explanatory variables included in the $M^{(0)}_{ref,C2}$ model (i.e. $T_{avg,july} \cdot DD_y$), alongside spillover events. The shaded area highlights the period before NiV surveillance was established in Bangladesh.

high Cook's distance, indicating its influence on the model's regression coefficients. Year 10, though less extreme, also warrants attention. In summary, while the model $M_{ref,C2}$ effectively capture key climatic drivers, particularly July temperatures, residual outliers and influential data points, such as those labeled 3 and 10, indicate areas for refinement. Notably, significant positive anomalies in July temperatures were recorded in these years, which the models struggle to fully account for.

This highlights both the strength and limitation of the simpler, non-interactive model. While generally robust, its inability to incorporate complex interactions may stem from either the data or the model's assumptions. Nonetheless, this model remain reliable overall, despite not addressing every anomaly.

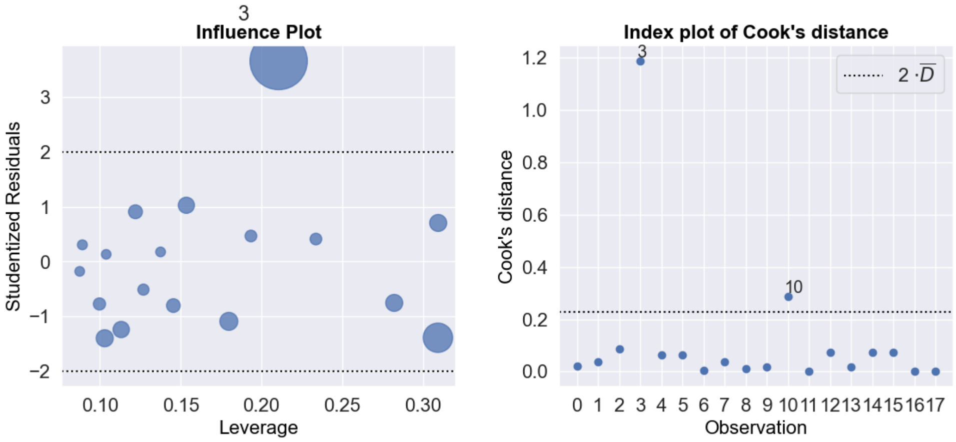


Figure 5.10: **Cluster 2:** Diagnostic analysis of the best-fit negative binomial GLM: enhanced influence plot (left) with scatter point sizes proportional to Cook's distance, and index plot of Cook's distance (right), both with indicated thresholds for identifying outliers and influential points, respectively.

To overcome these limitations, future research could explore alternative approaches, such as generalized additive models (GAMs), which offer greater flexibility in modeling relationships. Additionally, further analysis of influential data points from years 3, 8, and 10 in Cluster 1, and years 3 and 10 in Cluster 2, may reveal whether data transformation or other adjustments are required. Addressing these factors could improve the models' capacity to capture variability in spillover events, leading to more accurate predictions.

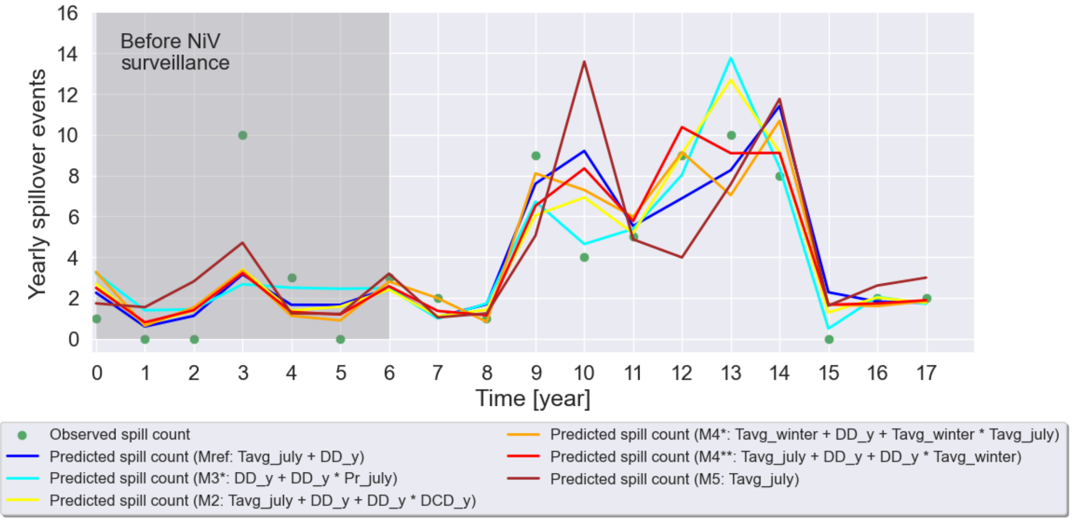


Figure 5.11: **Cluster 2:** Comparison between observed annual spillover counts (green dots) and predictions from the top six models.

5.2 Climate drivers of spillover events: current findings and future projections

The findings of this thesis have established a clear link between colder winters and warmer, drier monsoons in July with increased spillover events. These results also underscore the spatial heterogeneity of climate drivers, with harsh winter conditions being particularly critical in the northwestern region (Cluster 1), while in the central-western districts (Cluster 2), both hot, dry conditions in July and colder winters play a key role in driving spillovers. Given these findings, it is crucial to evaluate whether these conditions are more or less likely to persist in the future due to climate change, and whether they will be exacerbated or shift toward new patterns that could potentially alter spillover dynamics.

To explore this, the latest climate projections from a large ensemble of 27 Coupled Model Intercomparison Project phase 6 (CMIP6) Global Climate Models (GCMs) provide detailed seasonal and spatial insights into the future of rainfall and temperature in Bangladesh, based on a reference period of 1850–2014 [121], [122]. These models, which cover multiple Shared Socioeconomic Pathways (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5), predict significant changes in rainfall and temperature across different regions and timeframes: near-future (2015–2044), mid-future (2045–2074), and far-future (2075–2100) [121], [122].

Winter projections Projections show that both maximum (Tmax) and minimum (Tmin) temperatures are expected to rise across all future periods and SSPs, with the most significant warming occurring during the winter months (December–February), especially in the far future, and predominantly in the northern and coastal regions [121]. Tmin is projected to increase more rapidly than Tmax, with November showing the highest rise in Tmin (1.34–4.57 $°C$) and February seeing the largest rise in Tmax (1.31–3.88 $°C$) [121]. This reduction in the diurnal temperature range (DTR) could negatively impact crop yields, particularly in winter. Higher night temperatures (i.e., Tmin), combined with a projected 11.12% decrease in winter precipitation by the mid-future under SSP3-7.0, are expected to increase respiratory losses in crops and disrupt xylem refilling in plants like Date Palms [121]. This would likely reduce DPS production, a critical winter food source for bats (as discussed in Section 1.6.3), intensifying their nutritional stress and potentially increasing viral recrudescence and human-bat interactions during this critical period. Although warmer winters could reduce Nipah virus viability outside the host (see Section 1.6.3), food scarcity, exacerbated by decreased precipitation, could heighten the overall risk of spillover events. Additionally, the sharper temperature rise projected for the northwestern region (Cluster 1) could significantly alter the conditions that currently drive spillovers in this area [121].

Monsoon projections Projections suggest a significant increase in rainfall during the pre-monsoon (March-April), monsoon (June-September), and post-monsoon (October-November) seasons in Bangladesh [121]. Under SSP5-8.5, monsoon precipitation could rise by up to 20.8%, with post-monsoon rainfall increasing by as much as 41.98%, particularly in the northeastern and southeastern regions [121], [122]. Increased inter-annual variability is also projected, leading to more frequent stormy seasons and extreme events, such as typhoons, which could severely impact bat behavior and habitats [122]. Storm-related defoliation may reduce the concealment of roosting bats, making them more vulnerable to harsh weather, hunting, and predation (see Section 1.6.3), while flooding could further degrade habitats (restricting their available options), as bats tend to avoid flooded areas for roosting [41]. Notably, the districts of Rajbari and Faridpur in the central Nipah belt, which are particularly prone to flooding, have also reported some of the highest numbers of NiV spillover events during the 18-year analysed period, suggesting a link between flooding and heightened spillover risk in these areas.

In this thesis, drier July monsoons were suggested to contribute to food scarcity, potentially leading to nutritional stress in bats and delayed viral recrudescence with a six to nine-month lag, similar to observations of HeV in Australia [82], [83]. However, the changing climate patterns, characterized by wetter conditions, may led to similar outcomes through different mechanisms in the future. Indeed, also excess rainfall could lead to crop failures, particularly for crops sensitive to waterlogging, and exacerbate flash floods in already vulnerable low-lying areas. These disruptions to food resources could similarly intensify stress in bats, driving them closer to human environments in search of alternative food sources. This increased human-bat interaction could heighten the risk of spillover events, similar to those observed during drier conditions, though driven by different climate stressors.

Monsoon temperatures are also expected to rise significantly, especially in the northwestern and west-central regions, where increases could exceed 3.8 $^{\circ}C$ by the far future under SSP5-8.5 [122]. This is particularly concerning for Cluster 2, where higher July temperatures have already been linked to increased spillover events. Rising temperatures may introduce thermoregulatory stress for bats, as they struggle to cope with extreme heat (see Section 1.6.3), and combined with habitat disruptions and food scarcity, could further drive viral shedding and spillover events.

In summary, the projected warmer and wetter monsoons under future climate scenarios mark a shift from the drier July conditions, while simultaneously intensifying the warmer trends, currently associated with spillover risks. Although the specific triggers may change, new stressors—such as heat stress, flooding, and food scarcity—are likely to exacerbate human-bat interactions, increasing the overall risk of spillover events in the future.

In conclusion, the CMIP6-based multi-model downscaling ensemble indicates that Bangladesh will likely follow the “dry gets drier, and wet gets wetter” (DDWW) pattern, with drier winters and wetter monsoons projected for the future. Rising temperatures are expected across all seasons, with the most pronounced increase during winter. Rainfall variability will intensify in all periods, and all seasons except winter will become wetter, further exacerbating the country’s vulnerability to climate-related hazards such as floods, landslides, and other extreme events.

While conditions like colder winters and warmer, drier monsoons—currently associated with heightened spillover risks—may shift, new stressors such as flooding and extreme heat are likely to continue driving spillover events. Therefore, ongoing research and adaptive strategies are essential to monitor and mitigate these evolving climate-related NiV spillover drivers, particularly in the context of climate change.

Conclusions

Climate change, along with other human-induced environmental stressors, is driving the emergence and spread of infectious diseases, especially bat-borne zoonotic spillover, which pose a growing global health threat, as demonstrated by the COVID-19 pandemic [8]. In this context, the thesis focuses on understanding the climatic factors influencing Nipah Virus transmission from *Pteropus* bats to humans in Bangladesh, addressing critical research gaps. Through a comprehensive investigation of climatic and anthropogenic drivers, the study integrates a broad literature review with advanced temporal, spatial, and statistical analyses, revealing the complex interplay between climate, bat biology, and human-induced environmental changes that shape the dynamics of zoonotic disease transmission, especially under the pressures of climate change.

Literature findings unveil potentials for improvement in NiV research

The in-depth literature review revealed several gaps and opportunities for advancing our understanding of the relationship between climate and Nipah (NiV) spillovers. These insights guided the methodological choices of this thesis, resulting in significant improvements. A major strength of this research lies in the utilization of ERA5 reanalysis data, which offers higher spatial and temporal resolution and broader coverage, enhancing the accuracy and robustness of climate analyses. Although WorldClim and weather station data are valuable and widely used, particularly in the tropics, they are limited by reliance on sparse and unevenly distributed stations [101]. ERA5 addresses these limitations by not depending solely on ground-based networks, but rather integrating them with a physics-based model of the climate system, thereby offering a more reliable option for studies in such regions [98]. This thesis also expands beyond the typical winter focus (the usual NiV spillover season in Bangladesh) by incorporating year-round climatic influences and extreme weather events, providing a more complete picture of the conditions leading to spillovers. Furthermore, unlike prior studies that often focused on a single major transmission route (i.e., consumption of date palm sap), this work considers all potential exposure pathways—food-borne, intermediate-host, and human-to-human—offering a comprehensive understanding of NiV dynamics by quantifying and mapping the abundance of these established drivers.

Key results of this thesis

The literature review highlights the critical role of climate in influencing the biology, physiology, and behavior of bats, particularly in the context of zoonotic diseases emergence. Building on this knowledge, this thesis conducted a series of detailed temporal (Chapter 3) and spatial (Chapter 4) analyses that provide novel insights into the climatic and ecological drivers of NiV spillover in Bangladesh:

Temporal climatic analysis: A comprehensive analysis of standardized climate anomalies identified distinct climate patterns during years with high spillover events compared to years with lower spillover counts. Specifically, colder winters, warmer and drier monsoons, and drier post-monsoon periods were significantly associated with an increased frequency of spillovers over the 18-year reference period (2001-2018). Key months such as July (monsoon), December, and February (winter) exhibited critical anomalies correlating with higher spillover occurrences.

These climatic stressors likely influenced viral shedding and transmission in multiple ways. Colder average and minimum winter temperatures may extend the virus's survival in the environment, and increase date palm sap (DPS) production through climate-driven xylem refilling, potentially leading to greater consumption of contaminated sap by bats and humans. Additionally, these colder conditions might exacerbate thermoregulatory stress in bats, adding to the nutritional stress typically experienced during winter when fruit availability is at its lowest. This scarcity could alter bat foraging behaviour, pushing them toward alternative, sub-optimal food sources in anthropogenic environment. Conversely, higher monsoon temperatures and reduced rainfall may trigger food shortages during the bats' pregnancy phase, contributing to viral recrudescence later in the winter, with a delayed impact of 6 to 9 months [82], [83]. Although the precise mechanism behind this time lag remains unclear, the cumulative impacts of these external stressors mirrors observations of Hendra virus dynamics in Australia [82], [83]. This finding is particularly noteworthy, as it highlights a previously unreported correlation between monsoon climate conditions and the frequency of winter spillover events in Bangladesh. Supported by analogous patterns seen in Australia, this connection offers a fresh perspective on how seasonal climatic factors may drive NiV dynamics in bats, contributing to zoonotic spillover risks in Bangladesh.

Moreover, the study reveals a clear correlation between specific weather extremes—dry days, cold days, and sequences of dry days followed by cold nights—and increased NiV spillover events. The lack of statistical significance for prolonged sequences of these conditions suggests that it is not their duration, but the occurrence of intense, isolated events that correlates with increased spillover risks. This emphasizes the critical role of the timing and intensity of short-term weather events in facilitating Nipah virus transmission, rather than sustained climatic conditions. Brief threshold events can cluster temporally in ways that significantly alter bat behaviors—such as foraging and movement—and potentially compromise their immune systems, thereby increasing the likelihood of virus spillover (via viral recrudescence or enhanced interactions with humans). Identifying the months most susceptible to extreme conditions linked with spillover events enables more effective strategic intervention planning and optimizes resource distribution to monitor and control potential outbreaks.

Spatial analysis of climate and environmental stressors: identification of two geographic clusters The geospatial analysis conducted using QGIS and phyton effectively mapped and quantified the heterogeneity of NiV transmission drivers, reinforcing previous findings from the literature that identify the "Nipah Belt" as a particularly vulnerable region [36], [33]. This vulnerability stems from its dense population, extensive deforestation leading to bat habitat loss and fragmentation, and the prevalence of agricultural land use and built-up areas.

Within this region, advanced techniques such as the k-means algorithm, Principal Component Analysis (PCA), and Silhouette plots were employed to identify and validate distinct ecological clusters, each characterized by unique climatic and land-use stressors. Cluster 1, located in the north, is defined by colder and wetter conditions almost year-round, with particularly lower temperatures during winter (DJF). In contrast, Cluster 2 exhibits generally warmer and drier conditions, especially during the monsoon season, along with higher population density and a greater prevalence of flood-prone areas. Notably, the districts of Rajbari and Faridpur emerged as particularly susceptible to flooding. These areas also have high densities of date palms and have reported some of the highest numbers of NiV spillover events during the 18-year reference period. The identification of these distinct clusters highlights the spatial heterogeneity of spillover risk, underscoring the importance of targeted interventions that address the specific conditions of each region.

Regression analysis of NiV spillovers and climatic variables within spatial clusters: The regression analysis builds on previous climate and spatial findings to determine the extent to which annual Nipah virus spillover events can be explained by selected climate variables across the Nipah Belt and within its two distinct clusters. Separate univariate generalized linear models (GLMs) were developed using area-weighted climate statistics and annual spillover counts, assuming a Negative Binomial distribution to account for data dispersion. Model selection involved the exhaustive feature selection method and systematic addition of selected interaction terms, identifying the most influential variables based on Akaike's Information Criterion corrected for small sample sizes (AICc).

In Case A, where the Nipah Belt is treated as a single cluster, the analysis significantly advances over previous efforts [33] by achieving greater explanatory power and a more refined model fit. By integrating climate variables beyond the traditional winter months, the research uncovers complex patterns tied to NiV spillover events, revealing that not only colder winter, but also warmer, drier conditions during July's monsoon also play a crucial role in shaping the annual spillover variability. Notably, it establishes for the first time a direct link between July's climate conditions and a subsequent increase in spillover events during the winter season, suggesting a potentially delayed impact on bat behavior or immune responses. This newly identified pattern for the Nipah virus in Bangladesh mirrors dynamics observed with Australia's Hendra virus [82] [83], enhancing our understanding of how earlier climate conditions can influence viral transmission. The specific role of July, rather than the entire monsoon season, was uncovered through an in-depth temporal resolution analysis of variables in the pre-winter period, clarifying the timing of climatic impacts on disease transmission.

Further cluster-level analyses reinforce these findings and provide additional insights: harsh winters conditions are critical in the northern cluster (Cluster 1), while in the central cluster (Cluster 2), both hot, dry conditions in July coupled with colder winters are key drivers. These results underscore the spatial heterogeneity of climate factors that influence spillover events in different regions of Bangladesh.

Overall, these findings highlights the need for region-specific approaches in assessing and mitigating NiV spillover risks, and emphasize the importance of incorporating broader range of climate variables beyond winter conditions in future predictive models. This approach will lead to more accurate predictions and better-targeted interventions.

Limitations, future developments and implications

One limitation of this study is the constrained resolution of climate data, dictated by the granularity of accessible spillover data. The analysis was limited to district-level spatial resolution and monthly time intervals, which restricted the depth of investigation. Future research could benefit from more detailed spillover data, both spatially and temporally, allowing for better use of high-resolution ERA5 datasets. This would enable the examination of specific short-term climatic patterns leading up to spillover events, potentially revealing unknown mechanisms.

Enhancing the Generalized Linear Models (GLMs) used in the final regression analysis is another potential area for improvement. While effective, the current models may not fully capture the complexity of interactions between climate and environmental variables. Future studies could explore more advanced modeling techniques or include additional variables to better represent these interactions, providing a deeper understanding of the drivers of NiV spillovers.

A more precise understanding of the climatic triggers associated with spillover events could also facilitate predictive studies, supporting the development of interventions to prevent or mitigate NiV outbreaks. Such predictive capabilities are crucial for public health planning, especially in the context of climate change, enabling timely responses to emerging threats and reducing the risk of future pandemics.

Appendix A

A.1 Climate variability: ENSO

The **El Niño Southern Oscillation** (ENSO) is the most dominant mode of coupled atmosphere-ocean variability in the present climate (Neelin et al., 1998), and it has been identified as a critical driver of the emergence and outbreaks of several infectious diseases (such as cholera, Rift Valley fever, visceral leishmaniasis, dengue, Zika virus, and malaria) [17]. It is characterised by sea-surface temperature anomalies (SST) in the central-eastern equatorial Pacific, and it has the ability to modify the global atmospheric circulation, which in turn, influences temperature and convective precipitation across the globe [123] (refer to Figure A.1). Its warm phase with the positive SST anomaly is called El Niño and occurs in December-February period, and the opposite phase is called La Niña (i.e. cold phase) [123]. ENSO occur in irregular cycles of 3–7 years, and between the two extreme phases, there is a third one called ENSO-neutral. Notably, the persistence of the two phases can be attributed to the positive Bjerknes feedback, which represents the atmospheric response to the oceanic forcing and vice versa.

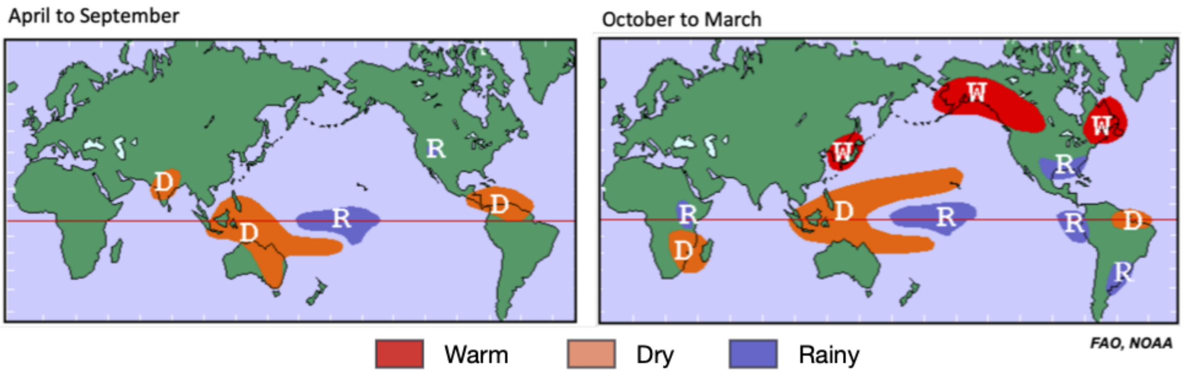


Figure A.1: Global impact of El Niño: temperature and precipitation patterns that are typical of April to September (left) and October to March (right)of El Niño conditions. Map by NOAA Climate.gov and FAO.

A.2 Bangladesh Case

A.2.1 Detailed insights into embolism and Xylem refilling phenomenon

High temperatures and low humidity increase evapotranspiration [1], reducing soil water potential with a subsequent drop in pressure within the *xylem* [125], the plant's vascular system responsible for transporting water and nutrients from roots to leaves. Water transport follows increasingly negative pressure gradients, so as soil water potential decreases, the potential in the root, stem, and leaf also drops. However, this decline has a limit, as the leaf must still transpire. Prolonged drought led to stomata closure, thereby reducing transpiration [126].

Sap flow is highly sensitive to sudden changes in intravascular pressure caused by increased atmospheric water demand [125]. Prolonged water stress in xylematic plants can lead to rapid pressure fluctuations resulting in the formation of air bubbles in the xylem vessels, a phenomenon known as *cavitation.* These air bubbles can disrupt sap flow, causing a complete blockage known as *embolism* [125]. Vascular plants can repair embolism through rapid physiological mechanisms that restore the functionality of embolized conduits. This process, known as *xylem refilling*, involves water entering the xylem vessel through positive pressure generated by the *phloem*, which transports nutrients from the leaves to other parts of the plant (known as phloem transport) [125], [127].

The *xylem refilling phenomenon* is particularly relevant for date palms, as it support sap production during drought periods. Like many other vascular plants, date palms experience diurnal cycles of xylem pressure, with negative pressure during the day and positive pressure at night [56]. These plants undergo cavitation and embolism due to the negative pressure generated by leaf transpiration and reduced soil water availability during droughts. At night, when temperatures drops and transpiration decreases, refilling occurs. Water from the soil and surrounding tissues is pushed through embolized xylem vessels, restoring their functionality and allowing the circulation of water and nutrients.

In Bangladesh, xylem refilling in date palms intensifies during the dry winter months [56]. The combination of high evaporative demand on sunny, low-humidity winter days, and clear, cold nights, increases xylem pressure, facilitating refilling and boosting sap production [56]. This process supplies essential nourishment to nectar-feeding fruit bats during a period of scarcity, providing a vital food provision when other resources are limited.

Previous studies have documented that in Bangladesh, sap production from date palms is highest from December to March, peaking in January, the coldest month of the year[56]. Additionally, reports from sap harvester indicate that sap flow increases on colder winter nights, further supporting this evidence [56].

Interestingly, human Nipah virus spillover events are more frequent during colder winters [56]. This correlation suggests that weather patterns influencing date palm physiology and sap production may contribute to annual variations in sap consumption by humans and bats, thereby affecting the risk of Nipah virus spillover [56].

---

[1] **Evapo-transpiration (ET)** is the process by which water is transferred from the land to the atmosphere by *evaporation (E)* from the soil and other surfaces (i.e. plant and open water bodies) and by *transpiration (T)* from plants' canopy through the stomata [124].

A.2.2 Reconstruction of Nipah virus spillover events

Methods

Year 2001 In 2001, of the 13 reported cases of Nipah encephalitis, only one was identified as a spillover event, as shown in Table A.2. Sources [128] and [93] consistently report on the number, date, and location of this event.

Year 2002, 2006 and 20016 According to all consulted sources, no Nipah virus cases were detected in the years 2002, 2006, and 2016.

Year 2003 The total number of cases and spillover events for 2003 is detailed in Table A.2.

Year 2004 Comprehensive details of the Nipah virus spillover events in 2004 are documented in Table A.2. That year saw 67 reported cases, the highest in the 18-year period, with two significant outbreaks: one in Rajbari (January to April) and the other in Faridpur (February to April).

The Rajbari outbreak began with a child displaying symptoms on January 11, suggesting a likely spillover date of January 4, given the assumed seven-day incubation period. Seven children and one adult from nearby Kazipara and Juran Molla Para villages were also affected, likely through exposure to bat-contaminated fruit or contact with bat secretions and excretions. The predominance of NiV cases among young boys during this first outbreak can be attributed to a specific childhood activity—tree climbing—which significantly increase their exposure to the virus. The cases, concentrated within three households, were identified as three separate index cases [129], [58], [130], [131].

In Manikganj, cases recorded between January 4 and February 8 align with early January spillovers [130]. Conversely, Faridpur witnessed an isolated event in April, with an index case showing symptoms on February 19, suggesting a February 12 spillover[130], [131], [132].

The second outbreak took place in Faridpur district, affecting seven villages from February 19 to April 17, and involved a substantial person-to-person transmission chain [130]. Nearly all infected individuals were either residents of Guha Laksmipur village or had direct contact with a local spiritual leader affected by the virus [130]. The virus underwent four distinct transmission cycles within the human population in this area. This significant outbreak resulted in 36 reported cases. The index case developed symptoms on February 19 [130], and the presumed spillover event is dated to February 12, based on the incubation period and initial symptom onset.

For the first outbreak, sources [129] and [131] indicate that spillovers were concentrated between January and February across six districts, with the additional cases occurring in March and April. The second outbreak's spillovers were noted from March to April.

The monthly distribution of spillover events in 2004, highlighted in blue in Table A.1, was meticulously reconstructed using sources [128], [93], [33], and [59]. This distribution was derived by comparing the annual spillover event count from [33], the monthly index cases reported in [93] (in *Figures 2 and 3*), and the patterns observed in previous years. This approach provided a coherent temporal framework, revealing nine unallocated spillover events in addition to the six confirmed occurrences documented in literature, and shown in black in Table A.2: two in January, one in February, four in March, one in April, and one in December. To determine the geographic distribution of these events, spillovers were assigned to districts with single reported NiV cases, considering geographical separation to avoid misattribution. This method facilitated the identification of one spillover event for each of the following districts: Rangpur, Natore, Dhaka, Gopalgonj.

Yearly number of spillover events	**Year**	**Number of Index cases by month**	**Spillover month**	**Notes**
1	2001	1	April	-
0	2002	0	-	-
1	2003	1	January	-
15	2004	1	December	-
		6	January	
		3	February	
		4	March	
		1	April	
2	2005	1	January	Information reported doesn't match with sources [93] and [33].
0	2006	0	-	-
5	2007	2	January	
		1	February	-
		2	March	-
Sources				
Fig. 1D from [33]	-	For years 2001 and 2003 are: [128], [93], [33] For year 2005 is [93] For year 2007 is the Figure S2 in S.I. of [59]		

Table A.1: Annual count of NiV cases and corresponding spillover events from 2001 to 2007.

The spatial allocation was further refined through a detailed analysis of *Figure 1* from [93], which maps the distribution of spillover events from 2001 to 2007. Using QGIS, the raster was imported, georeferenced, and reprojected into the EPSG:3106 - Gulshan 303 / TM 90 NE coordinate system (refer to Figure A.2). This process enabled the precise identification of the spillover events' locations for 2004 by cross-referencing the spatial distribution of events in other years. Based on this analysis, the final geographic distribution of the remaining unallocated spillovers was established as follows:

- One event in Faridpur district;
- One event in Rajshahi district;
- One event in Naogaon district.

After completing the spatial allocation of spillover events, each affected location was subsequently associated with a specific month, based on the previously established monthly distribution of spillover events (as outlined in Table A.2). The final geographic and temporal distribution was determined as follows:

- the December event was allocated to the Dhaka district, considering its proximity to Tangail, where the 2005 outbreak occurred;
- the two January events were attributed to Nagaon;

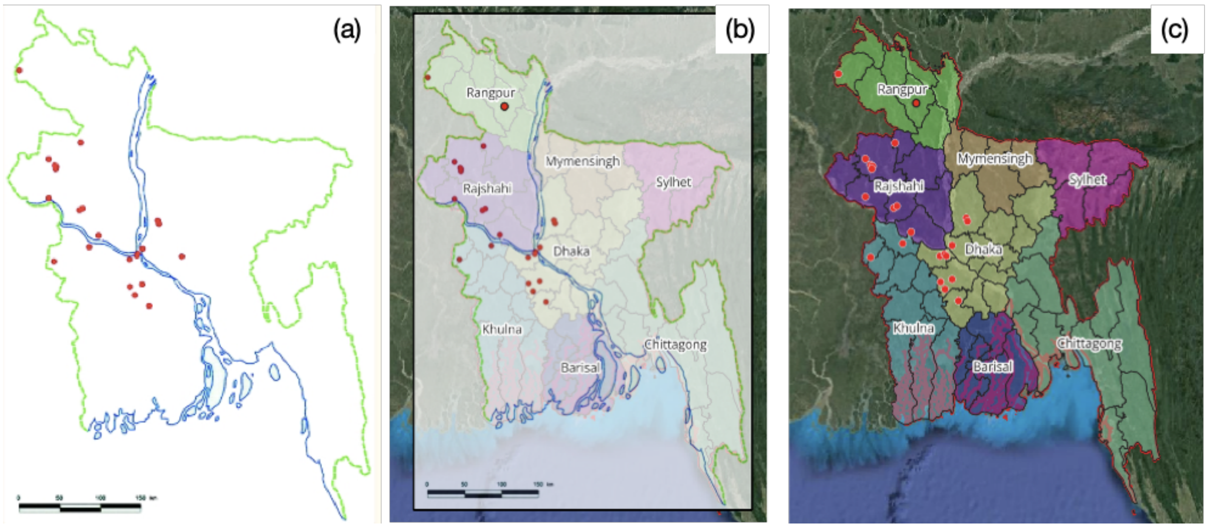


Figure A.2: (a) Figure 1 taken from [93] reporting the location of spillover events in years 2001-2007; (b) Figure 1 taken from [93] imported, georefrenced and re-projected in QGIS; (c) Resulting information.

- the February spillover was assigned to Joypurat, adjacent to Nagaon.
- the April event, part of Faridpur's secondary outbreak, was specifically allocated to that district.
- the March events were distributed among Rangpur, Natore, Rajshahi, and Gopalgonj.

Therefore, these allocations were guided by a thorough integration of literature data with the detailed spatial and temporal analyses previously discussed.

Year 2005 In 2005, 12 Nipah virus cases were reported, of which two were identified as spillover events. Detailed information is provided in Table A.2.

Year 2007 In 2007, two distinct Nipah virus outbreaks occurred, resulting in a total of 12 cases, including five spillover events (refer to Table A.2).

The first outbreak, in Thakurgaon District, spanned January and February. The index case, showing symptoms on January 21, died five days later, with the spillover event retrospectively dated to January 14, in Haripur Upazila.

The second outbreak occurred between March and April in Sadar Upazila, Kushtia District, involving two confirmed and five probable cases. The initial case manifested on March 17, with the spillover date inferred as March 10. Subsequent cases emerged 12-16 days later, indicating a single phase of secondary transmission. Temporal details of these events were derived from *Figure S2* in the Supplementary Information of [59].

Year 2008 In 2008, 10 Nipah virus cases were reported, with two identified as spillover events (Table A.2). These events, occurring within a two-week period in February, impacted the Rajbari and Manikgonj districts, which are located 44 kilometers apart at the confluence of the Padma and Jamuna Rivers [133]. The spillovers were linked to the consumption of raw Date Palm Sap (DPS) [133].

Year 2009 In 2009, two Nipah virus cases were documented, both identified as independent spillover events [59]. Detailed information is available in Table A.2.

Year 2010 In 2010, 18 Nipah virus cases were reported, with 12 classified as spillover events. These events were categorized as isolated or clustered based on the distribution and frequency of NiV cases within each affected district. Specific details are provided in Table A.2.

Year 2011 During year 2011, 43 Nipah virus cases were documented, including nine identified as distinct spillover events. The locations of two spillover events, one in Comilla and one in Kurigram, were determined [134], though the specific months and dates remain unclear due to the lack of available information. Events with missing information are highlighted in red in Table A.2.

Year 2012 In 2012, 10 out of 17 NiV cases were identified as spillover events, as documented in Table A.3:

- *Joypurhat District:* Three spillover events occurred in January, including two clustered and one isolated event. The clustered cases involved two children in Joypurhat Sadar who first exhibited symptoms on January 17, with the spillover estimated on January 10.

- *Rashashi District:* One spillover event occurred in February involved a tari producer who harvested date palm sap (DPS).

- *Multiple Districts in March:* One spillover event occurred in each of the following districts: Natore, Gopalgonj, Faridpur, and Rajbari.

- *Rangpur District:* One spillover event was documented in November.

- *Dinajpur District:* One spillover event took place in December.

Year 2013 In 2013, 31 cases of Nipah virus were reported, with 17 classified as spillover events (refer to Table A.3 for further details) according to data from [59] and [135].

Many events occurred in January across several districts. In Nilphamari, symptoms began on January 23, with the spillover estimated a week earlier. Kurigram followed a similar pattern, with symptoms appearing on January 26 after a January 19 spillover. Gaibandha and Naogaon reported early January spillovers, with symptoms appearing within a week. Additional January events occurred in Rajshahi, Natore, Pabna (which witnessed two events, including one occurred on January 16), Rajbari, and Jhinaidha.

Spillover events continued into February in Kushtia and Magura. March saw an event in Manikgonj, while October and December each witnessed a spillover event in Gopalgonj and Manikgonj, respectively. This distribution underscores significant temporal and geographical spread of the virus within the year.

Year 2014 In 2014, 11 out of 37 documented Nipah virus cases were designated as spillover events according to [94], compared to only 9 identified by [33]. Although [33] generally served as the primary reference (with the exception of 2011 when [59] was used), the more detailed dataset provided by [94] was selected for the 2014 analysis. This decision was driven by the need for a more detailed and comprehensive account, as [33] lacked sufficient data to fully capture all recorded incidents.

Certainty in event reporting Six spillover events confirmed by multiple sources included five occurrences in January and one in February. These events, specifically two in Rangpur and two in Faridpur, were detailed as follows:

- *Rangpur:* two Index cases were identified as part of a family cluster where the patients, who had consumed tari from a common source, showed symptoms on January 14th and 19th. While sources [135], [46], and [94] refer to these patients using different identifiers, the case details are consistent across these references. Specifically, *Figure 2* from [46] labels them as patients 'K' and 'L', whereas *Table 4* and *Figure 3* from the Technical Appendix of [94] discuss these cases as part of 'Cluster A'.
- *Faridpur (Saltha):* one case on January 31st was aligned with similar symptoms appearing in a family member the following day, indicating localized transmission from a shared source of raw date palm sap. Data concerning these Faridpur cases, sourced from [135], align with the descriptions of patients '3' and '6' of Cluster B in [94].

This consistency across sources not only confirms the detailed timing and connections of the events, but also aids in reconstructing additional spillover events.

Reconstruction of Uncertain Events To account for the 11 spillover events, seven additional incidents were identified, comprising two clusters and five isolated cases. The distribution of these events was guided by a detailed analysis of temporal data from [94] and spatial information from [135].

- *Cluster B:* One spillover was inferred to correspond to case-patient '14' of Cluster B [94], who developed symptoms on January 30 (spillover date assumed to be January 23).
- *Magura (Sadar):* An isolated case (patient '2') reported in [94] developed symptoms on January 14, coinciding with a primary case involving a 14-year-old in Magura at Sadar [135]. Assuming a one-week incubation period, the spillover date was set to January 7, matching timing and location across sources.

To be consistent with [94], the remaining spillover events were the following:

- Three isolated events of patients '7','8' and '9', who developed symptoms on January 20th, March 27th, and April 1st, respectively (spillover dates: January 13th, March 20th, and March 25th);
- One cluster of two events (Cluster C) with spillover dates on January 24th and 25th.

Further approximations for the spatial distribution of remaining spillover events were derived from *Figure 4* of [136], which highlights affected districts in 2014-2015 period (notably 2016 reported no cases). This spatial analysis facilitated the identification of districts where spillover events likely occurred:

- *Panchagarh:* WHO reports only one NiV case in 2014 and none in 2015, suggesting an isolated spillover event here, given its distance from Rangpur, the nearest affected district.
- *Kushtia:* With one reported case in 2014 and its distance from other affected districts, it was inferred that a spillover event occurred here.
- *Rajshahi:* Multiple cases indicated at least one spillover event, likely a cluster. It was assumed that Cluster C from [94] occurred here, with onset dates for cases '15' and '4' on January 31 and February 4, making the spillover dates January 24 and 27, 2014.
- *Manikganj:* Highlighted as an affected district in [136], where only one NiV case was reported by WHO in 2014, indicating a spillover event.

In summary, fragmented data dispersed across multiple publications often lacked complete details regarding dates and locations of spillover events. Where direct inferences were not feasible, assumptions were made guided by the spatial and temporal distribution of NiV cases from WHO and IECDR sources. Nevertheless, by integrating all available sources, a reasonable and comprehensive allocation of spillover events for 2014 was achieved, ensuring that all incidents were accounted for within the constraints of available data.

Year 2015 In 2015, 15 Nipah virus (NiV) cases were reported, with 9 identified as spillover events [135]. The distribution and timing of these events were reconstructed using detailed information from various sources, as outlined in Table A.3.
To account for the 9 spillover events reported in [33], three additional events were inferred based on geographical and temporal data.

Inferred spillover events Information from [136] and WHO reports was used to identify affected districts and their corresponding cases. This analysis indicated that the remaining districts with spillover events were Natore and Faridpur.

In Natore, where only one case was reported, it was inferred that a single spillover event occurred, likely in January, given its proximity to Naogaon, which experienced a cluster of cases in the same month.

In Faridpur, three cases were reported, leading to the attribution of two additional spillover events. Temporal data from [137] suggested these events likely occurred in January, coinciding with a period of increased spillover risk.

The integration of sources [135], [137], [136], and WHO enabled a comprehensive reconstruction of NiV spillover events in 2015, as summarized in Table A.3. Primary cases were confirmed in Naogaon, Gopalganj, Madaripur, and Rajshahi, with additional inferred events in Natore and Faridpur. This approach ensured that all occurrences were accurately represented within the limits of the available data, providing a detailed understanding of the outbreak dynamics for that year.

Year 2017 In 2017, three Nipah virus cases were reported, resulting from two spillover events. According to data from [135], these primary cases occurred in February: one in Faridpur, with symptoms appearing on February 15 in Joykalidangi (spillover date: February 8, 2017), and the other in Pabna on February 10 in Uttar Raghobpur (spillover date: February 3, 2017). The distribution of these spillover events is summarized in Table A.3.

Year 2018 In 2018, four Nipah virus cases were reported, linked to three spillover events [135], [46]. The events were distributed as follows:

- In Faridpur, one patient developed the first symptoms on February 8, 2018, with the spillover date estimated to be February 1, 2018 [135].
- In Rajshahi, one patient developed symptoms on April 3, 2018, with the spillover date estimated as March 27, 2018 [135].
- In Bogra, one encephalitis case was reported in February, as reported in [46].

The detailed spatio-temporal distribution of these spillover events is summarized in Table A.3.

Final results

Spillover events were classified as 'real' when corroborated by multiple sources or explicitly reported in literature, 'reconstructed' when inferred trough available information and 'dubious' when when precise information was lacking or based on general assumptions. The final results are reported in Tables A.2 and A.3.

Year	**Number of spillover events N_s**	**District**	**Division**	**Date**	**Source**
2001	1	Meherpur (Chandpur village)	Khulna	April 13th	[128], [93]
2003	1	Naogaon	Rajshahi	January 4th	[128], [93]
2004	1	Rangpur	Rangpur	March	Assumptions and sources [129]-[59]
	1	Joypurhat	Rajshahi	February	Assumptions and sources [129]-[59]
	2	Naogaon	Rajshahi	January	Assumptions and sources [129]-[59]
	1	Natore	Rajshahi	March	Assumptions and sources [129]-[59]
	1	Rajshahi	Rajshahi	March	Assumptions and sources [129]-[59]
	1	Faridpur	Dhaka	April	Assumptions and sources [129]-[59]
	1	Gopalgonj	Dhaka	March	Assumptions and sources [129]-[59]
	1	Dhaka	Dhaka	December	Assumptions and sources [129]-[59]
	1	Manikganj	Dhaka	January	[130]
	3	Rajbari (Goalando township, Kazipara and Juran Molla Para villages)	Dhaka	January 4th	[129], [58], [130], [131]
	1	Faridpur, Guha Laksmipur village	Dhaka	February 12th	[130], [131], [132]
	1	Faridpur	Dhaka	April	[130], [131], [132]
2005	2	Tangail	Dhaka	January	[93]
2007	1	Thakurgaon, Haripur Upazila subdistrict	Rangpur	January 14th	[59], [138]
	1	Naogaon	Rajshahi	March	[59]
	1	Natore	Rajshahi	February	[59]
	1	Pabna	Rajshahi	January	[59]
	1	Kushtia, Sadar Upazila	Khulna	March 10th	[59], [139]
2008	1	Manikganj	Dhaka	February	[59], [133]
	1	Rajbari	Dhaka	February	[59], [133]
2009	1	Rangpur	Rangpur	January	[59]
	1	Rajbari	Dhaka	January	[59]
2010	1	Kurigram	Rangpur	December	[59]
	2	Rajbari	Dhaka	one in March, one in December	[59]
	5	Faridpur	Dhaka	one in january, one in february two in March, one in December	[59]
	1	Madaripur	Dhaka	February	[59]
	2	Gopalgonj	Dhaka	one in February, one in March	[59]
	1	Khustia	Khulna	February	[59]
2011	1	Kurigram	Rangpur	December	[59]
	2	Rajbari	Dhaka	one in March, one in December	[59]
	5	Faridpur	Dhaka	one in january, one in february two in March, one in December	[59]
	1	Madaripur	Dhaka	February	[59]
	2	Gopalgonj	Dhaka	one in February, one in March	[59]
	1	Khustia	Khulna	February	[59]

Table A.2: Nipah virus spillover events in Bangladesh from 2001 to 2011, detailing district, date, and sources. Events classified as 'real' are shown in black, 'reconstructed' based on inferred information in blue, and 'dubious' in red.

Year	Number of spillover events N_s	District	Division	Date	Source
2012	1	Dinajpur	Rangpur	December	[59], [135], [46]
	1	Rangpur	Rangpur	November	[59], [135], [46]
	3	Joypurhat Sadar	Rajshahi	January	[59], [135], [46]
	1	Natore	Rajshahi	March	[59], [135], [46]
	1	Rajshahi	Rajshahi	February	[59], [135], [46]
	1	Rajbari	Dhaka	March	[59], [135], [46]
	1	Faridpur	Dhaka	March	[59], [135], [46]
	1	Gopalgonj	Dhaka	March	[59], [135], [46]
2013	1	Gaibandha (Polashbari)	Rangpur	January 7th	[59], [135]
	1	Kurigram (Fulbari)	Rangpur	January 19th	[59], [135]
	3	Nilphamari (Sayedpur)	Rajshahi	January 16th	[59], [135]
	1	Naogaon (Bodolgachi)	Rajshahi	January 12th	[59], [135]
	1	Natore	Rajshahi	January	[59], [135]
	2	Pabna (Chatmohor)	Rajshahi	January 9th	[59], [135]
	1	Rajshahi	Rajshahi	January	[59], [135]
	2	Manikganj	Dhaka	one in March, one in December	[59], [135]
	1	Rajbari	Dhaka	January	[59], [135]
	1	Gopalgonj	Dhaka	October	[59], [135]
	1	Magura	Khulna	February	[59], [135]
	1	Jhenaidah	Khulna	January	[59], [135]
	1	Kushtia	Khulna	February	[59], [135]
2014	2	Rangpur (Sedar)	Rangpur	January 7	[135], [46], [94]
	2	Faridpur (Saltha)	Dhaka	one on January 24th, one on January 25th	[135], [94]
	1	Faridpur (Saltha)	Dhaka	one on January 23th	[94]
	1	Magura (Sadar)	Khulna	January 7th	[94], [135]
	1	Panchagarh	Rangpur	March	[94], [136], WHO + Assumptions
	1	Khustia	Khulna	March	[94], [136], WHO + Assumptions
	2	Rajshahi	Rajshahi	1 on January 24th, 1 on January 27th	[94], [136], WHO + Assumptions
	1	Manikganj	Dhaka	January 20	[94], [136], WHO + Assumptions
2015	1	Naogaon, at Manda	Rajshahi	January 8th	[135]
	1	Rajshahi, at Bagha	Rajshahi	March 16th	[135]
	2	Madaripur, at Shibchar	Dhaka	February 28th, March 3rd	[135]
	2	Gopalganj, sub-district Mokshedpur	Dhaka	one on February 11; one in the same month and place	[135]
	1	Natore	Rajshahi	January	[136], WHO + Assumptions
	2	Faridpur	Dhaka	January	[137],[136], WHO + Assumptions
2017	1	Faridpur, at Joykalidangi	Dhaka	February 8	[135]
	1	Pabna, at Uttar Raghobpur	Rajshahi	February 3	[135]
2018	1	Faridpur	Dhaka	February 1st	[135]
	1	Rajshahi	Rajshahi	March 27th	[135]
	1	Bogra	Rajshahi	February	[46],[140]

Table A.3: Nipah virus spillover events in Bangladesh from 2012 to 2018, detailing district, date, and sources. Events classified as ’real’ are shown in black, ’reconstructed’ based on inferred information in blue, and ’dubious’ in red.

A.2.3 Data analysis Tools and Methods for annual spillover distribution

To identify the distribution that best fits the annual spillover data, which spans from May to April each year, the data were fitted to both Poisson and Negative Binomial distributions to determine which discrete probability distribution better describes the annual patterns. Given that the variance ($s = 30.25$) is significantly greater than the mean ($\bar{x} = 5.61$), it is unlikely that a Poisson distribution, which assumes equal mean and variance, would accurately describe the data. Nonetheless, a Poisson model was fitted to the data using `statsmodels.api.Poisson` to empirically confirm this expectation.

Poisson (discrete probability) distribution The Poisson distribution describes the likelihood of a given number of events (i.e. discrete count data) occurring in a fixed interval of time or space, assuming events occur independently and at a constant average rate. It is parameterized by λ, which represents the average rate of events, with the probability mass function (PMF) given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{A.1}$$

A key property of the Poisson distribution is that its mean (λ) equals its variance. Therefore, for data truly following a Poisson distribution, the dispersion statistic (variance/mean) should be close to 1.

In the context of this analysis, the dispersion statistic was 5.39, indicating significant overdispersion, and thus suggesting that the Poisson model may not be appropriate.
Despite this, a Poisson model was fitted to the data using `statsmodels.api.Poisson` with the Maximum Likelihood Estimation (MLE) for λ.

A chi-square test, comparing the observed vs. expected counts predicted, yielded a statistic of 26.16 and a p-value of 0.0002, leading to the rejection of the null hypothesis that the data follow a Poisson distribution.

A QQ plot further revealed deviations from the theoretical Poisson line, confirming the presence of overdispersion.

Together, the statistical test, descriptive statistics, and visual inspection of QQ plot suggest that alternative models, such as the Negative Binomial distribution, may be more appropriate for this dataset.

Negative Binomial (discrete probability) distribution The Negative Binomial distribution models the number of trials required to achieve a specified number of occurrences of an event (referred to as "success", such as thus occurrence of a spillover event with probability p) in a sequence of independent and identically distributed Bernoulli trials. The probability of observing k failures before achieving r successes is given by the PMF:

$$P(X = k) = f(k, r, p) = \binom{k + r - 1}{k} (1 - p)^k p^r \tag{A.2}$$

with mean $\mu = E(X) = r\left(\frac{1-p}{p}\right)$ and variance $\sigma^2 = \mathrm{Var}(X) = r\left(\frac{1-p}{p^2}\right)$. The parameters p and r can be expressed in terms of the first two moments as:

$$p = \frac{\mu}{\sigma^2}; \quad r = \frac{\mu^2}{\sigma^2 - \mu} \tag{A.3}$$

From equation A.3, the variance can also be re-parameterized as:

$$\sigma^2 = \mu + \alpha\mu^2, \quad \alpha = \frac{1}{r} \tag{A.4}$$

where α is the so called overdispersion parameter. Thus from equations A.3 and A.4:

$$p = \frac{\mu}{\mu + \alpha\mu^2} = \frac{1}{1 + \alpha\mu} \tag{A.5}$$

and substituting $\alpha = \frac{1}{r}$ into equation A.5:

$$r = \frac{p \cdot \mu}{1 - p} \tag{A.6}$$

The Negative Binomial is appropriate when the variance exceeds the mean, indicating overdispersion, making it a suitable alternative to the Poisson distribution. Given the rejection of the Poisson model, Negative Binomial regression was applied to the annual spillover data using `statsmodels`, with Maximum Likelihood Estimation (MLE) to estimate the model parameters $\log(\mu)$ (constant term) and α. The overdispersion parameter α was statistically significant (p-value < 0.01), confirming that the variance exceeds the mean.

Model parameters and fit The estimates for μ, p (i.e. probability of a spillover event) and r (i.e. total number of spillover events) were determined as follows:

$$\mu = \exp(\text{model.params[0]}) = 5.61$$

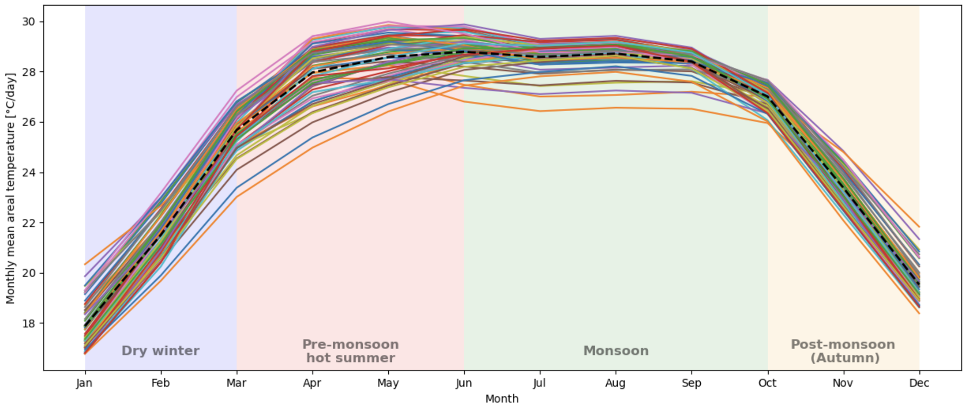
$$p = \frac{1}{1 + \mu \cdot \text{model.params[1]}} = 0.147 \quad (\text{model.params[1]} = \alpha = 1.035)$$

$$r = \frac{\mu \cdot p}{1 - p} = 0.966$$

Using these results, the histogram of the observed data was compared with the fitted Negative Binomial distribution, as illustrated in Figure 2.9a.

A.3 Climate analysis

(a) Average temperature



(b) Average precipitation

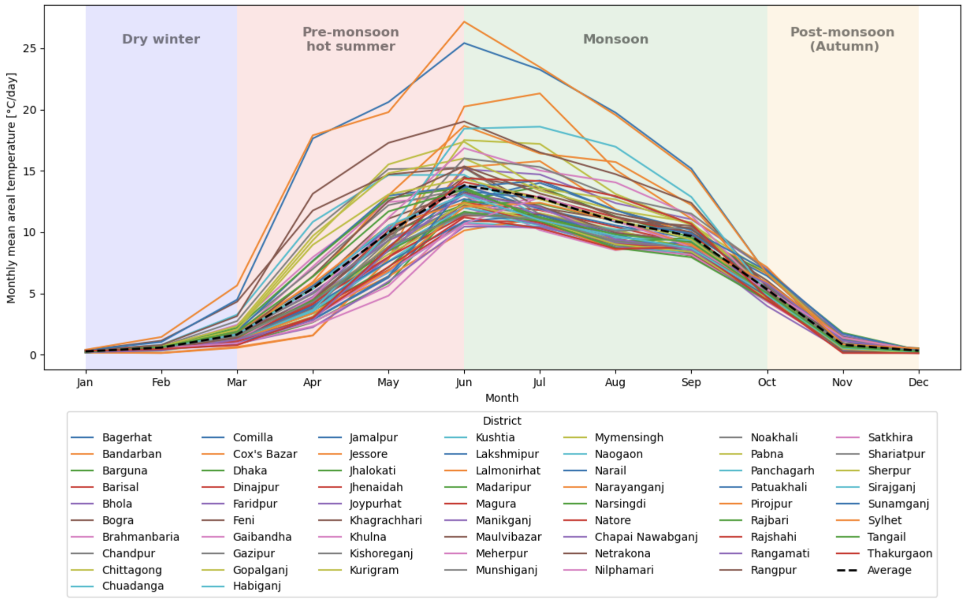


Figure A.3: *Spaghetti plot of (a) monthly mean areal temperature [°C/day] and (b) precipitation across 2000-2018 period and districts with seasonal highlights. Each colored line represents a district, with the black dashed line indicating the ensemble mean across all districts.*

A.3.1 Trend analysis

The *Theil-Sen estimator* and the *non-parametric Mann-Kendall test* are "non-parametric" methods, meaning they do not assume a specific distribution for the data, and are less sensitive to outliers, making them robust and well-suited for analyzing hydro-climatic data, as required for this study.

Sen's slope method The magnitude of the linear trend is estimated using the unbiased median-based slope estimator, known as the *Theil–Sen estimator*. A set of linear slopes are estimated as follow:

$$s_i = \frac{x_j - x_k}{t_j - t_k} \quad \text{for} \quad 1 \leq k < j < n \tag{A.7}$$

where s is the slope, x denotes the variable, N is the number of data and i and j are indexes. The n values of s_i are ranked and their median S_med is computed, resulting in the Sen's estimator of the slope.

$$s_{med} = median(s_i) \tag{A.8}$$

The intercepts are computed for each time step t as:

$$a_t = x_t - s_{med} \cdot t \tag{A.9}$$

The sign reflects direction of trend in the data, while the value indicates steepness of the trend [141].

The Mann-Kendall test is a non-parametric test for randomness against trend. The null hypothesis is that the data are Independent and Identically Distributed (IID). The alternative hypothesis is that the data follow a monotonic trend. It is a ranked based approach that compares each value of the time series with the remaining values in a sequential order [142]. The test statistic S is given by:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_k) \tag{A.10}$$

where the *sgn* function is given by:

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } (x_j - x_k) > 0 \\ 0 & \text{if } (x_j - x_k) = 0 \\ -1 & \text{if } (x_j - x_k) < 0 \end{cases} \tag{A.11}$$

S bearing positive or negative value indicates an upward or downward trend respectively [142].
The variance of S is given by:

$$Var(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} t_p(t_p - 1)(2t_p + 5)\right] \tag{A.12}$$

The probability associated with S and the sample size, n, is then computed to statistically quantify the significance of the trend. Normalized test statistic Z_{MK} (z-test) is performed as follows:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & \text{if } S < 0 \end{cases} \tag{A.13}$$

This value is compared with $Z_{\frac{\alpha}{2}}$ where $F_n\left(Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ and $F_n()$ denotes the cumulative distribution function (CDF) of a standard normal variate. The trend results in this study have been evaluated at 5% significant level, therefore the null hypothesis is rejected when

$|Z_{MK}| > Z_{\frac{\alpha}{2}}$. The alternative hypothesis is that a statistically significant trend exists in the data. The significance levels (p-values) for each trend test can be obtained from the relationship given as [141]:

$$p = 0.5 - F_n\left(|Z_{MK}|\right) \tag{A.14}$$

The $(1-\alpha)$ confidence interval of estimator s_{med} is (X_{m1}, X_{m2+1}), with:

$$n^* = Z_{\frac{\alpha}{2}} \cdot Var(S) \quad ; m_1 = \frac{n-n^*}{2} \quad ; m_2 = \frac{n+n^*}{2} \tag{A.15}$$

A.3.2 Extreme indexes

Dry Days and Cold Days calculation The dry day indicator d_i and the cold day indicator c_i are defined as:

$$\forall i \in \{1{,}2,...N\}: \quad d_i = \begin{cases} 1, & \text{if } Pr_{\text{cum, d}}(i) < 1 \left[\frac{\text{mm}}{\text{day}}\right] \\ 0, & \text{otherwise} \end{cases} \tag{A.16}$$

$$\forall i \in \{1{,}2,...N\}: \quad c_i = \begin{cases} 1, & \text{if } T_{\text{min, d}}(i) < 17\,[^\circ\text{C}] \\ 0, & \text{otherwise} \end{cases} \tag{A.17}$$

where $N = 6574$ [days] is the total number of days in the analysis period.

The monthly dry days $\{DD(m)\}$ and the monthly cold days $\{CD(m)\}$ are calculated by summing the number of days within each month m:

$$\begin{aligned} \forall m \in M, \forall d \in D: \quad DD_d(m) &= \sum_{i \in m} d_i\,; \\ CD_d(m) &= \sum_{i \in m} c_i \end{aligned} \tag{A.18}$$

where, $m \in M = \{May, June, July, ...April\}$ and $d \in D$ represents the district in Bangladesh.

The total annual (dry or cold) days for each district, are computed as:

$$DD_{y,d} = \sum_{m \in M} DD_d(m,y)\,; \quad CD_{y,d} = \sum_{m \in M} CD_d(m,y) \tag{A.19}$$

where y denotes the year in the 18-year reference period $Y = \{0{,}1{,}2,..17\}$.

The winter dry days (DD_{winter}) and cold days (CD_{winter}), are calculated by summing the counts of days in December, January, and February:

$$DD_{winter,d}(y) = \sum_{m \in \{D,J,F\}} DD_d(m,y)\,; \quad CD_{winter,d} = \sum_{m \in \{D,J,F\}} CD_d(m,y) \tag{A.20}$$

The average value of days for years with high spillover events (Y_H) and low spillover events (Y_L) are given by:

$$\begin{aligned} \forall y \in Y_H, \forall d \in D: \quad DD_d(Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} DD_{y,d}\,; \\ CD_d(Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} CD_{y,d} \end{aligned} \tag{A.21}$$

$$\begin{aligned} \forall y \in Y_L, \forall d \in D: \quad DD_d(Y_L) &= \frac{1}{n_L} \sum_{y \in Y_L} DD_{y,d}\,; \\ CD_d(Y_L) &= \frac{1}{n_L} \sum_{y \in Y_L} CD_{y,d} \end{aligned} \tag{A.22}$$

where $n_H = 7$ and $n_L = 11$ represent the number of years with high and low spillover events, respectively, as defined in Section 3.1.3.

Finally, the difference between these values ($\Delta DD_{obs,d}$ for dry days and $\Delta CD_{obs,d}$ for cold days) is computed for each district as:

$$\begin{aligned} \forall d \in D: \quad \Delta DD_{obs,d} &= DD_d(Y_H) - DD_d(Y_L)\,; \\ \Delta CD_{obs,d} &= CD_d(Y_H) - CD_d(Y_L) \end{aligned} \tag{A.23}$$

Dry days followed by cold nights The main steps for the calculation of "Dry days followed by cold nights" include:

1. **Date extraction:** the sequences of dates corresponding to dry days and cold nights are extracted separately:

$$\begin{aligned} \forall d \in D, \forall y \in Y: \quad \text{dates\_DD}_d &= \text{dates\_DD}_d(y)\,; \\ \text{dates\_CD}_d &= \text{dates\_CD}_d(y) \end{aligned} \tag{A.24}$$

2. **Date shifting and matching:** Dry day dates were incremented by one day and matched against cold night dates to identify periods when consecutive dry and cold conditions hold. This was achieved by shifting the dates_DD$_d$ array:

$$\forall d \in D: \quad \text{dates\_DD}_{d,shift} = \{\text{date} + 1\ \text{day} | \text{date} \in \text{dates\_DD}_d\} \tag{A.25}$$

and generating a boolean match array using `np.isin` method of `Numpy library` in Pyhton:

$$\text{match}_d = \text{np.isin}(\text{dates\_DD}_{d,shift}, \text{dates\_CD}_d) \tag{A.26}$$

The matching dates were then extracted and stored in match_dates$_d$, capturing the intersection of the two sequences:

$$\text{match\_dates}_d = \{\text{dates\_DD}_{d,shift}[j] \mid \text{match}_d[j] = \text{True}\} \tag{A.27}$$

3. **Count matching dates:** The number of annual consecutive dry days and cold nights was calculated at the district level, as follows:

$$D\_CD_{y,d} = \sum(\text{match}_d) \tag{A.28}$$

4. **Winter season analysis:** The process was repeated for the winter season (DJF) by filtering the matches to include only December, January, and February:
$\forall d \in D:$

$$\text{match\_dates}_{d,filt} = \{\text{date}|\text{date} \in \text{match\_dates}_d\ \&\ \text{date.month} \in \{12,1,2\}\} \tag{A.29}$$

$$\text{D\_CD}_{winter,d}(y) = \text{len}(\text{match\_dates}_{d,filt}) \tag{A.30}$$

where D_CD$_{winter,d}(y)$ is total number of these filtered dates within a given year y.

5. **Average calculation:** The results were then organized into a matrix (64 districts × 18 years) and the average number of "dry days followed by cold nights" was computed for years with high spillover events ($y \in Y_H$) and then compared to years with low spillover events ($y \in Y_L$).

$$\begin{aligned} \forall y \in Y_H, \forall d \in D: \quad \text{D\_CD}_d(Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} \text{D\_CD}_{y,d}\,; \\ \text{D\_CD}_d(Y_L) &= \frac{1}{n_L} \sum_{y \in Y_L} \text{D\_CD}_{y,d} \end{aligned} \tag{A.31}$$

6. **Difference calculation:** The difference in averages between the two groups was calculated for each district, highlighting distinctive climatic patterns.

$$\forall d \in D: \Delta\text{D\_CD}_{obs,d} = \text{D\_CD}_d(Y_H) - \text{D\_CD}_d(Y_L) \tag{A.32}$$

7. **Spatial distribution visualization:** The spatial distribution of these differences was visualized creating a `GeoDataFrame` of Bangladesh's districts.

Consecutive indexes calculation The main steps include:

1. **Data extraction** Daily cumulative precipitation and minimum temperature data are extracted for each district and year, along with the corresponding dates. Dry and cold days are identified by checking the specific criteria previously mentioned.

2. **Consecutive days detection** A function identifies and records the longest sequences of consecutive dry days CDD and consecutive cold days CCD for each district and year, employing a helper function, **`FindConsecutiveDays`** (algorithm 2) within the **`ConsecutiveDaysDetection`** procedure (algortihm 1). The sequence of dates for these consecutive days are also recorded, providing insights into the temporal distribution of these climatic events.

3. **Shifting and Matching** Dry day sequences are shifted by one day, and overlapping dates with cold day sequences are identified using the **`ShiftDates`** and **`FindMatches`** functions (refer to algorithm 3). The total number of consecutive "Dry and Cold" days (C_D_CD) is counted for each district and year.

4. **Average calculation** The average number of consecutive dry days, cold days, and "Dry and Cold" days is calculated for years with high (Y_H) and low (Y_L) spillover frequency:

$$\begin{aligned} \forall d \in D: \quad \text{CDD}_d(Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} \text{CDD}_{y,d}\,; \quad \text{CDD}_d(Y_L) = \frac{1}{n_L} \sum_{y \in Y_L} \text{CDD}_{y,d} \\ \text{CCD}_d(Y_H) &= \frac{1}{n_H} \sum_{y \in Y_H} \text{CCD}_{y,d}\,; \quad \text{CCD}_d(Y_L) = \frac{1}{n_L} \sum_{y \in Y_L} \text{CCD}_{y,d} \end{aligned} \tag{A.33}$$

$$\text{C\_D\_CD}_d(Y_H) = \frac{1}{n_H} \sum_{y \in Y_H} \text{C\_D\_CD}_{y,d}\,; \quad \text{C\_D\_CD}_d(Y_L) = \frac{1}{n_L} \sum_{y \in Y_L} \text{C\_D\_CD}_{y,d}$$

5. **Difference calculation** The difference between high and low spillover years is calculated for each district to determine whether high spillover years are characterized by longer periods of climatic extremes:

$$\begin{aligned} \forall d \in D: \quad & \Delta\mathrm{CDD}_{obs,d} = \mathrm{CDD}_d(Y_H) - \mathrm{CDD}_d(Y_L)\,; \\ & \Delta\mathrm{CCD}_{obs,d} = \mathrm{CCD}_d(Y_H) - \mathrm{CCD}_d(Y_L)\,; \\ & \Delta\mathrm{C\_D\_CD}_{obs,d} = \mathrm{C\_D\_CD}_d(Y_H) - \mathrm{C\_D\_CD}_d(Y_L) \end{aligned} \tag{A.34}$$

Algorithm 1 Consecutive Days Detection Algorithm

```
procedure ConsecutiveDaysDetection(data, dates)
    ▷ data is the daily climatic dataset from May 1, 2000, to April 30, 2018
    ▷ dates is the daily datetime series from May 1, 2000, to April 30, 2018
    ▷ Initialization
    $N$ ← length of data                                ▷ Total number of days in the dataset
    $D$ ← number of districts                           ▷ Number of districts analyzed
    $years$ ← unique years in data                      ▷ Years in the dataset
    $CDD$ ← zero matrix of size $D \times$ length of years    ▷ Matrix for consecutive dry days
    $CCD$ ← zero matrix of size $D \times$ length of years    ▷ Matrix for consecutive cold days
    $C\_D\_CD$ ← zero matrix of size $D \times$ length of years    ▷ Matrix for consecutive dry and cold days
    dates_CDD ← empty matrix of object data type of size $D \times$ length of years    ▷ Matrix for maximum sequence of dates for dry spell
    dates_CCD ← empty matrix of object data type of size $D \times$ length of years    ▷ Matrix for maximum sequence of dates for cold spell
    ▷ Execution
    for each year $y$ do
        drysp ← zero array of size $D$                  ▷ Array for dry-spell count at district level
        coldsp ← zero array of size $D$                 ▷ Array for cold-spell count at district level
        for each district $d$ do
            $time$ ← daily date time series for district $d$ in year $y$
            $pr$ ← daily cumulated precipitation data for district $d$ in year $y$
            $Tmin$ ← daily minimum temperature data for district $d$ in year $y$
            $(drysp[d], \text{dates\_CDD}[d,y])$ ← FindConsecutiveDays($pr < 1$ mm$, data, time$)    ▷ Detect consecutive dry days and store corresponding dates
            $(coldsp[d], \text{dates\_CCD}[d,y])$ ← FindConsecutiveDays($Tmin < 17°C, data, time$)    ▷ Detect consecutive cold days and store corresponding dates
            dates_CDD_shift[d,y] ← ShiftDates(dates_CDD[d,y])    ▷ Shift forward the dates by 1 day
            ($C\_D\_CD[d,y]$, matches[d,y]) ← FindMatches(dates_CDD_shift[d,y],
                dates_CCD[d,y])    ▷ Detect dry days followed by cold nights, returning the count and dates
        end for
        $CDD[:,y]$ ← $drysp$
        $CCD[:,y]$ ← $coldsp$
    end for
    return $CDD, CCD, C\_D\_CD$, dates_CDD, dates_CCD, matches    ▷ Return matrices of consecutive days and dates
end procedure
```

Algorithm 2 Find Consecutive Days Algorithm

```
procedure FindConsecutiveDays(condition, data, dates)
    ▷ Identifies longest sequence of consecutive days meeting the condition
    ▷ data is the daily climatic dataset from May 1, 2000, to April 30, 2018
    ▷ dates is the daily datetime series from May 1, 2000, to April 30, 2018
    ▷ Initialization
    $smax \leftarrow 0$                                    ▷ Maximum sequence length
    $sc \leftarrow 0$                                      ▷ Current sequence length
    $max\_sequence\_dates \leftarrow []$                   ▷ Dates of maximum sequence
    $current\_sequence\_dates \leftarrow []$               ▷ Dates of current sequence
    for each day $i$ in condition do
        if condition $data[i]$ is true then
            $sc \leftarrow sc + 1$
            current_sequence_dates.append($dates[i]$)
        else
            if $sc > smax$ then
                $smax \leftarrow sc$
                $max\_sequence\_dates \leftarrow$ current_sequence_dates
            end if
            $sc \leftarrow 0$
            current_sequence_dates ← []
        end if
    end for
    if $sc > smax$ then
        $smax \leftarrow sc$
        $max\_sequence\_dates \leftarrow$ current_sequence_dates
    end if
    return $smax, max\_sequence\_dates$        ▷ Return maximum sequence length and dates
end procedure
```

Algorithm 3 Shift Dates and Find Matches Algorithms

```
procedure ShiftDates(dates)
    ▷ Shift each date by one day forward to prepare for matching
    shifted_dates ← []
    for each date $d$ in dates do
        shifted_dates.append($d + 1$ day)              ▷ Shift date forward by one day
    end for
    return shifted_dates                               ▷ Return the list of shifted dates
end procedure
procedure FindMatches(dates1, dates2)
    ▷ Find matching dates between two lists of dates, for each district and year
    ▷ dates1 is the shifted series (i.e. shifted_dates)
    matches ← []
    for each date $d$ in dates1 do
        if $d$ is in dates2 then
            matches.append($d$)
        end if
    end for
    count ← len(matches)                               ▷ Count the number of matches
    return count, matches                 ▷ Return the count and list of matching dates
end procedure
```

A.4 Spatial analysis

Data source	**Data description**	**Data format and Horizontal resolution**	**Date**	**Source link**
Sentinel-2 L2A Esri Land Cover - ArcGIS Living Atlas of the World	Land Use and Land Cover (LULC). Thematic map (9 classes of Land cover): Water, Built area, Rangelands, Crops, Flooded vegetation, Trees, Bare ground, Snow/Ice, Cloud. Source Data Coordinate System: Universal Transverse Mercator (UTM) WGS84.	Raster file (.GeoTIFF); Cell size 10 m.	2018	Esri Sentinel-2 Land Cover Explorer
WorldPop	Estimated Population Density per grid-cell (2000-2018). The projection is Geographic Coordinate System, WGS84. The units are number of people per square kilometer.	Raster files (.GeoTIFF); ~30 arc sec (1 km^2 at equator)	2020	WorldPop
Bangladesh Bureau of Statistics (BBS)	• Area in acre of fruit crops and sugar crops (2010-2011 and 2022). Selected fruit crops (from [A] - 2010-2011): Ripe papaya, Guava, Jack fruit, Banana, Pineapple, Mango, Litchi, Blackberry • Selected fruit crops (from [B] - Area Under garden - 2019-2020): Jujube (Boroi/Kul), Star apple (Jamrul), Ata (Custard apple), Dewya, Lotkon, Carambola apple (Kamranga), Wood apple, Green Coconut (flowers), Dragon, Dalim fruit, Safeda (Sapodilla), Sharifa, Strawberry • Selected sugar crops (from [B] - Area Under garden - 2019-2020):Date-palm (juice), Date-palm (fruit), Ripe Palmyra (Tal), Palmyra palm (juice)	Tabular data	2013, 2022	A Bangladesh Bureau of Statistics (BBS), Districts statistics reports published in 2013, based on data of 2010-2011 which refers to the agriculture census 2008; B Bangladesh Bureau of Statistics (BBS) Yearbook of Agricultural Statistics 2022.
Gridded Livestock of the World (GLW4)	GIS file of the areal-weighted (AW) product. Raster files reporting the absolute number of animals per pixel (4320 by 2160 pixels): Cattle, Pigs, Goat, Sheep. GIS file of the area per pixel reporting the area in square km (4320 by 2160 pixels).	Raster files 5 arc-minutes; 0.08° x 0.08°; (~10 km^2 at equator)	2015	Gilbert, Marius; Cinardi, Giuseppina; Da Re, Daniele; Wint, William G. R.; Wisser, Dominik; Robinson, Timothy P., 2022, Harvard Dataverse, V1. Harvard GLW4
RBB SDI Bangladesh WFL1 - ArcGIS Hub	Flood prone areas: Severe Tidal Surge, Severe Flash Flooding, Severe River Flooding, Moderate Tidal Surge, Moderate Flash Flooding, Moderate River Flooding	Vector data (.shp)	2019	ArcGis - link1; ArcGis - link2

Table A.4: Description of data, data sources and associated details.

A.4.1 Detailed methodology of K-means clustering process

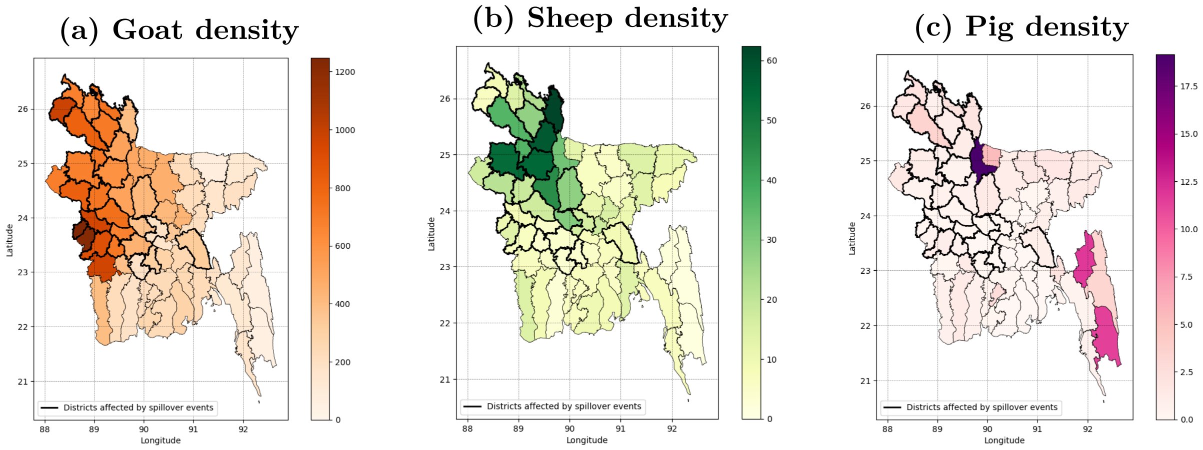
The K-means algorithm is a widely-used method for partitioning N observations into K non-overlapping clusters, each defined by a centroid that minimizes the within-cluster sum of squares (WCSS). It is particularly effective for grouping data with similar attributes into distinct categories.

The objective function for K-means clustering, known as inertia or WCSS, is mathematically expressed as:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{K} w_{ji} \left\| x_j - \mu_i \right\|^2 \tag{A.35}$$

where:

- $K$is the total number of clusters;



(a) Goat density (b) Sheep density (c) Pig density

Figure A.4: Spatial distribution of district-level densities of Livestock, specifically: (a) Goat, (b) sheep and (c) pigs. Black-bordered districts indicate those belonging to the Nipah belt, which were affected by spillover events during the 2000-2018 period.

- N is the total number of data points;
- X_j is the j-th data point with $j = 1{,}2, \dots, N$;
- μ_i is the centroid of the i-th cluster (C_i) with $i = 1{,}2, \dots, K$;
- $w_{ji} \in \{0, 1\}$ is a binary indicator that describing which of the K clusters the data point x_j is assigned to, so that if data point x_j is assigned to cluster i then $w_{ji} = 1$, and $w_{jh} = 0$ for $h \neq i$;
- $\|\cdot\|$ is the L2 norm, or Euclidean distance.

Inertia J measures how internally coherent the clusters are, with lower values indicating more homogeneous clusters.

For areal clusters, such as districts represented by polygons, the K-means algorithm treats each district as a point in a multidimensional feature space. Therefore the Euclidean distance is calculated in the high-dimensional space where the 35 features, serving as coordinates.

Therefore, the goal of the K-means algorithm is to find $\{w_{ji}\}$ and $\{\mu_i\}$ values to minimize the inertia J. This is achieved through an iterative procedure, called `Expectation-Maximization` (EM) approach, which involves an initialization and two successive optimization steps [108]:

1. **Initialization:** the number of clusters K are specified by the user and the corresponding cluster centroids μ_i are initialised using the `k-means++` method[2]
2. **Expectation-step (E-step):** Assign each observation x_i to the cluster C_i whose centroid μ_i minimizes the Eucliden distance $\|x_j - \mu_i\|^2$ [109]. This step involves

---

[2] The "k-means++" method selects initial cluster centroids using sampling based on an empirical probability distribution of the points contribution to the overall inertia. This technique, implemented in `scikit-learn` as "greedy k-means++", speeds up convergence by making several trials at each sampling step and choosing the best centroid among them. Source information: Scikit-learn.org

calculating the assignment matrix w_{ji}, where $w_{ji} = 1$ if observation j is closest to centroid i and $w_{ji} = 0$ otherwise. Therefore, J is minimized with respect to w_{ji} while treating μ_i as fixed. Since J is a linear function of w_{ji}, mathematically, the cluster assignment is expressed by:

$$w_{ji} = \begin{cases} 1 & \text{if } i = \arg\min_k \|x_j - \mu_k\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{A.36}$$

3. **Maximization Step (M-step):** Update each centroid μ_i to be the mean of all points assigned to cluster C_i. This step recalculates μ_i based on the current assignments, aligning the centroids with the cluster means to better reflect the data's distribution. Thus, in this second step J is minimized with respect to mu_i while treating w_{ji} as fixed. The objective function J, being a quadratic function of μ_i, is minimised by setting its derivative with respect to μ_i to zero giving [108]:

$$\frac{\delta J}{\delta \mu_i} = 2\sum_{j=1}^{N} w_{ji}\left(x_i - \mu_i\right) = 0 \quad \Rightarrow \quad \mu_i = \frac{\sum_{j=1}^{N} w_{ji} x_i}{\sum_{j=1}^{N} w_{ji}} = \frac{1}{|C_i|}\sum_{x_j \in C_i} x_j \tag{A.37}$$

The process alternates between the *E-step* and *M-step* until the changes in centroid positions are minimal, indicating convergence [108]. This iterative refinement ensures that the clusters are internally homogeneous and well-separated from each other, providing a robust grouping of the data points.

In this study, districts are treated as points in a 35-dimensional feature space, with features reflecting NiV risk factors such as climate, land use, and population densities. The clustering aims to identify regions with similar ecological characteristics, aiding in the understanding of NiV transmission dynamics.

A.5 Regression analysis: Generalized Linear Models (GLMs)

GLM: Negative binomial model definition Let $y = (y_1, y_2, ..y_n)$ represent the observations of the response variable, which is the annual count of spillover events. Each y_i corresponds to the number of spillovers observed in a given year within a given spatial cluster, with $n = 18$. For each year i (where $i = 1,2, ...n$), the known values of p explanatory variables, are represented by the row vector $x_i = \{x_{i1}, x_{i2}, ...x_{ip}\}$ [117].

A generalized linear model (GLM) is characterized by the following three components:

1. **Random component:** The response variables Y_i are assumed to be independent and follow a negative binomial distribution [117]. Hilbe provides a convenient parameterization of the negative binomial distribution [143] as a Poisson-gamma mixture [114]:

   $$p(y) = P(Y = y) = f(y; \mu, \alpha) = \frac{\Gamma(y + \alpha)}{\Gamma(y + 1)\Gamma(\alpha)} \left(\frac{\mu/\alpha}{1 + \mu/\alpha}\right)^y \left(\frac{1}{1 + \mu/\alpha}\right)^\alpha \tag{A.38}$$

   where $\mu > 0$ is the mean of Y, and $\alpha > 0$ is the *dispersion parameter*. Therefore:

   $$Y_i \sim nbinom(\alpha, \pi_i) \quad i = 1,2, ..n \tag{A.39}$$

   where π_i is the probability parameter expressed as:

   $$\pi_i = \frac{1}{1 + \alpha\mu_i}$$

   An equivalent re-parametrization gives:

   $$Y_i \sim nbinom(\alpha, \mu_i) \quad i = 1,2, ..n \tag{A.40}$$

   where μ_i is the mean. The first two moments of this distribution, mean and variance, are respectively retrieved as $E(Y_i) = \mu_i$ and $Var(Y_i) = \mu_i + \alpha\mu_i^2$ [117].

2. **Systematic component:** The linear predictor η is a linear combination of the explanatory variables:

   $$\eta = X\beta$$

   where $X = (x_1^T, x_2^T, ...x_n^T)^T$ is the $n \times p$ design matrix of the known explanatory variables, and β is the $p \times 1$ vector of regression coefficients that has to be estimated [117]. Specifically:

   $$X_{n \times p} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1p} \\ 1 & x_{22} & x_{23} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{np} \end{bmatrix}; \quad \beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$$

   Therefore, the linear predictor η is an $n \times 1$ vector:

   $$\eta = (\eta_1, \eta_2, ..., \eta_n)^T = (x_1\beta, ...x_n\beta)^T$$

   where $x_1 = (1,1, ...1)$, making β_0 the intercept of the model.

3. **Link function:** A link function $g(\cdot)$ relates the expected value of the response variable μ_i to the linear predictor η_i [117]. In this analysis, the log link function is used:
$$g(\mu_i) = log(\mu_i) = \eta_i = x_i\beta = \beta_0 x_{i1} + .. + \beta_p x_{ip}$$

Thus, the traditional negative binomial regression model is expressed as [114]:

$$log(\mu) = \eta = X\beta = \beta_0 + \beta_1 x_1 + .. + \beta_p x_p \tag{A.41}$$

Consequently, the mean μ is related to η through the inverse of the link function, which is the exponential function:

$$\mu = e^{\eta} = e^{(X\beta)} = e^{(\beta_0+\beta_1 x_1+..+\beta_p x_p)} \tag{A.42}$$

Estimation of model parameters Following the specification of the negative binomial model's distribution and the structure, the next step involves estimating the model parameters.

Referring to equation A.42, and letting x_i represents the $i - th$ row of the design matrix X, the distribution equation A.38 can be re-expressed as [114]:

$$p(y_i) = \frac{\Gamma(y_i + \alpha)}{\Gamma(y_i + 1)\Gamma(\alpha)} \left(\frac{e^{x_i \cdot \beta}/\alpha}{1 + e^{x_i \cdot \beta}/\alpha}\right)^{y_i} \left(\frac{1}{1 + e^{x_i \cdot \beta}/\alpha}\right)^{\alpha} \tag{A.43}$$

The estimation of the parameters α (dispersion parameter) and β (the regression coefficients) relies on the method of maximum likelihood estimation (MLE). The likelihood function is expressed as [114]:

$$L(\alpha, \beta) = \prod_{i=1}^{n} p(y_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \alpha)}{\Gamma(y_i + 1)\Gamma(\alpha)} \left(\frac{e^{x_i \cdot \beta}/\alpha}{1 + e^{x_i \cdot \beta}/\alpha}\right)^{y_i} \left(\frac{1}{1 + e^{x_i \cdot \beta}/\alpha}\right)^{\alpha} \tag{A.44}$$

The corresponding log-likelihood function is:

$$\begin{aligned} \log L(\alpha, \beta) = &\sum_{i=1}^{n} \Big[\log\Gamma(y_i + \alpha) - \log\Gamma(y_i + 1) - \log\Gamma(\alpha) \\ &+ y_i(x_i \cdot \beta - \log\alpha - \log(1 + e^{x_i \cdot \beta}/\alpha)) \\ &- \alpha \log(1 + e^{x_i \cdot \beta}/\alpha)\Big] \end{aligned} \tag{A.45}$$

The values of α and β that maximize the $\log L(\alpha, \beta)$ are the sought maximum likelihood estimates [114]. These estimates are essential for building a model that accurately reflects the observed data and can be used effectively for predictive and inferential purposes. However, due to the complexity of the likelihood equations in generalized linear models, explicit solutions are rare, requiring advanced numerical algorithms for parameter estimation [117]. A particularly robust method in this context is the *Iteratively Reweighted Least Squares* (IRLS), an adaptation of the *Newton-Raphson method* [117]. This method is well-suited for estimating β when a preliminary estimate of α is held constant [117]. The iterative nature of the IRLS algorithm effectively refines the estimates of the regression coefficients β until convergence is achieved, ensuring precise parameter estimation.

In the negative binomial model, the dispersion parameter α is not predefined and must be estimated from the data. The estimation approach varies depending on the context:

- **Univariate GLM:** In this approach, α is estimated by fitting a Negative Binomial model to the data using the `sm.NegativeBinomial` function from the *statsmodels* library. This function employs the 'nb2' parameterization (i.e. Variance = $\mu + \alpha \cdot \mu^2$), specifically designed for the Negative Binomial model. The log-likelihood method is used to estimate the dispersion parameter along with other model parameters, ensuring an accurate representation of data variability. Additional details are provided in Section 5.1.2.

- **Model selection:** In this context, to effectively determine the optimal value of α, an iterative optimization process is employed, which minimizes the negative log-likelihood function using the *limited memory Broyden-Fletcher-Goldfarb-Shanno* (`L-BFGS-B`) algorithm. For further details, refer to Section 5.1.2.

A.5.1 Steps in the univariate GLM process

A systematic approach was employed to estimate the dispersion parameter α and fit the negative binomial models, as outlined in the following steps. The complete code for this process is available (see Listing A.1). From this point onward, it is assumed that all analyses are conducted for each spatial cluster, and this will not be explicitly stated in each step.

1. **Step 1: Estimation of α** The first step involves estimating the dispersion parameter α, a key component in the negative binomial regression model. This is done using the `NegativeBinomial` function from the `'statsmodels'` library, which fits an auxiliary regression model to the data employing the 'nb2' parameterization and the Maximum Likelihood Estimation (MLE) method. Notably, the `‘statsmodels.formula.api‘` is designed to automatically calculate the dispersion parameter α as part of the MLE fitting process.

2. **Step 2: Check statistical significance of α** After fitting the auxiliary model, the summary is reviewed to determine the statistical significance of α. A low p-value (e.g., less than 0.05) suggests that α is statistically significant, confirming that the negative binomial model is more appropriate than a Poisson model due to overdispersion.

3. **Step 3: fitting the GLM** Once the optimal α (denoted as α_{opt}) is obtained and its statistical signficance is confirmed, it is used as an input for fitting the Generalized Linear Model (GLM) with a negative binomial distribution. Indeed, in this phase, the GLM considers α as a fixed parameter that needs to be specified by the user before fitting the model. This step ensures that the model accurately reflects the data's overdispersion, as indicated by α_{opt}.

4. **Step 4: Model evaluation** The model's fit is evaluated using pseudo-R^2 measures, which provide an indication of the model's explanatory power in a way adapted for GLMs. In this analysis, the *Cox-Snell pseudo-R^2* was employed for its ability to approximate the traditional R^2 in GLMs. It is calculated based on the likelihood of the fitted model (denote with ll_f) compared to the likelihood of the null model (a model with no predictors except for the intercept, denoted as ll_{null}), using the following formula:

$$pseudo\text{-}R^2_{CS} = 1 - \exp\left(\frac{2}{n_{obs}} \cdot (ll_{null} - ll_f)\right) \tag{A.46}$$

where n_{obs} is the number of observations in the dataset (i.e. 18).
Additionally, the p-values of the coefficients were considered to assess the model's validity.

5. **Step 5: Coefficients estimates and significance** Coefficients for the predictor variable and the intercept are estimated using MLE. Given the log link function, these coefficients represent the expected change in the logarithm of the response variable for a one-unit increase in the predictor variable (for the β_1 coefficient) and the expected log-value of the response variable when the predictor is zero (for the intercept, β_0). In `statsmodel`, these estimated coefficients can be accessed via the `.params` attribute on the model's result object. The standard error for each coefficient is then calculated as the square root of the diagonal element of the variance-covariance matrix, obtained by calling `.cov_params()` on the result object. This step is crucial because the standard error quantify the uncertainty in the coefficient estimates, reflecting the precision of the sample in representing the population.
   Subsequently, the *Wald test* is employed to assess the significance of the coefficients in the model. A test statistic Z, known as the *Z-score* or *Wald statistics*, is computed for each coefficient by dividing the estimated coefficient by its standard error. Finally, the p-values are calculated using the Z-scores to determine the statistical significance of each coefficient, indicating the strength of the association between the predictor and the response variable. The calculation uses the cumulative distribution function (CDF) of the standard normal distribution ((`norm.cdf`)), expressed as:

   $$p - value = 2 \cdot (1 - \text{norm.cdf}(\text{np.abs}(Z))) \tag{A.47}$$

   This formula computes the two-tailed p-value for each coefficient, representing the probability of observing the given result by chance if the null hypothesis of no effect is true. The statistical significance is denoted by asterisks placed next to the coefficient value, as follows:

   - `"***"` if the p-value < 0.001, indicating very strong evidence against the null hypothesis.
   - `"**"` if the p-value < 0.01, indicating strong evidence against the null hypothesis.
   - `"*"` if the p-value < 0.05, indicating significant evidence against the null hypothesis.

6. **Step 6: Coefficients interpretation** The coefficients are interpreted by calculating the change factor for the predictor variable, which represents the factor by which the response variable changes for a one-unit increase in the explanatory variable:

   $$\text{change\_factor} = \exp(\beta) \tag{A.48}$$

   This factor is then converted into a percentage change to facilitate practical interpretation:

   $$\text{change\_factor}_{perc} = \begin{cases} (\text{change\_factor} - 1) \cdot 100 & \text{if change\_factor} > 1 \\ (1 - \text{change\_factor}) \cdot 100 & \text{otherwise} \end{cases} \tag{A.49}$$

7. **Step 7: prediction values and confidence intervals** The analysis includes generating a range of predictor values using `np.linspace` to create an array of 100

evenly spaced values between the minimum and maximum observed values of the predictor variable. This expanded range facilitates a smoother visualization of the model's predictions, addressing the limitation of the small number of predictor values (i.e. 18 values, one for each year) in the original dataset. Additionally, this approach enables a detailed analysis and prediction at unobserved values, enhancing the model evaluation process. Indeed, by examining the model's behavior across a broad range of predictor values, a more robust analysis can be conducted, ensuring the predictions are both accurate and applicable to a wider set of conditions beyond the observed data points.
The `summary_frame` method from the `statsmodels` library provides a comprehensive summary of the model's predictions, including:

  - *Predicted mean* (`mean`), which indicates the central tendency of the expected spillover events for each predictor value;
  - *Standard error of predicted mean* (`mean_se`), which quantifies the uncertainty associated with the predicted mean;
  - *Confidence intervals for the predicted mean* (`mean_ci_lower`, `mean_ci_upper`), which provide a range within which the true mean is expected to lie with a specified level of confidence (typically 95%).

8. **Step 8: Plot univariate GLMs** The final step involves plotting the univariate GLMs for both the full and filtered periods to visualize the relationship between climate variables and spillover events.

The complete code for this process is reported below.

Listing A.1: Univariate GLM Code

```
import statsmodels.api as sm
import pandas as pd
import numpy as np
from scipy.stats import norm

data = pd.DataFrame({           # DataFrame pandas creation
        'predictor': Var,
        'response': yearly_spill
 })

y = data['response']            # y is the response variable

data['Intercept'] = 1
X = data['Intercept', 'predictor']  # X is the explanatory variable
plus intercept

# Step 1: Estimation of dispersion parameter by fitting an auxiliary
model
model = sm.NegativeBinomial(y, X, loglike_method='nb2')
results = model.fit()           # fit the auxiliary model
alpha_opt = results.params['alpha'] # get optimal alpha

# Step 2: Check the statistical significance of alpha
print(results.summary())

# Step 3: Proceed with fitting a GLM
```

```
    glm_nbinom = sm.GLM(y, X, family = sm.families.NegativeBinomial(alpha=alpha_opt))
    glm_result = glm_nbinom.fit()   # fit the GLM of negative binomial family using alpha_opt as fixed input parameter

    # Step 4: Model evaluation
    print(glm_result.summary())
    print(glm_result.summary2())
    R_CS = glm_result.pseudo_rsquared(kind='cs')  # Cox-Snell likelihood ratio pseudo R-squared

    # Step 5: Coefficients estimates and relative p-values
    coef_estimated = glm_result.params          # Get coefficients estimates
    ci_coef_estmated = glm_result.conf_int(alpha=0.05)  # Confidence interval at 95% of the coefficient estimates

    cov_matrix = glm_result.cov_params()        # variance-covariance matrix of the estimators
    standard_errors = np.sqrt(np.diag(cov_matrix))      # standard errors are the square roots of the diagonal elements of the variance-covariance matrix

    Z_scores = coef_estimated / standard_errors      # Wald (Z-scores) statistics
    p_values = 2 * (1 - norm.cdf(np.abs(Z_scores))) # p-values of the coefficient estimates
    # or equivalently from the model result object
    # p_values_glm = glm_result.pvalues

    # Step 6: Coefficients interpretation
    coef_predictor = coef_estimated['predictor_name']
    coef_intercept = coef_estimated['Intercept']

    p_value_predictor = p_values['predictor_name']

    change_factor_predictor = np.exp(coef_predictor)
    change_factor_predictor_perc = (change_factor_predictor -1)*100

    significance = ""
    if p_value_predictor < 0.001:          # statistical significance of coefficients, denoted with asterisks
        significance = "***"
    elif p_value_predictor < 0.01:
        significance = "**"
    elif p_value_predictor < 0.05:
        significance = "*"

    # Step 7: Get prediction values and confidence intervals
    predictor_range = np.linspace(data['predictor_name'].min(), data['predictor_name'].max(), 100)      # generating a spectrum of values for the predictor, ranging from its min to max values

    predictions = glm_result.get_prediction(pd.DataFrame({'Intercept': 1, 'predictor_name': predictor_range})).summary_frame(alpha=0.05)

    pred_mu = predictions['mean']           # Expected mean
```

```
    pred_mu_se = predictions['mean_se']      # Standard error of predicted mean
    conf_int_low = predictions['mean_ci_lower']      # Lower confidence interval for the predicted mean
    conf_int_high = predictions['mean_ci_upper']     # Higher confidence interval for the predicted mean
```

GLMs: Exhaustive feature selection

The next step involves a rigorous model selection process aimed at identifying the most parsimonious model that best fits the data. This process uses Akaike's Information Criterion corrected for small sample sizes (AICc) to compare models, prioritizing those that achieve a good fit with fewer parameters.

To identify the best-fitting combination of scaled climate covariates for explaining the response variable (i.e. yearly variation in spillover events), the *exhaustive feature selection* algorithm was employed. This algorithm is a *wrapper* approach that performs a brute-force evaluation of feature subsets by considering all possible combinations of features. It starts with a baseline model that includes only the intercept and progressively considers all possible feature combinations, beginning with single-variable models and incrementally adding more variables to assess their collective impact. This step is crucial for capturing the complex relationships between the predictors and the response variable.

At each iteration, a specific feature set is considered. Before fitting a GLM with a negative binomial distribution, an optimal dispersion parameter α_{opt} must be determined to accurately model overdispersed count data. This is achieved through an iterative optimization process that minimizes the negative log-likelihood function using the *limited memory Broyden-Fletcher-Goldfarb-Shanno* (`L-BFGS-B`) algorithm. Minimizing the negative log-likelihood, rather than maximizing the log-likelihood, simplifies the optimization problem (i.e. due to optimization conventions[3]) and enhances numerical stability. The `L-BFGS-B` algorithm efficiently navigates the parameter space within specified bounds, ensuring that α remains within a plausible range (i.e. α must be non-negative). This iterative process seeks the value of α that minimizes the negative log-likelihood, starting from an initial value of 0.1 to avoid computational issues.

For each combination of features, and with the optimized α_{opt} updated at each iteration, a Generalized Linear Model using the negative binomial family is fitted to the data using `statsmodel.api` in Python. The model's goodness of fit is then assessed using the Akaike Information Criterion corrected (AICc):

$$AIC = 2p - 2ln(\ell) \tag{A.50}$$

$$AICc = AIC + \frac{2p(p+1)}{n-p-1} \tag{A.51}$$

where ℓ is the maximized value of the model's log-likelihood function, obtained as `model.llf`. The coefficient p corresponds to the number of estimated parameters in the model, including the intercept (`model.df_model + 1`), and n denotes the total number of observations (retrieved as `model.nobs`).

All tested models are then ranked in ascending order of AICc values to identify the model

---

[3] **Optimization convention** refers to the standard practice in mathematical optimization where algorithms are typically designed to minimize objective function rather than maximize it

that offers the best balance between complexity and fit, with the lowest AICc value indicating the most suitable model for explaining the response variable. The complete code employed in this analysis is here reported, see Listing A.2.

Although the exhaustive feature selection algorithm is computationally demanding, it methodically evaluates all possible feature combinations while optimizing key parameters like α. This approach ensures the development of a highly accurate and interpretable model that is both parsimonious and robust, providing valuable insights into the factors influencing the yearly variation in spillover events. The use of AICc as a selection criterion helps prevent overfitting, ensuring that the model's complexity is justified by its explanatory power.

Listing A.2: Exhaustive feature selection code

```python
import numpy as np
import statsmodels.api as sm
import itertools
import warnings
from scipy.optimize import minimize

def calculate_aicc(aic, n, k):
    """Calculates AICc given AIC, number of observations n, and number of parameters k."""
    return aic + (2 * k * (k + 1)) / (n - k - 1)

def neg_llf(alpha, y, X):
    """Calculates the negative log-likelihood for a given alpha (objective function to minimize)."""
    try:
        model = sm.GLM(y, X, family=sm.families.NegativeBinomial(alpha=alpha)).fit()
        return -model.llf
    except Exception as e:
        print(f"Error in neg_llf for alpha {alpha}: {e}")
        return np.inf

def find_optimal_alpha(y, X):
    """Finds the optimal value of alpha for the Negative Binomial model.

    Utilizes the L-BFGS-B optimization algorithm via the 'minimize' function, which is applied when bounds are specified for the parameter.

    Parameters:
    - y: The dependent variable (target outcomes).
    - X: The independent variables (predictors).

    Returns:
    - The optimal alpha value if the optimization converges successfully; otherwise, None.
    """
    result = minimize(neg_llf, x0=[0.1], args=(y, X), bounds=[(0.1, None)])
    if not result.success:
        print("Warning: Optimization failed to converge.")
        return None
    return result.x[0]

def evaluate_combinations(data, dependent_var, independent_vars):
```

```
    """
    Evaluates all combinations of independent variables using AICc criteria to select the best set.

    This function performs exhaustive feature selection with Negative Binomial models, determining the optimal alpha for each configuration.

    Parameters:
    - data: DataFrame with the dataset.
    - dependent_var: Column name of the dependent variable.
    - independent_vars: List of potential independent variables.

    Returns:
    - List of tuples with AICc value, variable combination, optimal alpha, and fitted model.
    """
    results = []
    y = data[dependent_var]
    for L in range(0, len(independent_vars) + 1):
        for subset in itertools.combinations(independent_vars, L):
            if not subset:
                continue
            X = data[list(subset)]
            X = sm.add_constant(X)  # Add a constant term for the intercept
            try:
                with warnings.catch_warnings(record=True) as w:
                    warnings.simplefilter("always")
                    alpha = find_optimal_alpha(y, X)
                    model = sm.GLM(y, X, family=sm.families.NegativeBinomial(alpha=alpha)).fit()
                    aicc = calculate_aicc(model.aic, model.nobs, len(model.params))
                    results.append((aicc, subset, alpha, model))
                    if w:
                        print(f"Warnings encountered with variables {subset}: {w}")
            except Exception as e:
                print(f"Error with variables {subset}: {e}")
                continue
    return results

    # Call the evaluate_combinations function
    results = evaluate_combinations(
    var_df_scaled,  # DataFrame containing the scaled data (explanatory variables)
    dependent_var,  # Dependent variable
    independent_vars  # List of independent variables
    )

    # Sort the results by AICc
    sorted_results = sorted(results, key=lambda x: x[0])
```

A.5.2 GLMs evaluation

Deviance and model residuals Referring to the notation from Section A.5, let $\hat{\eta} = X\hat{\beta}$ denote the estimated linear predictor. Given the relationship A.42 between μ and η, where

$\mu = E(Y) = (\mu_1, ...\mu_n)^T$ represents the expected values, the vector of predicted values is defined as $\hat{\mu} = (exp(\hat{\eta}_1), ...exp(\hat{\eta}_n))$ [117]. This follows from the exponential function serving as the inverse of the log link function used in the model. Thus, $\hat{y}$ is set equal to $\hat{\mu}$, where $\hat{\mu}_i$ represents the predicted value for the i-th year, and y_i indicates the observed value in the same year, with $i = 1, ...n$. In the ensuing discussion, the maximum likelihood estimate of the generic dispersion parameter ϕ will be denoted by $\hat{\phi}$.

In the context of a generalized linear model the *deviance* D is defined as [117], [114]:

$$D(y; \hat{\mu}) = 2\phi \left(\ell_\mu(y, \phi) - \ell_\mu(\hat{\mu}, \phi)\right) \tag{A.52}$$

where ℓ_μ is the log-likelihood function of the GLM's parameterization (μ, ϕ). The term $\ell_\mu(y, \phi)$ represents the log-likelihood of the saturated model, where each predicted value $\hat{\mu}_i$ equals the corresponding observed value y_i. This configuration results in the maximum possible log-likelihood for the data under the assumed distribution, as the saturated model has as many parameters as observations n (i.e. n = p, where p is the number of the explanatory variables), ensuring a perfect fit [117] [114]. Conversely, $\ell_\mu(\hat{\mu}, \phi)$ is the log-likelihood of the current model under evaluation, where $\hat{\mu}$ represents the predicted means calculated from the model parameters. Typically, this model has fewer parameters than the saturated model (i.e. p < n).
Deviance quantifies the loss of fit when moving from the saturated model to a more parsimonious model with fewer explanatory variables. Consequently, high deviance values indicate a substantial disparity between the predicted values ($\hat{\mu}_i$) and the observed values (y_i), indicating that the simpler model may not adequately capture the data's underlying structure [117].

For the negative binomial distribution with the dispersion α fixed at the maximum likelihood estimate $\hat{\alpha}$, the deviance simplifies as follows [117], [114]:

$$D(y; \hat{\mu}) = 2\sum_{i=1}^{n} \left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i + \hat{\alpha}) \ln\left(\frac{\hat{\alpha} + y_i}{\hat{\alpha} + \hat{\mu}_i}\right) \right) \tag{A.53}$$

This deviance metric is instrumental in defining a specific type of residuals known as *deviance residuals*, which, along with *raw residuals* and *Pearson residuals*, are crucial for conducting a comprehensive diagnostic analysis of the fitted model. These residuals collectively facilitate the evaluation of the model's performance by assessing the fit's adequacy and the validity of the underlying statistical assumptions. The types of residuals used in model diagnostics are as follows [117], [114]:

- **Raw or response residuals** represent the simple difference between the observed and predicted values:

  $$r_i = y_i - \hat{\mu}_i, \quad i = 1, \ldots, n \tag{A.54}$$

  They can be obtained using the `.resid_response` attribute in `statsmodel` results. These residuals are direct and useful for straightforward evaluations, such as identifying outliers and assessing general fit.

- **Pearson residuals** are standardised residuals, calculated by dividing the raw residuals by the standard deviation of the predicted responses:

  $$r_i^P = \frac{r_i}{\sqrt{\sigma^2(\hat{\mu}_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \alpha\hat{\mu}_i^2}} \tag{A.55}$$

In this analysis, they are retrieved using the `.resid_pearson` attribute. Pearson residuals are used to assess whether the residual variance is constant (i.e. homoscedasticity).

- **Deviance residuals** are specific to GLMs, and measure the discrepancy between observed and predicted values in terms of deviance:

  $$r_i^D = \text{sgn}(y_i - \hat{\mu}_i)\sqrt{D_i}, \quad i = 1, \ldots, n \tag{A.56}$$

  where D_i represents the contribution of a single observation to the deviance. These residuals, accessible via the `.resid_deviance` attribute, are useful to assessing the contribution of each observation to the overall deviance, providing insight into individual influence and leverage within the model.

Outlier, leverage and influence diagnostics Most standard diagnostics for linear models extend straightforwardly to GLMs [144], leveraging maximum likelihood and quasi-likelihood estimates obtained through iteratively weighted least squares (IWLS). These diagnostics include *hat-values* and *Cook's distances* [144].

Hat-values or leverages In a linear model, the so called *Hat matrix* allows to transform the observations y to their estimated values $\hat{y}$ as follows [145]:

$$\hat{y} = X\hat{b} = X(X'X)^{-1}X'y = Hy$$

The diagonal elements of H, h_{ii} are called *leverages* and satisfy:

$$0 \leq h_{ii} \leq 1; \quad \sum_{i=1}^{n} h_{ii} = k$$

where k is the number of coefficients in the regression model, and n is the number of observations. Since the predicted response can be written as:

$$\hat{y}_i = h_{i1}y_1 + h_{i2}y_2 + \ldots + h_{ii}y_i + \ldots + h_{in}y_n \quad \text{for } i = 1,2, ..n$$

the leverage, h_{ii}, quantifies the influence of the observation y_i on its predicted value $\hat{y}_i$ [145]. Generally, a larger h_{ii} indicates that the i^{th} data point is further from the mean of the predictor values (i.e., the center of the predictor space), giving it more leverage [145]. An observation i can be considered an *outlier* if its leverage substantially exceeds the mean leverage value $\bar{h}$:

$$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n} = k/n$$

where k is the sum of the leverage values. A typical cut-off value is $h_{cut-off} = 2 \cdot k/n$, so that an observation x in the *predictor space* is flagged as unusual if its leverage $h_{ii} > h_{cut-off}$.

In GLMs, hat-values h_i can be obtained directly from the final iteration of the IWLS procedure, and have the same interpretation as in linear models, although in GLMs, they depend on both the response variable Y and the predictor variables X [144]. The generalised *Hat matrix* is defined as:

$$H = W^{1/2}X(X'WX)^{-1}X'W^{1/2} \tag{A.57}$$

where W is the weight matrix from the final IWLS iteration, a $n \times n$ diagonal matrix with $\frac{\mu_i}{1+\alpha\mu_i}$ is the i^{th} diagonal element [114].

Influence measure: Cook's distance Cook's distance D_i measures the normalised change in fitted response values $\hat{y}_j$ resulting from the exclusion of observation i, thus showing the influence of each observation on the fitted response values [146].

$$D_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{k\,MSE}$$

where $\hat{y}_{j(i)}$ is the j^{th} fitted response value, where the fit does not include observation i, and MSE is the mean square error [146]. This equation can be also be expressed as:

$$D_i = \frac{r_i^2}{k\,MSE}\left(\frac{h_{ii}}{(1-h_{ii})^2}\right)$$

where r_i is the i^{th} residual, and h_{ii} is the i^{th} leverage value [146]. A data point is flagged as *Influential point* if it has a large D_i, indicating that it strongly influences the fitted values.

A generalization of the Cook's distances is also possible [144]:

$$D_i = \frac{(r_{i,st}^P)^2}{\hat{\alpha}\cdot p}\cdot\frac{h_i}{1-h_i} \tag{A.58}$$

where $r_{i,st}^P$ is the *Standardised Pearson residuals* retrieved from equation A.55 of the Pearson residual as follows [114]:

$$r_{i,st}^P = \frac{r_i^P}{\sqrt{1-h_i}} \tag{A.59}$$

Cook's distance is typically compared against a predetermined cut-off value, $D_{cut-off}$, often chosen as $D_{cut-off} = 2\cdot\overline{D}$, where $\overline{D}$ is the mean of Cook's distances across all observations. Therefore, if the Cook's distance D_i of an observation i exceeds $D_{\text{cut-off}}$, then it is flagged as having a significant influence on the regression model's parameters.

A.5.3 GLMs results - Case A: single spatial cluster

Figure A.5 displays the Pairwise Pearson's correlation matrix between explanatory variables for the unique cluster.

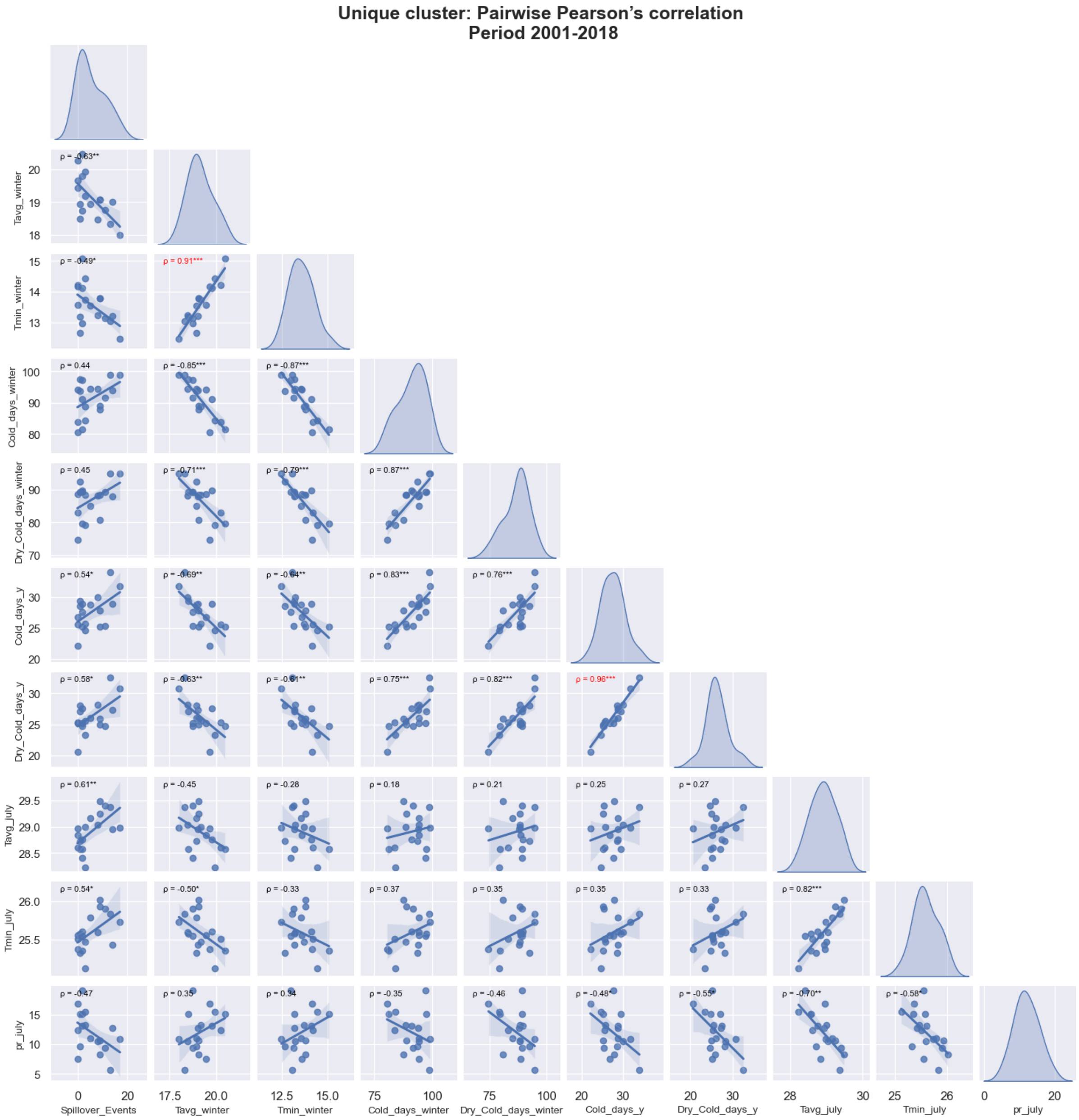


Figure A.5: *Case A - Unique cluster:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variables (i.e. mean temperature, minimum temperature, percentage of cold days, percentage of dry days followed by cold nights), annual indexes (i.e. annual dry days, annual cold days, annual dry days followed by cold nights) and July variables (i.e. mean temperature, minimum temperature and precipitation). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***).

Inclusion of interaction factors

The five top-performing models $M_j^{(0)}$ ($j = 1,2,..5$), each with a $\Delta AICc < 2$ compared to the best reference model M_{ref} (i.e. the top one $M_1^{(0)}$), were selected from the outcome of the exhaustive feature selection process, as shown in Table A.5. Interaction terms were then systematically added to these models, as described in Section 5.1.2 and detailed in Table A.6, to explore potential synergies between variables, with the goal of enhancing the models' explanatory power.

For each selected model $M_j^{(0)}$, a set of n interaction-enhanced models $M_j^{(l)}$ was tested,

Models $M_j^{(0)}$ with $\Delta AICc < 2$					
Model name	**AICc**	**pseudo-R^2**	**Explanatory variables**	α_{opt}	$\Delta AICc$
$M_1^{(0)}$	95.60	0.75	$T_{avg,winter}, T_{avg,july}$	0.225	0.0
$M_2^{(0)}$	96.27	0.84	$DCD_y, T_{avg,july}, Pr_{july}$	0.149	0.67
$M_3^{(0)}$	96.56	0.67	$DCD_y, T_{avg,july}$	0.305	0.96
$M_4^{(0)}$	97.15	0.47	$T_{avg,winter}$	0.502	1.56
$M_5^{(0)}$	97.37	0.65	$CD_{winter}, T_{avg,july}$	0.309	1.78

Table A.5: Outcome of the exhaustive feature selection process: the top five models with $\Delta AICc < 2$, ranked by ascending AICc values, along with their associated statistics.

where $l \in \{0,3,9\}$ corresponds to the number of candidate models depending on the number of explanatory variables (k) in the model:

- If $k = 1$, no interaction terms were added, and no new models were tested, so $l = 0$; the original $M_j^{(0)}$ was retained.
- If $k = 2$, one interaction term was added ($X_1 \cdot X_2$), and three new models were tested, so $l = 3$ (see "Candidate model" column in Table A.6 for further details).
- If $k = 3$, three interaction terms were added, and nine new models were tested, so $l = 9$ (refer to the "Candidate model" column in Table A.6).

Diagnostic plots

Top model without interaction terms To assess the robustness and validity of the top model without interaction terms, various diagnostic plots were generated (see Figure A.6). These diagnostics confirmed that the model largely satisfies the key assumptions of linearity, independence, and homoscedasticity of residuals. However, three specific years—3, 4, and 12—showed deviations that warrant further investigation. Specifically:

- **Linearity:** The observed versus predicted plot aligns well with the diagonal, indicating that the model accurately predicts the majority of the observations. However, years 3 and 12 show noticeable deviations from this line, suggesting potential misfit for those specific years.
- **Independence:** The deviance residuals exhibit no clear patterns, supporting the assumption of independence among observations. This is further corroborated by the Durbin-Watson test, which shows no significant autocorrelation.
- **Homoscedasticity:** Most Pearson residuals, except for those corresponding to years 3 and 4, fall within the ±2 standard deviation range, suggesting consistent variance across residuals.
- **Normality of deviance residuals:** The Q-Q plot and the high p-value from the Shapiro-Wilk test suggest that the deviance residuals are normally distributed with no substantial outliers.

Candidate models $M_j^{(l)}$ with $j = 1,2,\ldots 5$ and $l \in \{0,3,9\}$			
Starting model	**Explanatory variables**	**Candidate models**	**Results: Ranked models**
$\boldsymbol{M_1^{(0)}}$	$X_1 = T_{avg,winter}$ $X_2 = T_{avg,july}$	1) $M_1^{(1)} = X_1 + X_1 \cdot X_2$ 2) $M_1^{(2)} = X_2 + X_1 \cdot X_2$ 3) $M_1^{(3)} = X_1 + X_2 + X_1 \cdot X_2$	$\boldsymbol{M_1^{(2)*}, \quad AICc = 92.85}$ $\boldsymbol{M_1^{(0)}, \quad AICc = 95.60}$ $M_1^{(3)}, \quad AICc = 96.15$ $M_1^{(1)}, \quad AICc = 99.04$
$\boldsymbol{M_2^{(0)}}$	$X_1 = DCD_y$ $X_2 = T_{avg,july}$ $X_3 = Pr_{july}$	$M_2^{(1)} = X_1 + X_2 + X_1 \cdot X_3$ $M_2^{(2)} = X_1 + X_2 + X_2 \cdot X_3$ $M_2^{(3)} = X_1 + X_2 \cdot X_1 + X_3$ $M_2^{(4)} = X_1 + X_2 \cdot X_3 + X_3$ $M_2^{(5)} = X_1 \cdot X_2 + X_2 + X_3$ $M_2^{(6)} = X_1 \cdot X_3 + X_2 + X_3$ $M_2^{(7)} = X_1 + X_2 + X_3 + X_1 \cdot X_2$ $M_2^{(8)} = X_1 + X_2 + X_3 + X_1 \cdot X_3$ $M_2^{(9)} = X_1 + X_2 + X_3 + X_2 \cdot X_3$	$\boldsymbol{M_2^{(1)*}, \quad AICc = 93.46}$ $\boldsymbol{M_2^{(0)}, \quad AICc = 96.26}$ $M_2^{(6)}, \quad AICc = 96.87$ $M_2^{(8)}, \quad AICc = 96.93$ $M_2^{(5)}, \quad AICc = 97.51$ $M_2^{(2)}, \quad AICc = 98.74$ $M_2^{(3)}, \quad AICc = 99.57$ $M_2^{(7)}, \quad AICc = 100.04$ $M_2^{(9)}, \quad AICc = 100.19$ $M_2^{(4)}, \quad AICc = 102.98$
$\boldsymbol{M_3^{(0)}}$	$X_1 = DCD_y$ $X_2 = T_{avg,july}$	1) $M_3^{(1)} = X_1 + X_1 \cdot X_2$ 2) $M_3^{(2)} = X_2 + X_1 \cdot X_2$ 3) $M_3^{(3)} = X_1 + X_2 + X_1 \cdot X_2$	$\boldsymbol{M_3^{(0)}, \quad AICc = 96.56}$ $M_3^{(2)}, \quad AICc = 97.04$ $M_3^{(1)}, \quad AICc = 97.69$ $M_3^{(3)}, \quad AICc = 99.91$
$\boldsymbol{M_4^{(0)}}$	$X = T_{avg,winter}$	-	$\boldsymbol{M_4^{(0)}, \quad AICc = 97.15}$
$\boldsymbol{M_5^{(0)}}$	$X_1 = CD_{winter}$ $X_2 = T_{avg,july}$	1) $M_5^{(1)} = X_1 + X_1 \cdot X_2$ 2) $M_5^{(2)} = X_2 + X_1 \cdot X_2$ 3) $M_5^{(3)} = X_1 + X_2 + X_1 \cdot X_2$	$\boldsymbol{M_5^{(2)*}, \quad AICc = 95.64}$ $\boldsymbol{M_5^{(1)*}, \quad AICc = 95.68}$ $\boldsymbol{M_5^{(0)}, \quad AICc = 97.38}$ $M_5^{(3)}, \quad AICc = 98.96$

Table A.6: Results of model improvement efforts incorporating interaction factors for each candidate model selected based on the exhaustive model selection process, with a selection criterion of $\Delta AICc < 2$. Models identified as improvements over the original baseline model (highlighted in bold) are indicated in blue.

In summary, while the overall diagnostic checks support the validity of the model, the identified anomalies associated with the years 3, 4, and 12 need further investigation to ensure the model's robustness and accuracy.

Referring to Figure A.7, years 3 and 4 show high standardized residuals (i.e., exceed the threshold +2), indicating that the model underpredicts spillover events for these years. Despite this, these years exhibit low leverage, meaning they do not have unusual predictor values that would unduly influence the model. However, their Cook's distance values are above the threshold, indicating that even though their leverage is low, they still significantly impact the regression coefficients. This suggests that the observed data for these years do not fit the model well but do not distort the model's overall structure.

On the other hand, year 12 presents a different situation. While it does not exhibit problematic standardized residuals, it has high leverage, meaning its predictor values are more extreme compared to the rest of the dataset. This high leverage, combined with its Cook's distance, suggests that year 12 exerts a substantial influence on the model's parameters. Year 12 coincides with an extreme negative winter temperature anomaly (refers to Figure 3.4), which could explain its significant impact on spillover dynamics. This suggests that the model may not fully capture the complexities associated with such climatic anomalies.

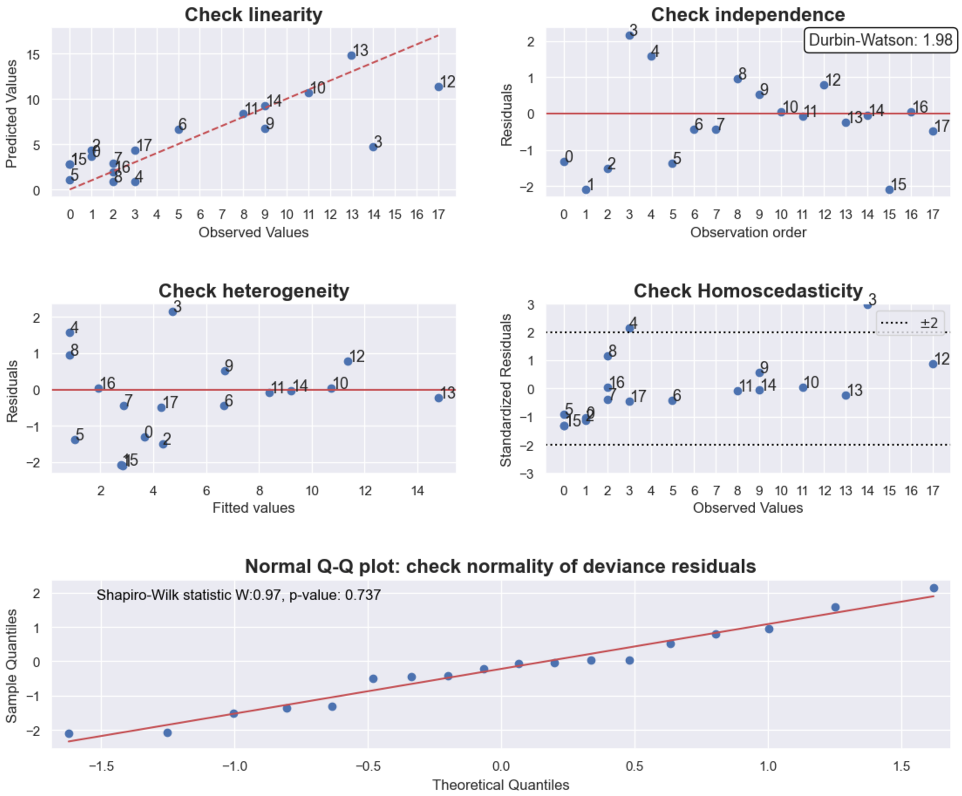


Figure A.6: Diagnostic plots assessing the validity and verifying the assumptions of the top GLM model without interaction terms.

Therefore, while the model remains valid and robust overall, further refinements—such as incorporating interaction terms or addressing non-linearities—could improve its ability to account for these specific anomalies, particularly in extreme climatic conditions.

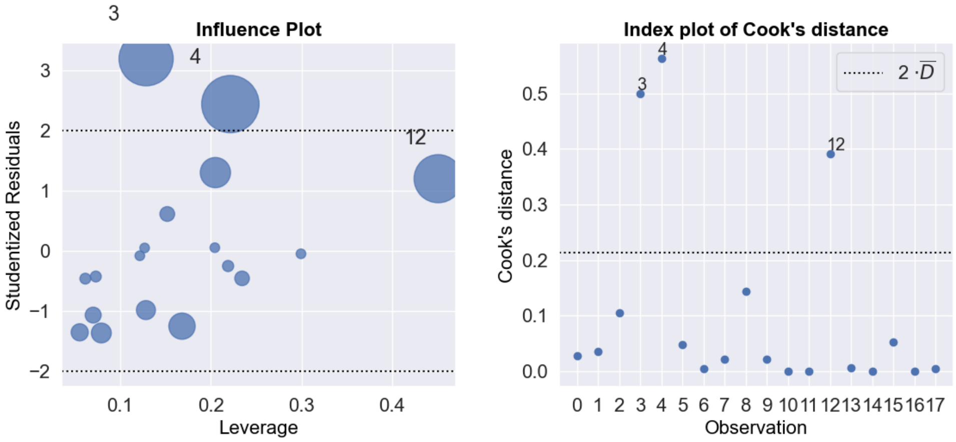


Figure A.7: Diagnostic analysis of the best-fit negative binomial GLM: (Left) enhanced influence plot with scatter point sizes proportional to Cook's distance, and (Right) index plot of Cook's distance, both with indicated thresholds for identifying outliers and influential points, respectively.

A.5.4 GLMs results - Case B: two spatial clusters

Figures A.8 and A.9 display the Pairwise Pearson's correlation matrices between explanatory variables for Cluster 1 and Cluster 2, respectively.

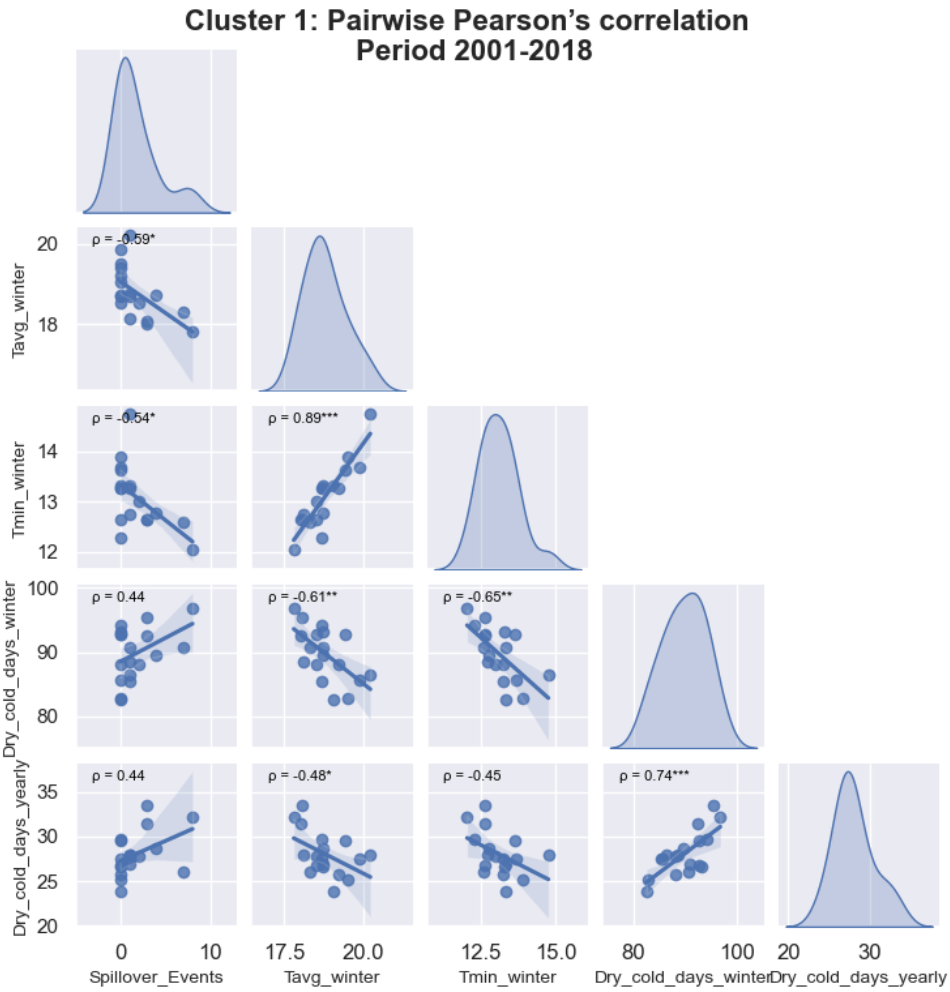


Figure A.8: *Case B - Cluster 1:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variables (i.e. mean temperature, minimum temperature, percentage of cold days, percentage of dry days followed by cold nights), annual index (i.e. annual dry days followed by cold nights) and July variable (i.e. mean temperature). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***).

Inclusion of interaction factors

The outcome of the exhaustive feature selection process for Cluster 1 and Cluster 2 is reported in Table A.7. Table A.8 presents the initial models (from the exhaustive feature selection process), the interaction terms introduced, the procedure for incorporating these terms, and the corresponding candidate models tested to evaluate whether combining variables could improve the models' explanatory power by revealing synergies (see the procedure detailed in Section A.5.3).

Figure A.9: *Case B - Cluster 2:* Pairwise Pearson's correlation between yearly Nipah spillover events, and winter climate variable (i.e. mean temperature), annual indexes (i.e. annual dry days, annual cold days, annual dry days followed by cold nights) and July variables (i.e. mean temperature, minimum temperature and precipitation). Correlations with asterisks are statistically significant at the 0.05 level (*), 0.01 level (**) and 0.001 level (***).

Diagnostic plots

Top model of Cluster 1 Figure A.10 presents the diagnostic plots used to assess the validity and check the assumptions of the top GLM model $M_{ref,C1}$.

Top models of Cluster 2 Figure A.11 shows the diagnostic plots used to assess the validity and check the assumptions of the top GLM model $M_{ref,C2}$.

Cluster 1: Models $M^{(0)}_{j,C1}$ **with** $\Delta AICc < 2$				
Model name	**AICc**	**Explanatory variables**	α_{opt}	$\Delta AICc$
$M^{(0)}_{1,C1}$	59.07	$T_{avg,winter}$	0.325	0.0
$M^{(0)}_{2,C1}$	61.62	$T_{min,winter}$	0.53	2.56
Cluster 2: Models $M^{(0)}_{j,C2}$ **with** $\Delta AICc < 2$				
Model name	**AICc**	**Explanatory variables**	α_{opt}	$\Delta AICc$
$M^{(0)}_{1,C2}$	80.57	$T_{avg,july}, DD_y$	0.132	0.0
$M^{(0)}_{2,C2}$	81.33	$T_{avg,july}, DD_y, DCD_y$	0.1	0.76
$M^{(0)}_{3,C2}$	81.91	Pr_{july}, DD_y	0.11	1.34
$M^{(0)}_{4,C2}$	82.45	$T_{avg,winter}, T_{avg,july}, DD_y$	0.1	1.88
$M^{(0)}_{5,C2}$	82.49	$T_{avg,july}$	0.30	1.92

Table A.7: Outcome of the model selection process for Cluster 1 and 2: top models ranked by ascending AICc values, with associated statistics reported. In Cluster 1 only one model presents a $\Delta AICc < 2$ (highlighted in blue).

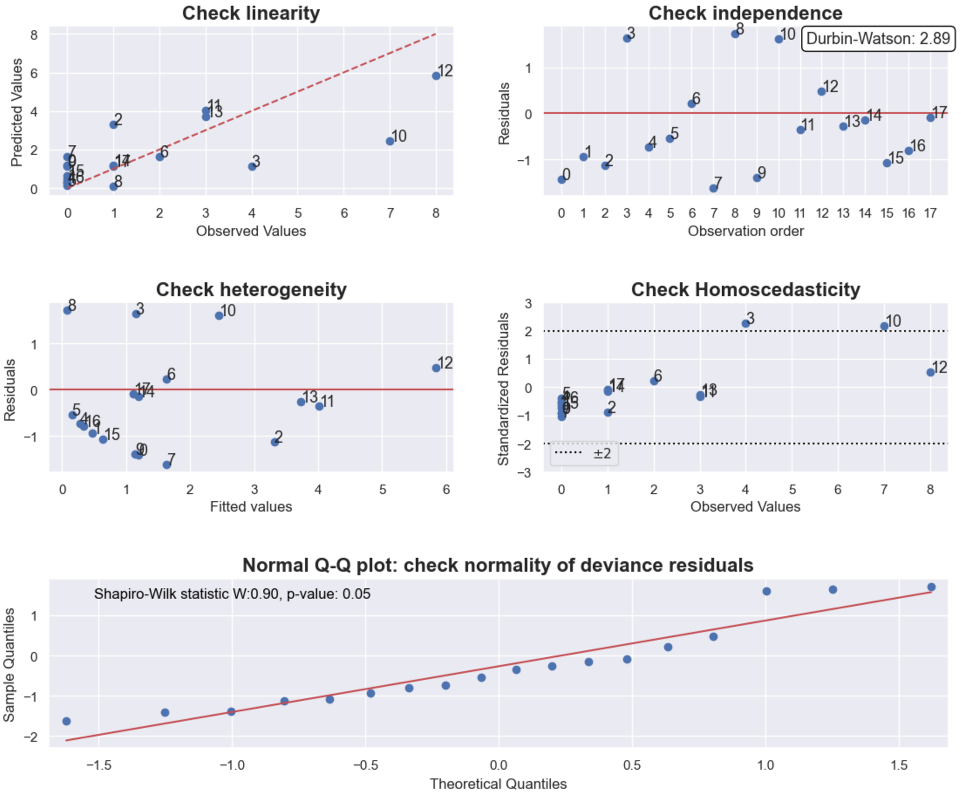


Figure A.10: Cluster 1: Diagnostic plots to evaluate the best reference model's validity and check GLM assumptions.

Cluster C1: Candidate models $M^{(l)}_{j,C1}$ with j = 1,2, ...5 and $l \in \{0,3,9\}$			
Starting model	**Explanatory variables**	**Candidate models**	**Results: Ranked models**
$\boldsymbol{M^{(0)}_{1,C1}}$	$X = T_{avg,winter}$	-	$\boldsymbol{M^{(0)}_{1,C1}}$, $\boldsymbol{AICc = 59.07}$
Cluster C2: Candidate models $M^{(l)}_{j,C1}$ with j = 1,2, ...5 and $l \in \{0,3,9\}$			
Starting model	**Explanatory variables**	**Candidate models**	**Results: Ranked models**
$\boldsymbol{M^{(0)}_{1,C2}}$	$X_1 = T_{avg,july}$ $X_2 = DD_y$	1) $M^{(1)}_{2,C2} = X_1 + X_1 \cdot X_2$ 2) $M^{(2)}_{2,C2} = X_2 + X_1 \cdot X_2$ 3) $M^{(3)}_{2,C2} = X_1 + X_2 + X_1 \cdot X_2$	$\boldsymbol{M^{(0)}_{1,C2}}$, $\boldsymbol{AICc = 80.57}$ $M^{(2)}_{2,C1}$, $AICc = 80.96$ $M^{(1)}_{2,C1}$, $AICc = 81.07$ $M^{(3)}_{2,C1}$, $AICc = 83.90$
$\boldsymbol{M^{(0)}_{2,C2}}$	$X_1 = T_{avg,july}$ $X_2 = DD_y$ $X_3 = DCD_y$	$M^{(1)}_{2,C2} = X_1 + X_2 + X_1 \cdot X_3$ $M^{(2)}_{2,C2} = X_1 + X_2 + X_2 \cdot X_3$ $M^{(3)}_{2,C2} = X_1 + X_2 \cdot X_1 + X_3$ $M^{(4)}_{2,C2} = X_1 + X_2 \cdot X_3 + X_3$ $M^{(5)}_{2,C2} = X_1 \cdot X_2 + X_2 + X_3$ $M^{(6)}_{2,C2} = X_1 \cdot X_3 + X_2 + X_3$ $M^{(7)}_{2,C2} = X_1 + X_2 + X_3 + X_1 \cdot X_2$ $M^{(8)}_{2,C2} = X_1 + X_2 + X_3 + X_1 \cdot X_3$ $M^{(9)}_{2,C2} = X_1 + X_2 + X_3 + X_2 \cdot X_3$	$\boldsymbol{M^{(0)}_{2,C2}}$, $\boldsymbol{AICc = 81.33}$ $M^{(2)}_{2,C2}$, $AICc = 81.65$ $M^{(4)}_{2,C2}$, $AICc = 81.97$ $M^{(1)}_{2,C2}$, $AICc = 82.06$ $M^{(5)}_{2,C2}$, $AICc = 82.29$ $M^{(3)}_{2,C2}$, $AICc = 82.44$ $M^{(6)}_{2,C2}$, $AICc = 83.74$ $M^{(9)}_{2,C2}$, $AICc = 84.89$ $M^{(8)}_{2,C2}$, $AICc = 85.15$ $M^{(7)}_{2,C2}$, $AICc = 85.16$
$\boldsymbol{M^{(0)}_{3,C2}}$	$X_1 = Pr_{july}$ $X_2 = DD_y$	1) $M^{(1)}_{3,C2} = X_1 + X_1 \cdot X_2$ 2) $M^{(2)}_{3,C2} = X_2 + X_1 \cdot X_2$ 3) $M^{(3)}_{3,C2} = X_1 + X_2 + X_1 \cdot X_2$	$\boldsymbol{M^{(2)}_{3,C2}}$, $\boldsymbol{AICc = 80.77}$ $\boldsymbol{M^{(0)}_{3,C2}}$, $\boldsymbol{AICc = 81.91}$ $M^{(1)}_{3,C2}$, $AICc = 84.00$ $M^{(3)}_{3,C2}$, $AICc = 84.11$
$\boldsymbol{M^{(0)}_{4,C2}}$	$X_1 = T_{avg,winter}$ $X_2 = T_{avg,july}$ $X_3 = DD_y$	$M^{(1)}_{4,C2} = X_1 + X_2 + X_1 \cdot X_3$ $M^{(2)}_{4,C2} = X_1 + X_2 + X_2 \cdot X_3$ $M^{(3)}_{4,C2} = X_1 + X_2 \cdot X_1 + X_3$ $M^{(4)}_{4,C2} = X_1 + X_2 \cdot X_3 + X_3$ $M^{(5)}_{4,C2} = X_1 \cdot X_2 + X_2 + X_3$ $M^{(6)}_{4,C2} = X_1 \cdot X_3 + X_2 + X_3$ $M^{(7)}_{4,C2} = X_1 + X_2 + X_3 + X_1 \cdot X_2$ $M^{(8)}_{4,C2} = X_1 + X_2 + X_3 + X_1 \cdot X_3$ $M^{(9)}_{4,C2} = X_1 + X_2 + X_3 + X_2 \cdot X_3$	$\boldsymbol{M^{(3)}_{4,C2}}$, $\boldsymbol{AICc = 81.94}$ $\boldsymbol{M^{(6)}_{4,C2}}$, $\boldsymbol{AICc = 82.07}$ $\boldsymbol{M^{(0)}_{4,C2}}$, $\boldsymbol{AICc = 82.45}$ $M^{(4)}_{4,C2}$, $AICc = 83.17$ $M^{(2)}_{4,C2}$, $AICc = 83.25$ $M^{(5)}_{4,C2}$, $AICc = 83.57$ $M^{(1)}_{4,C2}$, $AICc = 84.37$ $M^{(7)}_{4,C2}$, $AICc = 85.75$ $M^{(8)}_{4,C2}$, $AICc = 85.82$ $M^{(9)}_{4,C2}$, $AICc = 86.36$
$\boldsymbol{M^{(0)}_{5,C2}}$	$X = T_{avg,july}$	-	$\boldsymbol{M^{(0)}_{5,C2}}$, $\boldsymbol{AICc = 82.49}$

Table A.8: Results of model improvement efforts for Cluster 1 and 2, incorporating interaction factors for each candidate model selected based on the exhaustive model selection process, with a selection criterion of $\Delta AICc < 2$. Models identified as improvements over the original baseline model (highlighted in bold) are indicated in blue.

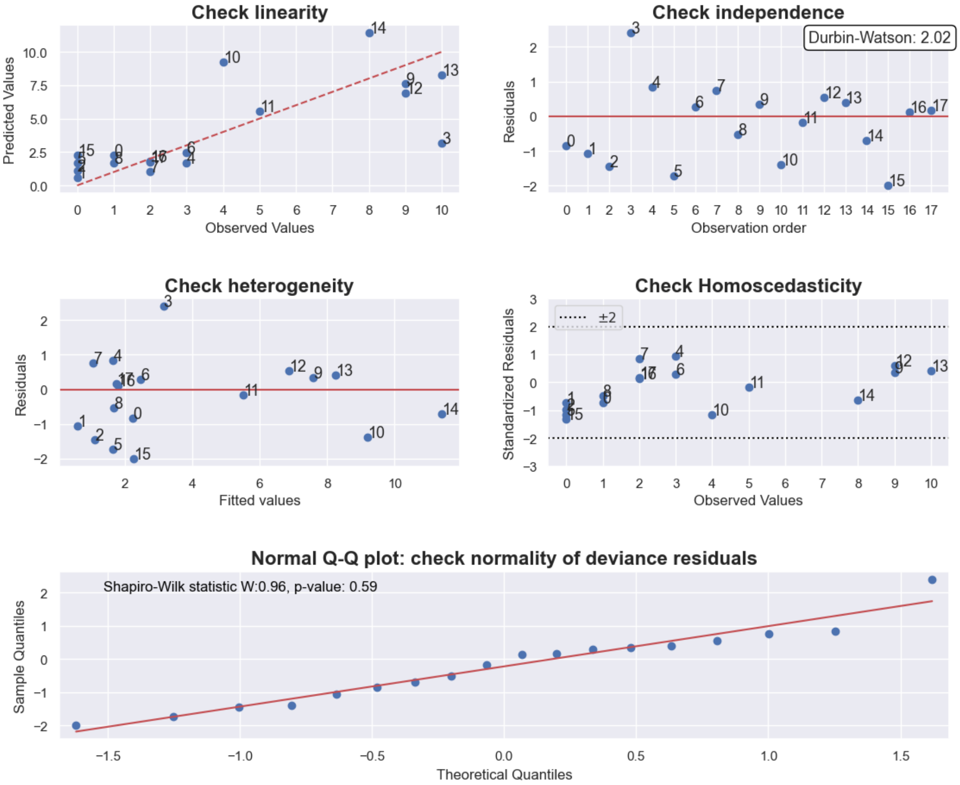


Figure A.11: **Cluster 2:** Diagnostic plots to evaluate the best reference model's validity and check GLM assumptions.

Bibliography

[1] S. Guth, N. Mollentze, K. Renault, D. G. Streicker, E. Visher, M. Boots, and C. E. Brook. «Bats host the most virulent-but not the most dangerous-zoonotic viruses». In: *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 119.14 (2022), e2113628119. DOI: `10.1073/pnas.2113628119` (cit. on pp. xvi, 1, 24).

[2] National Institutes of Health (US) and Biological Sciences Curriculum Study. *Understanding Emerging and Re-emerging Infectious Diseases*. Bethesda, MD: National Institutes of Health (US), 2007 (cit. on p. xvi).

[3] R.K. Plowright et al. «Ecological dynamics of emerging bat virus spillover». In: *Proc Biol Sci* 282.1798 (2015). DOI: `10.1098/rspb.2014.2124` (cit. on pp. xvi, 7, 10, 12–17).

[4] Raina K. Plowright, Colin R. Parrish, Hamish McCallum, Peter J. Hudson, Albert I. Ko, Andrea L. Graham, and James O. Lloyd-Smith. «Pathways to zoonotic spillover». In: *Nature Reviews Microbiology* 15 (2017), pp. 502–510. DOI: `10.1038/nrmicro.2017.45`. URL: `https://doi.org/10.1038/nrmicro.2017.45` (cit. on pp. xvi, 3, 4).

[5] J. Louten. «Virus Transmission and Epidemiology». In: *Essential Human Virology* (2016). Epub 2016 May 6, pp. 71–92. DOI: `10.1016/B978-0-12-800947-5.00005-3` (cit. on p. xvi).

[6] C.J. Burrell, C.R. Howard, and F.A. Murphy. «Epidemiology of Viral Infections». In: *Fenner and White's Medical Virology* (2017). Epub 2016 Nov 11, pp. 185–203. DOI: `10.1016/B978-0-12-375156-0.00013-8` (cit. on p. xvi).

[7] Barbara A. Han, Andrew M. Kramer, and John M. Drake. «Global Patterns of Zoonotic Disease in Mammals». In: *Trends in Parasitology* 32.7 (2016). Epub 2016 Jun 14, pp. 565–577. DOI: `10.1016/j.pt.2016.04.007`. URL: `https://doi.org/10.1016/j.pt.2016.04.007` (cit. on p. xvi).

[8] Claire Heffernan. «Climate change and multiple emerging infectious diseases». In: *The Veterinary Journal* 234 (2018). Epub 2018 Jan 1, pp. 43–47. DOI: `10.1016/j.tvjl.2017.12.021` (cit. on pp. 1, 102).

[9] Kate E. Jones, Nikkita G. Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L. Gittleman, and Peter Daszak. «Global trends in emerging infectious diseases». In: *Nature* 451 (2008), pp. 990–993. DOI: `10.1038/nature06536`. URL: `https://doi.org/10.1038/nature06536` (cit. on p. 1).

[10] Ruwini Rupasinghe, Bruno B. Chomel, and Beatriz Martínez-López. «Climate change and zoonoses: A review of the current status, knowledge gaps, and future trends». In: *Acta Tropica* 226 (2022), p. 106225. ISSN: 0001-706X. DOI: `10.1016/j.actatropica.2021.106225`. URL: `https://www.sciencedirect.com/science/article/pii/S0001706X21004034` (cit. on pp. 1, 20, 23, 24).

[11] Reuters. *Jump zones: How a pandemic might start.* `https://www.reuters.com/investigates/special-report/global-pandemic-bats-jumpzones/`. 2022 (cit. on pp. 1, 5, 10).

[12] C.F.J. Meyer, M.J. Struebig, and M.R. Willig. «Responses of Tropical Bats to Habitat Fragmentation, Logging, and Deforestation». In: *Bats in the Anthropocene: Conservation of Bats in a Changing World.* Ed. by C. Voigt and T. Kingston. Cham: Springer, 2016, pp. 63–103. DOI: `10.1007/978-3-319-25220-9_4` (cit. on pp. 1, 6, 11, 13).

[13] C.J. Carlson, G.F. Albery, C. Merow, et al. «Climate change increases cross-species viral transmission risk». In: *Nature* 607.7917 (2022), pp. 555–562. DOI: `10.1038/s41586-022-04788-w` (cit. on pp. 1, 3, 4, 22, 23, 26).

[14] F. Festa et al. «Bat responses to climate change: a systematic review». In: *Biological Reviews of the Cambridge Philosophical Society* 98.1 (2023), pp. 19–33. DOI: `10.1111/brv.12893` (cit. on pp. 1, 25, 28).

[15] P. Devnath and H.M.A.A. Masud. «Nipah virus: a potential pandemic agent in the context of the current severe acute respiratory syndrome coronavirus 2 pandemic». In: *New Microbes and New Infections* 41 (2021). DOI: `10.1016/j.nmni.2021.100873` (cit. on pp. 1, 6, 7, 35).

[16] S. Gazal, N. Sharma, S. Gazal, M. Tikoo, D. Shikha, G.A. Badroo, M. Rashid, and S.J. Lee. «Nipah and Hendra Viruses: Deadly Zoonotic Paramyxoviruses with the Potential to Cause the Next Pandemic». In: *Pathogens* 11.12 (2022). DOI: `10.3390/pathogens11121419` (cit. on pp. 1, 6, 7, 17, 35).

[17] A. Latinne and S. Morand. «Climate Anomalies and Spillover of Bat-Borne Viral Diseases in the Asia-Pacific Region and the Arabian Peninsula». In: *Viruses* 14.5 (2022). DOI: `10.3390/v14051100` (cit. on pp. 1, 6, 23, 41, 106).

[18] World Health Organization. *Nipah Virus.* `https://www.who.int/news-room/fact-sheets/detail/nipah-virus`. 2018 (cit. on pp. 1, 7, 8).

[19] Nathan D. Wolfe, Claire P. Dunavan, and Jared Diamond. «ORIGINS OF MAJOR HUMAN INFECTIOUS DISEASES». In: *Improving Food Safety Through a One Health Approach: Workshop Summary.* Ed. by Institute of Medicine (US). Available from: `https://www.ncbi.nlm.nih.gov/books/NBK114494/`. Washington (DC): National Academies Press (US), 2012. Chap. A16 (cit. on p. 3).

[20] Ian Smith and Lin-Fa Wang. «Bats and their virome: an important source of emerging viruses capable of infecting humans». In: *Current Opinion in Virology* 3.1 (2013). Epub 2012 Dec 21, pp. 84–91. DOI: `10.1016/j.coviro.2012.11.006` (cit. on p. 3).

[21] M. Letko, S.N. Seifert, K.J. Olival, et al. «Bat-borne virus diversity, spillover and emergence». In: *Nat Rev Microbiol* 18.6 (2020), pp. 461–471. DOI: `10.1038/s41579-020-0394-z` (cit. on pp. 4, 5, 8).

[22] M.K. Kessler et al. «Changing resource landscapes and spillover of henipaviruses». In: *Ann N Y Acad Sci* 1429.1 (2018), pp. 78–99. DOI: `10.1111/nyas.13910` (cit. on pp. 4, 5, 8, 10–16, 18, 26).

[23] T. Schountz. «Immunology of bats and their viruses: challenges and opportunities». In: *Viruses* 6.12 (2014), pp. 4880–4901. DOI: `10.3390/v6124880` (cit. on p. 4).

[24] Valeria Gonzalez and Abhishek Banerjee. «Molecular, ecological, and behavioral drivers of the bat-virus relationship». In: *iScience* 25.8 (2022). Epub 2022 Jul 20, p. 104779. DOI: `10.1016/j.isci.2022.104779`. URL: `https://doi.org/10.1016/j.isci.2022.104779` (cit. on p. 4).

[25] Tigga Kingston. «Response of Bat Diversity to Forest Disturbance in Southeast Asia – Insights from Long-Term Research in Malaysia». In: *Bat Evolution, Ecology and Conservation*. Ed. by R.A. Adams and S.C. Pedersen. Springer Science Press, 2013, pp. 169–195. DOI: `10.13140/2.1.2121.3125` (cit. on p. 5).

[26] R.C. Estoque, M. Ooba, V. Avitabile, et al. «The future of Southeast Asia's forests». In: *Nat Commun* 10 (2019). DOI: `10.1038/s41467-019-09646-4` (cit. on pp. 5, 20).

[27] Reuters. *Global Pandemic: Bats*. 2023. URL: `https://www.reuters.com/investigates/special-report/global-pandemic-bats-methodology/` (cit. on p. 5).

[28] World Health Organization. *Prioritizing diseases for research and development in emergency contexts*. `https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts`. 2022 (cit. on p. 5).

[29] Richard J. Hatchett. *Nipah Virus – A Continuing Public Health Threat*. `https://cepi.net/nipah`. Richard J. Hatchett, CEO of the Coalition for Epidemic Preparedness Innovations (CEPI). 2022 (cit. on p. 6).

[30] A.R. Spickler. *Nipah Virus Infection*. 2016. URL: `http://www.cfsph.iastate.edu/DiseaseInfo/factsheets.php` (cit. on pp. 6, 7, 11, 16–18).

[31] Jonathan H. Epstein et al. «Nipah virus dynamics in bats and implications for spillover to humans». In: *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 117.46 (2020). Epub 2020 Nov 2, pp. 29190–29201. DOI: `10.1073/pnas.2000429117`. URL: `https://doi.org/10.1073/pnas.2000429117` (cit. on p. 6).

[32] H. Constable. *The other virus that worries Asia*. 2021. URL: `https://www.bbc.com/future/article/20210106-nipah-virus-how-bats-could-cause-the-next-pandemic` (cit. on pp. 7, 18, 23, 24).

[33] C.D. McKee, A. Islam, S.P. Luby, H. Salje, P.J. Hudson, R.K. Plowright, and E.S. Gurley. «The Ecology of Nipah Virus in Bangladesh: A Nexus of Land-Use Change and Opportunistic Feeding Behavior in Bats». In: *Viruses* 13.2 (2021). DOI: `10.3390/v13020169` (cit. on pp. 7, 10, 12–14, 16–19, 24, 26, 35–38, 41, 50, 53, 54, 73, 77–79, 81, 85, 87, 89, 103, 104, 108, 109, 112, 113).

[34] L. Bruno, M.A. Nappo, L. Ferrari, R. Di Lecce, C. Guarnieri, A.M. Cantoni, and A. Corradi. «Nipah Virus Disease: Epidemiological, Clinical, Diagnostic and Legislative Aspects of This Unpredictable Emerging Zoonosis». In: *Animals* 13.1 (2023). DOI: `10.3390/ani13010159` (cit. on pp. 7, 9, 16, 17).

[35] Robyn Fogarty, Kerrie Halpin, Alex D. Hyatt, Peter Daszak, and Bruce A. Mungall. «Henipavirus susceptibility to environmental variables». In: *Virus Research* 132.1-2 (2008). Epub 2007 Dec 31, pp. 140–144. DOI: `10.1016/j.virusres.2007.11.010`. URL: `https://doi.org/10.1016/j.virusres.2007.11.010` (cit. on pp. 7, 17).

[36] M.A. Deka and N. Morshed. «Mapping Disease Transmission Risk of Nipah Virus in South and Southeast Asia». In: *Trop Med Infect Dis* 3.2 (2018). DOI: `10.3390/tropicalmed3020057` (cit. on pp. 7, 10, 11, 13, 18, 19, 35, 103).

[37] Encyclopaedia Britannica. *Flying Fox*. `https://www.britannica.com/animal/flying-fox`. 2024 (cit. on p. 7).

[38] Reuters. *Global Pandemic: Bats Spillover*. 2024. URL: `https://www.reuters.com/investigates/special-report/global-pandemic-bats-spillover/` (cit. on pp. 8, 10, 18, 26, 27, 35).

[39] P. Daszak, C. Zambrana-Torrelio, T.L. Bogich, M. Fernandez, J.H. Epstein, K.A. Murray, and H. Hamilton. «Interdisciplinary approaches to understanding disease emergence: the past, present, and future drivers of Nipah virus emergence». In: *Proc Natl Acad Sci U S A* 110.Suppl 1 (2013), pp. 3681–3688. DOI: `10.1073/pnas.1201243109` (cit. on p. 9).

[40] Utpal Singha Roy Somenath Dey and Sanjib Chattopadhyay. «Effect of heat wave on the Indian Flying Fox Pteropus giganteus (Mammalia: Chiroptera: Pteropodidae) population from Purulia District of West Bengal, India». In: *Journal of Threatened Taxa* 7.3 (2015), pp. 7029–7033. DOI: `10.11609/jott.1905.7029-7033` (cit. on pp. 9, 25, 50).

[41] Micah B. Hahn, Jonathan H. Epstein, Emily S. Gurley, Md Saiful Islam, Stephen P. Luby, Peter Daszak, and Jonathan A. Patz. «Roosting behaviour and habitat selection of *Pteropus giganteus* reveals potential links to Nipah virus epidemiology». In: *Journal of Applied Ecology* 51.2 (2014), pp. 376–387. DOI: `10.1111/1365-2664.12212`. URL: `https://doi.org/10.1111/1365-2664.12212` (cit. on pp. 10, 100).

[42] K. Choden et al. «Pteropus lylei primarily forages in residential areas in Kandal, Cambodia». In: *Ecology and Evolution* 9.16 (2019), pp. 1108–1119. DOI: `10.1002/ece3.5046` (cit. on pp. 10, 63).

[43] Keren Egert-Berg, Meital Handel, Alon Goldshtein, Omer Eitan, Ivailo Borissov, and Yossi Yovel. «Fruit bats adjust their foraging strategies to urban environments to diversify their diet». In: *BMC Biology* 19.1 (2021), p. 123. DOI: `10.1186/s12915-021-01060-x` (cit. on p. 10).

[44] Jeremy Tait, Humberto L. Perotto-Baldivieso, Annabel McKeown, and David A. Westcott. «Are flying-foxes coming to town? Urbanisation of the spectacled flying-fox (*Pteropus conspicillatus*) in Australia». In: *PLoS One* 9.10 (2014), e109810. DOI: `10.1371/journal.pone.0109810`. URL: `https://doi.org/10.1371/journal.pone.0109810` (cit. on p. 10).

[45] S. Win and M. Khin. «The diet of the Indian Flying Fox Pteropus giganteus (Brünnich. 1782) (Chiroptera: Pteropodidae) in Myanmar - conflicts with local people?» In: *Journal of Threatened Taxa* 7.4 (2015), pp. 7568–7572. DOI: `10.11609/jott.2182.7568-7572` (cit. on pp. 10, 63).

[46] M.S. Islam et al. «Nipah Virus Transmission from Bats to Humans Associated with Drinking Traditional Liquor Made from Date Palm Sap, Bangladesh, 2011-2014». In: *Emerg Infect Dis.* 22.4 (2016). PMID: 26981928; PMCID: PMC4806957, pp. 664–670. DOI: `10.3201/eid2204.151747` (cit. on pp. 10, 17–19, 112, 114, 116).

[47] M. Döring. *Pteropodidae, 1821 in English Wikipedia - Species Pages.* Checklist dataset `https://doi.org/10.15468/c3kkgh` accessed via GBIF.org. Global Biodiversity Information Facility (GBIF). 2022 (cit. on p. 10).

[48] Elizabeth D. Pierson and William E. Rainey. «The Biology of Flying Foxes of the Genus Pteropus: A Review». In: *Pacific Island Flying Foxes: Proceedings of an International Conservation Conference.* Ed. by Don E. Wilson and Gary Lynn Graham. Vol. 90. Biological Report. Washington, D.C.: U.S. Department of the Interior, Fish and Wildlife Service, 1992, pp. 1–17 (cit. on pp. 10, 11, 26, 27, 51, 58).

[49] Beng Siang P. Ang, Tze Chean C. Lim, and Linfa Wang. «Nipah Virus Infection». In: *Journal of Clinical Microbiology* 56.6 (2018). Epub 2018 May 25, e01875–17. DOI: `10.1128/JCM.01875-17` (cit. on p. 10).

[50] L.M. Looi and K.B. Chua. «Lessons from the Nipah virus outbreak in Malaysia». In: *Malays J Pathol* 29.2 (2007), pp. 63–67 (cit. on p. 11).

[51] Juliet R. C. Pulliam et al. «Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis». In: *Journal of the Royal Society Interface* 9.66 (2012). Epub 2011 Jun 1, pp. 89–101. DOI: `10.1098/rsif.2011.0223` (cit. on p. 11).

[52] Bronwyn J. Roberts, Carla P. Catterall, Peggy Eby, and John Kanowski. «Long-distance and frequent movements of the flying-fox *Pteropus poliocephalus*: implications for management». In: *PLoS One* 7.8 (2012). Epub 2012 Aug 3, e42532. DOI: `10.1371/journal.pone.0042532`. URL: `https://doi.org/10.1371/journal.pone.0042532` (cit. on p. 11).

[53] J.A. Welbergen, J. Meade, H. Field, et al. «Extreme mobility of the world's largest flying mammals creates key challenges for management and conservation». In: *BMC Biol* 18 (2020). DOI: `10.1186/s12915-020-00829-w` (cit. on p. 11).

[54] Anne Seltmann, Gábor Á. Czirják, Alexandre Courtiol, Henry Bernard, Matthew J. Struebig, and Christian C. Voigt. «Habitat disturbance results in chronic stress and impaired health status in forest-dwelling paleotropical bats». In: *Conservation Physiology* 5.1 (2017), cox020. DOI: `10.1093/conphys/cox020` (cit. on pp. 11, 25).

[55] M.B. Hahn, E.S. Gurley, J.H. Epstein, M.S. Islam, J.A. Patz, P. Daszak, and S.P. Luby. «The role of landscape composition and configuration on *Pteropus giganteus* roosting ecology and Nipah virus spillover risk in Bangladesh». In: *American Journal of Tropical Medicine and Hygiene* 90.2 (2014), pp. 247–255. DOI: `10.4269/ajtmh.13-0256` (cit. on pp. 13, 35, 63, 73).

[56] A. Islam, C. McKee, P. K. Ghosh, J. Abedin, J. H. Epstein, P. Daszak, S. P. Luby, S. U. Khan, and E. S. Gurley. «Seasonality of Date Palm Sap Feeding Behavior by Bats in Bangladesh». In: *Ecohealth* 18.3 (Sept. 2021). Epub 2021 Oct 5, pp. 359–371. DOI: `10.1007/s10393-021-01561-9` (cit. on pp. 16, 19, 26, 27, 50, 52, 107).

[57] Javier Faus-Cotino, Gonzalo Reina, and Jorge Pueyo. «Nipah Virus: A Multidimensional Update». In: *Viruses* 16.2 (2024). Epub 2024 Jan 25, p. 179. DOI: `10.3390/v16020179` (cit. on pp. 17–19).

[58] Joel M. Montgomery, Mohamed J. Hossain, E. Gurley, D.S. Carroll, A. Croisier, et al. «Risk Factors for Nipah Virus Encephalitis in Bangladesh». In: *Emerg Infect Dis.* 14.10 (2008), pp. 1526–1532. DOI: `10.3201/eid1410.060507` (cit. on pp. 18, 24, 108, 115).

[59] M.C. Cortes et al. «Characterization of the Spatial and Temporal Distribution of Nipah Virus Spillover Events in Bangladesh, 2007-2013». In: *J. Infect. Dis.* 217 (2018), pp. 1390–1394. DOI: `10.1093/infdis/jiy015` (cit. on pp. 19, 36–38, 41, 108–112, 115, 116).

[60] Mannava V. K. Sivakumar and Robert Stefanski. *Climate Change in South Asia.* Springer, Dordrecht, 2010, pp. 13–30. DOI: `10.1007/978-90-481-9516-9_2` (cit. on pp. 19–22).

[61] B. B. Singh, M. Singh, and D. Singh. «An Overview of Climate Change Over South Asia: Observations, Projections, and Recent Advances». In: *Practices in Regional Science and Sustainable Regional Development.* Ed. by R. B. Singh, S. Chatterjee, M. Mishra, and A. J. de Lucena. Singapore: Springer, 2021, pp. 275–293. DOI: `10.1007/978-981-16-2221-2_12` (cit. on p. 19).

[62] V. Nath, G. Kumar, S. D. Pandey, and S. Pandey. «Impact of Climate Change on Tropical Fruit Production Systems and its Mitigation Strategies». In: *Climate Change and Agriculture in India: Impact and Adaptation.* Ed. by Sheraz Mahdi. Cham: Springer, 2019, pp. 273–298. DOI: `10.1007/978-3-319-90086-5_11`. URL: `https://doi.org/10.1007/978-3-319-90086-5_11` (cit. on pp. 19, 50).

[63] John A. Weiss. *The Economics of Climate Change in Southeast Asia: A Regional Review.* Climate change, Promoting climate resilient development. Asian Development Bank, 2009. ISBN: 978-971-561-787-1 (cit. on pp. 20, 22).

[64] Renata L. Muylaert, Tigga Kingston, Jinhong Luo, Maurício Humberto Vancine, Nikolas Galli, Colin J. Carlson, Reju Sam John, Maria Cristina Rulli, and David T. S. Hayman. «Present and future distribution of bat hosts of sarbecoviruses: implications for conservation and public health». In: *Proceedings of the Royal Society B* 289.20220397 (2022). DOI: `10.1098/rspb.2022.0397` (cit. on p. 20).

[65] Alice C Hughes, Chutamas Satasook, Paul J. J. Bates, Sara Bumrungsri, and Gareth J. F. Jones. «The projected effects of climatic and vegetation changes on the distribution and diversity of Southeast Asian bats». In: *Global Change Biology* 18 (2012). URL: `https://api.semanticscholar.org/CorpusID:85621727` (cit. on p. 20).

[66] Aditya Srinivasulu, Matt R. K. Zeale, Bhargavi Srinivasulu, Chelmala Srinivasulu, Gareth Jones, and Mónica González-Suárez. «Future climatically suitable areas for bats in South Asia». In: *Ecology and Evolution* 14.5 (2024). DOI: `10.1002/ece3.11420` (cit. on p. 20).

[67] Neha M. Vora, Lee Hannah, Susan Lieberman, Mariana M. Vale, Raina K. Plowright, and Aaron S. Bernstein. «Want to prevent pandemics? Stop spillovers». In: *Nature* 605.7910 (2022), pp. 419–422. DOI: `10.1038/d41586-022-01312-y` (cit. on pp. 20, 24).

[68] Climate Technology Centre and Network. *IPCC's Fifth Assessment Report: What's in it for South Asia?* `https://www.ctc-n.org/resources/ipccs-fifth-assessment-report-whats-it-south-asia`. 2014 (cit. on pp. 20, 21).

[69] Dennis L. Hartmann. *Global Physical Climatology*. Vol. 56. International Geophysics. Elsevier Science, 1994, p. 411. ISBN: 0123285305 (cit. on p. 21).

[70] A. Cherchi and A. Turner. «Annex V: Monsoons». In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte, P. Zhai, A. Pirani, et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 2193–2204. DOI: `10.1017/9781009157896.019` (cit. on p. 21).

[71] Hui Luo, Zhaobo Wang, Chao He, et al. «Future changes in South Asian summer monsoon circulation under global warming: role of the Tibetan Plateau latent heating». In: *npj Climate and Atmospheric Science* 7 (2024), p. 103. DOI: `10.1038/s41612-024-00653-x` (cit. on p. 21).

[72] V. Masson-Delmotte, P. Zhai, A. Pirani, et al., eds. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. In press. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021. DOI: `10.1017/9781009157896` (cit. on pp. 21, 22).

[73] Walter Leal Filho, Liliia Ternova, Samir A. Parasnis, Marina Kovaleva, and Gustavo J. Nagy. «Climate Change and Zoonoses: A Review of Concepts, Definitions, and Bibliometrics». In: *International Journal of Environmental Research and Public Health* 19.2 (2022), p. 893. DOI: `10.3390/ijerph19020893`. URL: `https://doi.org/10.3390/ijerph19020893` (cit. on p. 22).

[74] K. Marie Mcintyre, Christian Setzkorn, Philip Hepworth, Serge Morand, Andrew Morse, and Matthew Baylis. «Systematic Assessment of the Climate Sensitivity of Important Human and Domestic Animals Pathogens in Europe». In: *Scientific Reports* 7 (2017), Article 7137. DOI: `10.1038/s41598-017-06948-9` (cit. on p. 22).

[75] G. Cissé et al. «Health, Wellbeing, and the Changing Structure of Communities». In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by H.-O. Pörtner, D. C. Roberts, M. Tignor, et al. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022, pp. 1041–1170. DOI: `10.1017/9781009325844.009` (cit. on pp. 22, 23).

[76] Eva Alessi, Marco Galaverni, Mariagrazia Midulla, Isabella Pratesi, and Giulia Prato. *Climate Change and Zoonotic Infectious Diseases*. Rome, Italy, 2020. URL: `https://www.wwf.it` (cit. on pp. 22, 23).

[77] T. Allen, K. A. Murray, C. Zambrana-Torrelio, et al. «Global hotspots and correlates of emerging zoonotic diseases». In: *Nature Communications* 8 (2017), p. 1124. DOI: `10.1038/s41467-017-00923-8` (cit. on p. 23).

[78] Kenneth L. Gage et al. «Climate and Vectorborne Diseases». In: *American Journal of Preventive Medicine* 35.5 (2008), pp. 436–450. DOI: `10.1016/j.amepre.2008.08.030` (cit. on p. 23).

[79] Vishesh L. Diengdoh, Stefania Ondei, Mark Hunt, and Barry W. Brook. «Predicted impacts of climate change and extreme temperature events on the future distribution of fruit bat species in Australia». In: *Global Ecology and Conservation* 37 (2022), e02181. ISSN: 2351-9894. DOI: `10.1016/j.gecco.2022.e02181`. URL: `https://www.sciencedirect.com/science/article/pii/S2351989422001834` (cit. on p. 25).

[80] Rob Stones and Jacob E. Wiebers. «A Review of Temperature Regulation in Bats (Chiroptera)». In: *American Midland Naturalist* 74 (1965), p. 155. URL: `https://api.semanticscholar.org/CorpusID:87339205` (cit. on pp. 25, 50).

[81] Matthew Mo, Jessica Meade, Janina Price, Jacquie Maisey, and Justin A. Welbergen. «Synchronous abortion events in the grey-headed flying-fox (Pteropus poliocephalus)». In: *Pacific Conservation Biology* 29 (2023), pp. 110–118. DOI: `10.1071/PC21060` (cit. on p. 25).

[82] Peggy Eby, Alison J. Peel, Amanda Hoegh, et al. «Pathogen spillover driven by rapid changes in bat ecology». In: *Nature* 613 (2023), pp. 340–344. DOI: `10.1038/s41586-022-05506-2`. URL: `https://doi.org/10.1038/s41586-022-05506-2` (cit. on pp. 26, 27, 51, 101, 103, 104).

[83] Daniel J. Becker, Peggy Eby, William Madden, Alison J. Peel, and Raina K. Plowright. «Ecological conditions predict the intensity of Hendra virus excretion over space and time». In: *Ecology Letters* 26.1 (2023), pp. 23–36. DOI: `10.1111/ele.14007`. URL: `https://doi.org/10.1111/ele.14007` (cit. on pp. 27, 51, 101, 103, 104).

[84] R. K. Plowright, H. E. Field, C. Smith, A. Divljan, C. Palmer, G. Tabor, P. Daszak, and J. E. Foley. «Reproduction and nutritional stress are risk factors for Hendra virus infection in little red flying foxes (Pteropus scapulatus)». In: *Proceedings of the Royal Society B: Biological Sciences* 275.1636 (2008), pp. 861–869. DOI: `10.1098/rspb.2007.1260`. URL: `https://doi.org/10.1098/rspb.2007.1260` (cit. on p. 27).

[85] Winifred F. Frick, Tigga Kingston, and Jessa Flanders. «A review of the major threats and challenges to global bat conservation». In: *Annals of the New York Academy of Sciences* 1469.1 (2020), pp. 5–25. DOI: `10.1111/nyas.14045` (cit. on p. 28).

[86] Paolo Agnelli, Alessandro Martinoli, Elena Patriarca, Danilo Russo, Dario Scaravelli, and Piero Genovesi. *Guidelines for bat monitoring: methods for the study and conservation of bats in Italy.* Vol. 19 bis. Quaderni di Conservazione della Natura. Rome and Ozzano dell'Emilia (Bologna), Italy: Ministero dell'Ambiente - Istituto Nazionale per la Fauna Selvatica, 2006 (cit. on p. 28).

[87] Ahsan Uddin Ahmed. *Bangladesh Climate Change Impacts and Vulnerability: A Synthesis.* Bangladesh: Climate Change Cell, Department of Environment, 2006, p. 70. ISBN: 9843234677 (cit. on pp. 30–33, 65).

[88] Simon et al. *Climate: observations, projections and impacts.* Tech. rep. Institution or Organization, 2011 (cit. on p. 30).

[89] Md. Reaz Akter Mullick, Ridwan Mohammed Nur, Md. Jahangir Alam, and K.M. Ashraful Islam. «Observed trends in temperature and rainfall in Bangladesh using pre-whitening approach». In: *Global and Planetary Change* 172 (2019), pp. 104–113. ISSN: 0921-8181. DOI: `10.1016/j.gloplacha.2018.10.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0921818117302552` (cit. on pp. 30, 32, 33, 48).

[90] Kaosar Afsana et al. «Health care for poor people in the urban slums of Bangladesh». In: *The Lancet* 382.9910 (2013), pp. 2049–2051 (cit. on pp. 33, 34).

[91] Siddharth T. Hegde et al. «Using healthcare-seeking behaviour to estimate the number of Nipah outbreaks missed by hospital-based surveillance in Bangladesh». In: *International Journal of Epidemiology* 48.4 (2019). Epub 2019 Aug 1, pp. 1219–1227. DOI: `10.1093/ije/dyz057` (cit. on p. 36).

[92] B. Nikolay et al. «Transmission of Nipah Virus - 14 Years of Investigations in Bangladesh». In: *N Engl J Med.* 380.19 (2019). PMID: 31067370; PMCID: PMC6547369, pp. 1804–1814. DOI: `10.1056/NEJMoa1805376` (cit. on p. 36).

[93] S.P. Luby, M.J. Hossain, E.S. Gurley, B.N. Ahmed, S. Banu, S.U. Khan, et al. «Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001-2007». In: *Emerg. Infect. Dis.* 15.8 (2009), pp. 1229–1235. DOI: `10.3201/eid1508.081237` (cit. on pp. 36, 108–110, 115).

[94] M.Z. Hassan et al. «Nipah Virus Contamination of Hospital Surfaces during Outbreaks, Bangladesh, 2013-2014». In: *Emerg Infect Dis.* 24.1 (2018). PMID: 29260663; PMCID: PMC5749460, pp. 15–21. DOI: `10.3201/eid2401.161758` (cit. on pp. 37, 112, 113, 116).

[95] Graham Martin, Carlos Yanez-Arenas, Chen Chen, Raina K Plowright, Richard J Webb, and Lee F Skerratt. «Climate Change Could Increase the Geographic Extent of Hendra Virus Spillover Risk». In: *EcoHealth* 15.3 (2018). Epub 2018 Mar 19, pp. 509–525. DOI: `10.1007/s10393-018-1322-9` (cit. on p. 41).

[96] European Centre for Medium-Range Weather Forecasts (ECMWF). *ERA5 Reanalysis (ECMWF Reanalysis v5)*. `https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5`. Accessed: 2024-07-30. 2021 (cit. on p. 42).

[97] Copernicus Climate Change Service (C3S). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. 2017. URL: `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview` (cit. on p. 42).

[98] M. Tarek, F. P. Brissette, and R. Arsenault. «Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America». In: *Hydrology and Earth System Sciences* 24.5 (2020), pp. 2527–2550. DOI: `10.5194/hess-24-2527-2020`. URL: `https://hess.copernicus.org/articles/24/2527/2020/` (cit. on pp. 42, 102).

[99] C. Vega, G. Pertierra, and M. Olalla-Tárraga. «MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling». In: *Sci Data* 4 (2017), p. 170078. DOI: `10.1038/sdata.2017.78`. URL: `https://doi.org/10.1038/sdata.2017.78` (cit. on pp. 42, 43).

[100] WorldClim. *WorldClim Version 1.4: Methods*. `https://www.worldclim.org/data/v1.4/methods.html`. 2020 (cit. on p. 43).

[101] Vincent Deblauwe et al. «Remotely sensed climate data for tropical species distribution models». In: *Global Ecology and Biogeography* 25.4 (2016), pp. 443–454. DOI: `10.1111/geb.12426`. URL: `https://doi.org/10.1111/geb.12426` (cit. on pp. 43, 102).

[102] International Research Institute for Climate and Society. *Climatologies and Standardized Anomalies.* `https://iridl.ldeo.columbia.edu/dochelp/StatTutorial/Climatologies/index.html#Intro` (cit. on p. 44).

[103] University of Maine Climate Reanalyzer Climate Change Institute. *2m temperature standardized anomaly.* `https://climatereanalyzer.org/wx/todays-weather/?var_id=t2anom&ortho=1&wt=1` (cit. on p. 44).

[104] Millennia Young. *Permutation Testing and Bootstrapping: Non-parametric Approaches to Statistical Testing and Estimation.* NASA Technical Report 20150001882. Wyle Science, Technology and Engineering Group, 2015. URL: `https://ntrs.nasa.gov/api/citations/20150001882/downloads/20150001882.pdf` (cit. on p. 47).

[105] Jonathan Pillow. *Bootstrap Methods.* Lecture slides from Math Tools for Neuroscience, Princeton University. Available online. 2016. URL: `https://pillowlab.princeton.edu/teaching/mathtools16/slides/lec21_Bootstrap.pdf` (cit. on p. 47).

[106] Bhuwon Sthapit, V. Ramanatha Rao, and Sajal Sthapit, eds. *Tropical Fruit Tree Species and Climate Change.* Office for South Asia, National Agricultural Science Centre, DPS Marg, Pusa Campus, New Delhi 110 012, India: Bioversity International, 2012. ISBN: 978-92-9043-909-7 (cit. on p. 50).

[107] Jennifer J. Openshaw, Sapana Hegde, Hossain M. S. Sazzad, S. M. Khan, M. Jahangir Hossain, Jonathan H. Epstein, Peter Daszak, Emily S. Gurley, and Stephen P. Luby. «Bat Hunting and Bat-Human Interactions in Bangladeshi Villages: Implications for Zoonotic Disease Transmission and Bat Conservation». In: *Transboundary and Emerging Diseases* 64.4 (Aug. 2017). Epub 2016 Apr 29, pp. 1287–1293. DOI: `10.1111/tbed.12505`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/tbed.12505`. URL: `https://doi.org/10.1111/tbed.12505` (cit. on p. 50).

[108] Giuseppe Boccignone. *k-means clustering.* `https://boccignone.di.unimi.it/IN_2019_files/w4_kmeans.pdf`. Accessed: 2024-09-05. 2019 (cit. on pp. 70, 127, 128).

[109] Imad Dabbura. *K-Means Clustering Algorithm: Applications, Evaluation Methods, and Drawbacks.* `https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a`. 2018 (cit. on pp. 70, 127).

[110] Ankita Banerji. *K-Mean: Getting the Optimal Number of Clusters.* `https://www.analyticsvidhya.com/blog/2021/05/k-mean-getting-the-optimal-number-of-clusters/`. 2024 (cit. on p. 70).

[111] scikit-learn. *K-Means Clustering Silhouette Analysis.* `https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html#sphx-glr-auto-examples-cluster-plot-kmeans-silhouette-analysis-py` (cit. on p. 70).

[112] Michael Greenacre, Patrick J.F. Groenen, Trevor Hastie, et al. «Principal component analysis». In: *Nature Reviews Methods Primers* 2 (2022), p. 100. DOI: `10.1038/s43586-022-00184-w`. URL: `https://www.nature.com/articles/s43586-022-00184-w` (cit. on p. 71).

[113] J. A. Nelder and R. W. M. Wedderburn. «Generalized Linear Models». In: *Journal of the Royal Statistical Society, Series A* 135 (1972), pp. 370–384 (cit. on p. 80).

[114] Michael L. Zwilling. «Negative Binomial Regression». In: *The Mathematica journal* 15 (2013). URL: `https://api.semanticscholar.org/CorpusID:124513556` (cit. on pp. 81, 83, 85, 129, 130, 138, 140).

[115] Yosiyuki Sakamoto, Makio Ishiguro, and G. Kitagawa. *Akaike Information Criterion Statistics*. illustrated. Vol. 1. Mathematics and its Applications, Japanese Series. Original from the University of Michigan, digitized February 3, 2010. Springer Netherlands, 1986, p. 290. ISBN: 9027722536 (cit. on p. 82).

[116] N. Sugiura. «Further analysis of the data by Akaike's information criterion and the finite corrections». In: *Communications in Statistics - Theory and Methods* 7 (1978), pp. 13–26. DOI: `10.1080/03610927808827599` (cit. on p. 82).

[117] Federico Castellan. «Metodi di verosimiglianza per la stima della sovradispersione nel modello binomiale negativo». `https://hdl.handle.net/20.500.12608/27872`. Università degli Studi di Padova, 2018 (cit. on pp. 83, 129, 130, 138).

[118] Omar Sosa. *Poisson Regression in Python*. `https://omarfsosa.github.io/poisson_regression_in_python`. Accessed on the web via GitHub repository. 2020 (cit. on p. 84).

[119] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. 5th ed. Vol. 821. Wiley Series in Probability and Statistics. John Wiley & Sons, 2012, p. 672 (cit. on pp. 84, 85).

[120] J. Hinde and C. Demétrio. «Half-normal plots and overdispersion». In: *GLIM Newsletter* 27 (1997), pp. 19–26 (cit. on p. 84).

[121] Mohammad Kamruzzaman et al. «Predicted changes in future precipitation and air temperature across Bangladesh using CMIP6 GCMs». In: *Heliyon* 9.5 (2023), e16274. ISSN: 2405-8440. DOI: `10.1016/j.heliyon.2023.e16274`. URL: `https://www.sciencedirect.com/science/article/pii/S2405844023034813` (cit. on p. 100).

[122] A. Bhattacharjee, S. M. Q. Hassan, P. Hazra, T. Kormoker, S. Islam, E. Alam, and A. R. Md. Towfiqul Islam. «Future changes of summer monsoon rainfall and temperature over Bangladesh using 27 CMIP6 models». In: *Geocarto International* 38.1 (2023). DOI: `10.1080/10106049.2023.2285342`. URL: `https://doi.org/10.1080/10106049.2023.2285342` (cit. on pp. 100, 101).

[123] Yoshihiro Fukuda, Masahiro Watanabe, and Fei-Fei Jin. «Mode of Precipitation Variability Generated by Coupling of ENSO With Seasonal Cycle in the Tropical Pacific». In: *Geophysical Research Letters* 48.16 (2021). DOI: `https://doi.org/10.1029/2021GL095204`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL095204` (cit. on p. 106).

[124] ScienceDirect. *Evapotranspiration - Earth and Planetary Sciences*. `https://www.sciencedirect.com/topics/earth-and-planetary-sciences/evapotranspiration` (cit. on p. 107).

[125] Accademia dei Georgofili Silverio Pachioli. *Cavitazione, embolia, tillosi nelle piante agrarie e forestali: cause, conseguenze e possibili rimedi.* `https://www.georgofili.info/contenuti/cavitazione-embolia-tillosi-nelle-piante-agrarie-e-forestali-cause-conseguenze-e-possibili-rimedi/15355`. 2023 (cit. on p. 107).

[126] Crisafulli. *Cavitazione - Fisiologia vegetale.* `https://biologiawiki.it/wiki/cavitazione/`. 2024 (cit. on p. 107).

[127] Crisafulli. *Refilling dello Xilema - Fisiologia vegetale.* `https://biologiawiki.it/wiki/refilling-dello-xilema/`. 2024 (cit. on p. 107).

[128] V.P. Hsu et al. «Nipah virus encephalitis reemergence Bangladesh». In: *Emerg. Infect. Dis.* 10 (2004), pp. 2082–2087. DOI: `10.3201/eid1012.040701` (cit. on pp. 108, 109, 115).

[129] World Health Organization. *Nipah virus outbreak(s) in Bangladesh, January-April 2004.* Available from: `https://www.who.int/publications/i/item/WER7917`. 2004 (cit. on pp. 108, 115).

[130] World Health Organization. *Disease Outbreak News 2004 - Bangladesh.* Available from: `https://www.who.int/emergencies/disease-outbreak-news/item/2004_02_12-en`. 2004 (cit. on pp. 108, 115).

[131] M. Jahangir Hossain, Emily S. Gurley, Joel M. Montgomery, et al. «Clinical Presentation of Nipah Virus Infection in Bangladesh». In: *Clin Infect Dis* 46.7 (2008), pp. 977–984. DOI: `10.1086/529147` (cit. on pp. 108, 115).

[132] Emily S. Gurley, Joel M. Montgomery, M. Jahangir Hossain, et al. «Person-to-Person Transmission of Nipah Virus in a Bangladeshi Community». In: *Emerg Infect Dis.* 13.7 (2007), pp. 1031–1037. DOI: `10.3201/eid1307.061128` (cit. on pp. 108, 115).

[133] M.K. Lo et al. «Characterization of Nipah virus from outbreaks in Bangladesh, 2008-2010». In: *Emerg Infect Dis.* 18.2 (2012). PMID: 22304936; PMCID: PMC3310473, pp. 248–255. DOI: `10.3201/eid1802.111492` (cit. on pp. 110, 115).

[134] IEDCR One Health Event Based Surveillance System Dashboard. *Year Wise NIPAH Case Distribution.* `https://ohebsdd.org/dash`. 2024 (cit. on p. 111).

[135] M.Z. Rahman et al. «Genetic diversity of Nipah virus in Bangladesh». In: *Int J Infect Dis.* 102 (2021). Epub 2020 Oct 28. PMID: 33129964, pp. 144–151. DOI: `10.1016/j.ijid.2020.10.041` (cit. on pp. 111–114, 116).

[136] B. Nikolay et al. «A Framework to Monitor Changes in Transmission and Epidemiology of Emerging Pathogens: Lessons From Nipah Virus». In: *J Infect Dis.* 221.Suppl 4 (2020). PMID: 32392322; PMCID: PMC7213557, S363–S369. DOI: `10.1093/infdis/jiaa074` (cit. on pp. 112, 113, 116).

[137] A. Islam et al. «Nipah Virus Exposure in Domestic and Peridomestic Animals Living in Human Outbreak Sites, Bangladesh, 2013-2015». In: *Emerg Infect Dis.* 29.2 (2023). PMID: 36692447; PMCID: PMC9881791, pp. 393–396. DOI: `10.3201/eid2902.221379` (cit. on pp. 113, 116).

[138] N. Homaira et al. «Nipah virus outbreak with person-to-person transmission in a district of Bangladesh, 2007». In: *Epidemiol Infect.* 138.11 (2010). Epub 2010 Apr 12. PMID: 20380769; PMCID: PMC10071638, pp. 1630–1636. DOI: `10.1017/S0950268810000695` (cit. on p. 115).

[139] N. Homaira et al. «Cluster of Nipah virus infection, Kushtia District, Bangladesh, 2007». In: *PLoS One* 5.10 (2010). PMID: 21042407; PMCID: PMC2958840, e13570. DOI: `10.1371/journal.pone.0013570` (cit. on p. 115).

[140] Disease Control Institute of Epidemiology and Research. *One Health Event Based Surveillance System Dashboard*. Available from: `https://ohebsdd.org/dash`. 2023 (cit. on p. 116).

[141] G. Okafor, O. Jimoh, and K. Larbi. «Detecting Changes in Hydro-Climatic Variables during the Last Four Decades (1975-2014) on Downstream Kaduna River Catchment, Nigeria». In: *Atmospheric and Climate Sciences* 7 (2017), pp. 161–175. DOI: `10.4236/acs.2017.72012` (cit. on pp. 120, 121).

[142] Md. Reaz Akter Mullick, Ridwan Mohammed Nur, Md. Jahangir Alam, and K.M. Ashraful Islam. «Observed trends in temperature and rainfall in Bangladesh using pre-whitening approach». In: *Global and Planetary Change* 172 (2019), pp. 104–113. ISSN: 0921-8181. DOI: `10.1016/j.gloplacha.2018.10.001` (cit. on p. 120).

[143] Joseph M. Hilbe. *Negative Binomial Regression*. 2nd. New York: Cambridge University Press, 2011 (cit. on p. 129).

[144] Unknown Author. *Chapter 15*. `https://www.sagepub.com/sites/default/files/upm-binaries/21121_Chapter_15.pdf`. Published by SAGE (cit. on pp. 139, 140).

[145] The Pennsylvania State University. *Regression method - Using Leverages to Help Identify Extreme x Values*. `https://online.stat.psu.edu/stat501/lesson/11/11.2` (cit. on p. 139).

[146] MathWorks. *Cook's Distance*. `https://it.mathworks.com/help/stats/cooks-distance.html` (cit. on p. 140).