1
2

Inferring livestock movement networks from archived data to support infectious disease control in developing countries

3

4 A. Muwonge[1,4], P.R. Bessell[1], T. Porphyre[1,5], P. Motta[1,6], G. Rydevik [1] , G.
5 Devailly[1,3], N.F. Egbe[2], R.F. Kelly[1], I.G. Handel[1,4], S. Mazeri[1,4], B.M.deC.
6 Bronsvoort[1,4]

7
8
9

10 1. The Roslin Institute and the Royal (Dick) School of Veterinary Studies,
11   University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK
12
13 2. School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS
14   United Kingdom.

15
16 3. GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan,
17   France

18
19 4. Epidemiology Economics and Risk Assessment group at The Roslin Institute
20   and the Royal (Dick) School of Veterinary Studies, University of Edinburgh,
21   Easter Bush, Midlothian, EH25 9RG, UK
22
23 5. Université Lyon 1, CNRS, VetAgro Sup, Laboratoire de Biométrie Et Biologie
24   Évolutive, Université de Lyon, Villeurbanne Cedex, France

25
26 6. The European Commission for the Control of Foot-and-Mouth Disease
27   (EuFMD), Food and Agriculture Organization of the United Nations, Rome,
28   Italy

29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 *Key words: Gravity modelling, phylogeography, M.bovis, Livestock, Cameroon*
46 **Abstract**

47

48 The use of network analysis to support livestock disease control in low middle-
49 income countries (LMICs) has historically been hampered by the cost of generating
50 empirical data in the absence of animal movement recording schemes. To fill this
51 gap, methods which exploit freely available demographic and archived molecular
52 data can be used to generate livestock networks based on gravity and
53 phylogeographic modelling techniques, respectively. However, questions remain on
54 the performance of these methods in capturing the topology of empirical networks.
55 Here, we compare output from these network methodologies to a network
56 constructed from either empirical data or randomly generated data. To facilitate this
57 comparison, the spread of infectious diseases was simulated, it is this evaluation that
58 demonstrates their potential utility to inform robust livestock disease control
59 strategies.
60 The molecular network was the closest approximation to the empirical network, both
61 in relation to topological and epidemic characteristics, whereas size of epidemics in
62 the gravity network tended to be larger, better agreement across all three networks
63 was observed when; a) total nodes infected, b) percentage infection take off were
64 compared. These methods consistently identified the same important animal
65 movement and trade hotspots as the empirical networks. We therefore consider this
66 proof-of-concept that demographic data such as censuses and archived molecular
67 data could be repurposed to inform livestock disease management in LMICs.

68

69 **Author summary**

70

71 Live animal movements in Africa represent a significant risk of transmission and
72 spread of infectious diseases in livestock populations, and therefore, have direct
73 implications on the food security of the continent. Here we explore the potential utility
74 of available data to support control strategies, by comparing movement networks
75 inferred from such data i.e. census and pathogen molecular data using gravity
76 modelling and phylogeography respectively. Their utility is evaluated by comparing
77 their topology and disease spread characteristics to empirical live animal movement.
78 Based on our results, we posit that archived data can be repurposed to support
79 infectious disease control on the African continent.

80

81

Introduction

83

84 Network-based approaches enable users to describe livestock movements as a

85 spatial network where nodes may represent villages, markets, herds, or individual

86 animals, whereas edges define the movement of at least one animal from one node

87 to another [1,2]. The utility of this approach has been exploited for epidemiological

88 purposes, in particular, to understand the role of live animal movements in disease

89 spread and control strategies [3–5]. While the use of network methods has

90 predominantly been in developed countries due to presence of large live animal

91 movement traceability data sets [6], relatively little has been done in sub-Saharan

92 Africa (SAA) due the paucity of traceability data [7].

93 To date, empirical livestock movement networks in SSA have been constructed from

94 purpose-made surveys, records of official permits granting movement licences or

95 gathering trade records from key locations [8,9]. However, such data gathering

96 efforts are expensive and may lead to data restricted in scope and most often of poor

97 quality. Our previous work has re-constructed a high quality live animal network from

98 such data in Cameroon but it took two years to gather the data and eight months of

99 cleaning and analysis [8]. One of the approaches than can be used to infer networks

100 is gravity modelling, here the amount of interactions that are likely to exist between

101 any two locations are quantified as a function of their population sizes and the

102 distance that separates them [10–13]. As such, it has been used to better

103 understand multi-lateral dynamics behind migrations and flow of goods [10] and can

104 be easily extended to infer a gravity network representing the movements of

105 livestock [14] and humans [10] between populations.

106 Alternatively, livestock movements networks can be inferred from the interactions

107 between livestock hosts and pathogens they carry. This approach exploits the

108 principals of "measurably evolving populations" (MEP) of pathogens through the use

109 of phylogeographic modelling [15]. This assumes that pathogen populations carried

110 by livestock exhibit detectable amounts of *de novo* evolutionary changes in time and

111 space [15]. In other words, the pathogens hitch-hiking on the hosts provides the

112 evolutionary signal (molecular distance) and host locations provides the physical

113 distance. A linear relationship between these two parameters determines the data

114 points to infer a molecular livestock network. In this regard, considerable amount of

115 molecular data has been collected over the years as part of research activities [16–

116 18]. In particular, molecular data on *Mycobacterium bovis*, the causative agent of

117 bovine tuberculosis which currently exist in form of spoligotypes and MIRU-VNTR

118 types for cattle in Cameroon [16,17,19].

119 Leveraging such data not only advances data driven disease control but safe-guards

120 sustainable food security on the continent. However, questions remain on the

121 performance of methods used to capture/infer the topology of empirical networks.

122 Therefore, in this study we aim to evaluate the performance and potential utility of

123 these alternatives, i.e. gravity and phylogeographic modelling methods, to infer

124 livestock movement networks and capture topological and epidemiological features

125 that emerge from empirical networks. To do so, we used data from Cameroon as an

126 exemplar because of its unique location and by extension role in linking cattle

127 movements in central and west Africa[20,21].

128

129 **Materials and Methods**

130

131 **Study area**

132

133 Cameroon is a middle-income country in Central Africa, which strategically connects

134 livestock in Central Africa to West Africa by trading live cattle to neighboring

135 countries such as Nigeria, the Congo and Gabon (**Figure 1**). Cameroon is home to

136 approximately eight million head of cattle providing nutrients to twenty-three million

137 people in Cameroon [17]. Over three quarters of the cattle is reared in the north of

138 Cameroon which supplies live animals for slaughter to abattoirs in the larger cities in

139 the south as well as to some neighboring countries. Pastoral seasonal mobility and

140 Sahel transhumant migration [20], where people and cattle move from as far east as

141 Central African Republic through the northern half of Cameroon to as far west as

142 Mali also occurs [20]. The study focuses on the main cattle rearing areas in the

143 north, as this represents the majority of the cattle movements in Cameroon.

144 Administratively, the country is divided into fifty-eight divisions and then into sub-

145 divisions, it is the latter that form our unity of analysis.

146

147 **Data sources**

148 The empirical data is available from Motta et al 2017 [7], this dataset contains

149 22,698, representing 9% of cattle movements in twelve months of market trading

150   data collected in 2014-2015 from across Cameroon[7] with volumes and direction of

151   flow between markets (**Supplementary DBSX1**). Details of the data collection

152   procedures are described in Motta et al. [10] and in Supplementary DBSX1.

153   The molecular data is available from Egbe et al [17], this dataset contains

154   spoligotyping and MIRU/VNTR data on *Mycobacterium bovis* from 2346 cattle

155   slaughtered in four abattoirs in north of Cameroon between 2012-2014

156   (**Supplementary DBSX2**).

157   The demographic data came from the Cameroonian human and livestock census

158   from 2005-2007 which is available from http://www.statistics-cameroon.org

159   (**Supplementary DBSX3**).To account for the long onset of pathology in bovine

160   tuberculosis, we allow for that a temporal lag of 4-7 years between the census and

161   pathogen molecular datasets [22,23]. We also obtained cartographic information

162   from GADM database of Global Administrative Areas (www.gadm.org) and merged it

163   with respective sub-division.

164

165   **Network construction**

166

167   The network construction is based on twenty sub-divisions common to all three

168   datasets. These sub-divisions account for 5% and 20% of the human and cattle

169   population of Cameroon (**Supplementary DBSX3**).

170

171   *The Empirical Network*

172   Full details of the methods for the development of the empirical network are

173   described in Motta et al. [10]. Briefly, a weighted static directed network was built.

174   Markets formed the nodes of the network and the links between markets were

175   defined as the movement of at least one animal from an origin market to a

176   destination market. For the purposes of this current analysis, data from different

177   market were aggregated at sub-division level, i.e. network nodes were sub-division

178   centroids [10].

179

180   *The Phylogenetic Network*

181   The phylogenetic network was constructed using molecular data for *M. bovis,* a slow

182   evolving pathogen that is endemic in Cameroon [17]. *M. bovis* genotypes were

183   defined as a pattern generated from spoligotyping and a standard 24 Loci MIRU-

184   VNTR [24]. A mutation event was defined as a difference in a MIRU-VNTR type

185   between any two isolates with the same spoligotype [24]. The molecular distance

186   between any two *M. bovis* genotypes was then computed as the number loci to

187   which they were different. The physical distance between the *M. bovis* samples was

188   calculated using the Euclidean distances[25] between the centroids of the sub-

189   division from which the *M. bovis* sample was obtained. Here we assumed that; a)

190   animal-pathogen data with similar temporal and spatial scales contain detectable

191   "historical record/signal of movement ", b) there is a linear relationship between

192   physical distances of cattle (host-host transmission) and molecular distances

193   (pathogen mutation events) [15]. It is this relationship that we exploit to generate the

194   molecular network i.e. all data points that satisfy this linear relationship. The

195   phylogenetic network is henceforth referred to as the molecular network and at this

196   stage, the generated network is undirected, we then introduce directionality in the

197   network using the difference in *M.bovis* genetic diversity between a pairs since any

198   location with a high diversity of *M.bovis* is more likely to be the source of for other

199   locations [17]. A more in-depth description can be found in supplementary materials

200   (**Figure S3&4**).

201

202   *The Gravity Network*

203   Gravity modelling exploits the attraction and repulsive forces between two bodies

204   [10,26,27]. These forces are used to define probable links (i.e. animal movements)

205   between any two sub-divisions with a specific population density of humans and

206   heads of cattle. The attractive and repulsive forces are modelled using a modified

207   gravity formula (eq1); where the gravitation force (δ) here represents the attraction of

208   cattle as a protein source for a given human population in a sub-division (**Figure**

209   **S2&5)**.

210

211    $Gravitation force (\delta) = \left(\frac{Human_D \times Cattle_O}{r_{OD}^2} \right)$ (eq1)

212

213   Where $Cattle_O$ is the size of the cattle population at the sub-division of origin,

214    $Human_D$ is the size of the human population in destination sub-division and r_{OD}^2 is

215   the square distance between the sub-divisions (origin and destination).

216   In order to estimate the relationship betweenδ and cattle movements, we define an

217   optimal threshold by fitting values of δ with presence of a link in the empirical

218   network **Figure S6**. Since we expect an overlap in movements reflected by these

219   two datasets, we then determine the threshold that best reflects an empirical link.

220   This was evaluated by maximising the estimate of the area under the curve (AUC)

221   comparing the presence of links in the network constructed using the gravity model

222   with those present in the empirical network. To examine the practical utility of the

223   gravity modelling approach in the absence of empirical surveillance data, we use the

224   full census data of Cameroon (2005-2007) to generate a country wide cattle

225   movement network. We used the optimised thresholds of δ from the gravity

226   modelling framework as described in the methods section. The resultant directed

227   network topology was then plotted on the Cameroon map using the ggmap package

228   in R version 3.6.

229

230   *Random network topology*

231   To evaluate the extent to which different network inference methodologies reflect

232   systematic differences rather than random noise, a set of random networks were

233   generated with equivalent characteristics to that of the empirical network [28]. We

234   use the Erdos-Renyi model for random sampling of networks and fixed the network

235   parameters (i.e. number of edges and nodes) equivalent to the parameters of the

236   empirical, gravity, and molecular network topologies [28]. We generated 1000

237   random equivalents of each network [28].

238
239   **Networks comparisons**

240   The network-level metrics, described below, were used to compare the phylogenetic

241   and gravity networks to the empirical network. In addition, diseases spread between

242   nodes of these networks was simulated to gain insight on how well each network

243   generating methods may captures "specific epidemics parameters" (Time and

244   number of nodes affected at epidemic peak) and "non-specific epidemics

245   parameters" (e.g. total number of nodes infected) epidemic features relevant for

246   informing mitigation strategies.

247
248   *Network and node level measures*

249   The topological measures computed to describe the structure of the networks[29]

250   included: the total number of links existing in each network, the network density (*d)*,

251   average shortest path length (*PL),* diameter (*di)* and the clustering coefficient (CC)

252   These are defined in **Table S1**.

253   We also calculated node-level metrics for each network including; in-degree, out-

254   degree, betweenness centrality and eigenvector centrality. These were then used to

255   characterise the role of each node.

256   To define the role of each node, we characterised them as; (a) "gate-keeper",

257   characterised by a high betweenness centrality with a low eigenvector centrality, (b)

258   "pulse-taker", a well-connected node characterized by a high eigenvector centrality

259   and low betweenness centrality, or (c) a "dual function" node, characterized by both

260   a high eigenvector and high betweenness centrality [7].

261   To evaluate the degree of information provided by the variability characterized

262   between networks, we compared (topologically, network indices and overlap

263   analysis) the three networks' metrics with those computed from a set of 1000

264   randomly generated networks with the same number of edges and nodes as the

265   considered networks. Here, all networks were generated and analysed using the

266   'igraph' package in the statistical software R version 3.6.

267

268   *Temporal representativeness*

269   The steps above will have generated static molecular and gravity networks which

270   represent all paths defined in a specific temporal window[30]. In reality all paths do

271   not exist at all times, so static networks tend to create accessibility pathways in the

272   network which are not always there, such an error is quantifiable as a comparison

273   between the number of paths in a static network and a time-series network[30]. In

274   this case, we split the empirical network into Quarters 1 to 4 of a year as reported in

275   [7]. We then define temporal fidelity (C) as be shown below;

276

$$C_{molecular} = \frac{Number\,of\,paths \in the\,temporal\,empirical\,network\,(Q1-Q4)}{Number\,of\,paths \in the\,static\,molecular\,network}$$

277
278

$$C_{gravity} = \frac{Number\,of\,paths \in the\,temporal\,empirical\,network\,(Q1-Q4)}{Number\,of\,paths \in the\,static\,gravity\,network}$$

279
280

281   A temporal fidelity(C) of 1 means the temporal network is well represented by the

282   static network, on the other hand a lower (C) means the opposite. The causal error

283   defined as overall overestimation of disease outbreaks if the static network is use is

284   approximated as $Causal\ error = \frac{1}{C}$

285

286

287   *Disease transmission potential for networks*

288

289   To evaluate the disease implications of the contact structure inferred by the

290   generated networks, we simulate disease spread using a discrete SEIR model in

291   *Epinet* package. The SEIR model included parameters for the infection rate (i.e. the

292   probability of contagion after contact per unit time, β), the recovery rate, (defined as

293   the probability of recovery per unit time, γ), and the rate of removal of infectious

294   nodes (θ_i).

295   To seed the infection, we fixed $\beta = 0.3$, $\gamma=1$ and θ_i as this was the lowest parameter

296   values at which the disease propagates through the network regardless of where it is

297   seeded on the 20 nodes. The disease spread simulation was repeated for 200

298   separate runs. The latency period was set to zero. In order to evaluate real-life

299   scenario, we repeated the above simulation across a range of parameters i.e. $\beta = 0$ -

300   0.31 and $\gamma\ = 0$ - 2 simulating rapidly spreading diseases such as foot-and-mouth

301   disease (FMD) and contagious bovine pleural pneumonia (CBPP) in a super herd

302   setting [31,32]. In all cases, a fixed recovery rate, $\theta_i = 5$, was used. The same

303   process was done for each of the random networks. We analyse the overlap of

304   epidemic characteristics from each network presented as a density plot in *ggplot2*

305   package in R version 3.6.

306

307   We also generated 1000 random-equivalents for each network topology and

308   repeated the simulation process on each network [28]. The coefficient of overlap

309   between networks[33] on a density plot for "specific" epidemic parameters was

310   assessed between the empirical and all the other network topologies including the

311   random network. Total overlap was interpreted as 100% agreement between

312   networks. Furthermore, we examined the impact of removing a node on disease

313   spread in each of network. Here we simulate disease spread ($\beta = 0.2$, $\gamma = 1$, $\theta_i = 5$)

314   and sequentially dropping one node and seeding the infection at each of the
315   remaining nodes for 200 iterations. We then compare epidemic statistics from these
316   simulations with dropped nodes to a baseline full network. The epidemic statistics
317   here were the mean number of infected nodes and the percentage of model runs
318   where there was take off (>1 node infected)

319
320
321
322
323   **Results**
324
325   ***A comparison of network and node level measures***
326   **Figure 2** highlights the topological similarities, in term of presence and direction of
327   flow, of the inferred Gravity network and Molecular network with the empirical
328   network. However, Gravity network contained more edges (routes) when compared
329   to empirical networks and Molecular network. The gravity network (134) captures
330   nearly three times as many edges as the empirical (43) and twice as many as the
331   molecular networks (53), **Table 1**.
332   A comparison of un-directed and directed networks shows that 40% and 22% of
333   edges in molecular network are present in the empirical network, respectively.
334   Similarly, the gravity and empirical networks share 41% and 27% of undirected and
335   directed edges, respectively (**Table S1**). However, better agreement is observed
336   between molecular and gravity i.e. 64.8% and 92.5% in directed and un-directed
337   networks respectively. Interestingly, all three networks identified the same critical link
338   between cattle populations of Cameroon. On the other hand, all three random
339   networks exhibited topological characteristics which were considerably different from
340   their respective network equivalents. For example; route in terrain that could not
341   support animal movement (**Figure 2 & Table 1**).
342   Compared to the density of the empirical network (d =11%), the molecular network
343   was of similar density (d =14%) whereas the gravity network was denser (d =35%).
344   While these findings highlight a higher clustering coefficient (CC) for Molecular
345   network and Gravity network, it is noteworthy that the former exhibits the shortest
346   path length. In general, we observe considerable differences between these and
347   their random equivalents on almost all measures **(Table 1)**.

348 By plotting the relationship between the local centrality measures eigenvector and
349 betweenness, we can characterize the role of each node in the network (**Figure S6**).
350 In doing so, we identify nodes that are "pulse-takers" (a well-connected node
351 characterized by a high eigenvector centrality and low betweenness centrality), and
352 those that have a dual-purpose (a node characterized by both a high eigenvector
353 and high betweenness centrality). Here, Banyo and Ngaoundere played a dual-
354 purpose role, with the former exhibiting this in both empirical and gravity networks.
355 However, the same was not true for Banyo in the molecular network, this role is
356 instead taken on by Fundong. We note that most "pulse-takers" were located in the
357 North West region of Cameroon. At national level (applying the gravity model to the
358 full Cameroonian census dataset), an emerging trend of animal movements are from
359 Northwest and the Adamawa towards the populous cities in the south like Yaoundé,
360 Douala and Buea is evident. Here we note that in addition to Banyo, Tibati acts as a
361 critical link between the Northwest and the Adamawa. Furthermore, Touboro and
362 Ngale are likely important transboundary hubs in the West and East of Cameroon
363 respectively (**Figure 5A & B**).

364 ***Temporal representativeness***
365 To evaluate the temporal representativeness of our networks i.e. if they consistently
366 capture pathways when they exist, we compare them with a quarterly empirical
367 network. Here we note that Gravity network's temporal fidelity ranged between 0.201
368 & 0.238 while that of Molecular network varied between 0.111 & 0.222 (**Table 2**).
369 Note that the activity of pathways varied across the year which is why the temporal
370 fidelity, although higher than the rest, ranged between 0.76 for first quarter and 0.88
371 for the fourth quarter. The causal error i.e. the propensity to overestimating an
372 outbreak when the static gravity and molecular network were used was on average
373 at 4.5 and 5.8 respectively. This error was lowest for the static empirical network
374 (1.17).
375

376 ***A comparison of network disease transmission potential***
377 Comparison based on epidemic characteristics (time to peak infection versus nodes
378 infected at peak), showed that it took the same time to infect nodes in the empirical
379 network, molecular and gravity network, but the later had more nodes infected at the
380 end **Figure 3.** Indeed, the level of overlap between molecular and empirical on these
381 epidemic characteristics was on average over 98% and approximately 59% between

382 empirical and gravity networks. The random networks over-estimates both time to

383 infections peak and total nodes infected when compared to the empirical. When this

384 comparison is done over a range of epidemic parameters i.e. Rate of spread versus

385 Rate of recovery as shown in **Figure 4A**, the overlap between the molecular and

386 empirical network remains high, on average 0.75. Note that the higher the spread

387 and recovery rates the lower overlap coefficients between molecular and empirical

388 networks. The comparison between gravity and empirical exhibits poor overlap. On

389 the other hand, when we consider epidemic parameters such as the mean number of

390 nodes infected at the end of the epidemic and the percentage disease take off

391 (**Figure 4B & C**), the characteristics captured are comparable. Here the overlap is

392 higher at high spread and recovery rates

393

394

395 **Discussion**

396

397 We set out to compare methods used to infer livestock movement networks using

398 Cameroon as an exemplar of a low middle-income setting with limited empirical data

399 availability. We used phylogeographic and gravity modelling to repurpose freely

400 available and archived data. We show that molecular data from tools regarded as

401 out-dated in developed countries infers network structures and disease

402 characteristics similar to empirical data. In comparison, although there was lesser

403 agreement between Gravity and empirical network, valuable higher scale inferences

404 can be obtained from it. Therefore, repurposing data with these approaches could

405 provide and/or complement empirical data to unravel livestock disease dynamics in

406 developing countries.

407

408 ***The molecular network and disease dynamics***

409 The molecular network output provided the best approximation to the empirical

410 network in terms of topology and epidemic characteristics. This finding suggests

411 added potential of using archived spoligotype and MIRU-VNTR *M. bovis* data to infer

412 networks [19]. Furthermore, the molecular network captured more routes than the

413 empirical network. While the empirical network is restricted to information captured at

414 the cattle markets, the molecular network can be used to infer unobserved events

415 through the clonal evolutionary pathway of *Mycobacterium bovis* [15,34]. Since *M.*

416 *bovis* is a slow evolving clonal bacteria, mutational changes in MIRU-VNTR and
417 spoligotyping take between 5-12 years to occur [35,36][19]. Agreement between the
418 empirical and molecular network suggest the following; a) there is a temporal
419 consistency in cattle movements in Cameroon and b) bovine tuberculosis must be an
420 endemic disease in this setting. Indeed, Motta et al[7] suggests consistency in cattle
421 movement between markets in Cameroon [7] and it has long been suggested that
422 bovine tuberculosis is endemic in this region [37]. The high costs of generating this
423 molecular data hampers its utility for this purpose in LMICs, for example; it costs
424 approximately €60 to genotype a sample using spoligotype and 24 loci MIRU-VNTR
425 [17], this unit cost would be too high for the tool's routine use for surveillance.
426 However, where this data exists in archives possibly generated through research or
427 surveillance, it can be repurposed to support regional and national livestock disease
428 control. Note that we propose its use at coarse rather than a fine scale because our
429 method involved merging animal market data per subdivision.
430

431 ***The gravity network and disease dynamics***
432 The gravity network captured nearly three times the number of routes as the
433 empirical network. Since the gravity network is generated from census data, it likely
434 represents a wide range of movements across an animal's whole life time hence the
435 high number of generated movements/edges. However, not all the routes/edges
436 captured by the gravity network will be actual movement routes. This false positive
437 limitation has been reported elsewhere [12,13], this could be minimized by fitting,
438 socio-economic and environmental parameters to improve the model
439 predications[5,10,38]. However, this model refinement was beyond the scope of this
440 work.
441 The difference in number of routes between the gravity and empirical networks is
442 characterized by a low level of overlap of epidemic characteristics, particularly when
443 we consider epidemic parameters such as size and speed to peak infection. This
444 could be because we assumed that all movements in the gravity networks are driven
445 by the demand of the population of consumers, i.e. the human demand for meat and
446 dairy products being the central pull for cattle in the region. In reality, cattle move
447 due to a variety of reasons including, the search for pasture, breeding, political
448 insecurity as well as trade [20]. We however observe better agreement when
449 epidemics parameters such as the total size of epidemic or proportion of infections

450   that take off are considered, in such cases the gravity network captures similar

451   characteristics. This therefore suggests that the utility of this network ought to be

452   limited to non-specific aspects of disease control and management.

453   The large number of routes the gravity network generates inherently comes with

454   some level of reduced specificity. Therefore, prudence should be exercised when

455   interpreting such results. In particular, we note that the gravity network can fail to

456   capture key nodes because of the alternative routes it generates around these key

457   nodes. It is more than likely that the gravity network provides a worst-case scenario

458   for preparedness, i.e. provides guidance on the maximum required resources to

459   manage an outbreak. Refining these broad inferences requires extra-information, in

460   this case a molecular or empirical network would be needed to guide specific

461   strategies like nodes to target for disease control. It is noteworthy that the empirical

462   network was used as a reference for optimizing the gravity model, this practically

463   means the utility of the former requires some basic empirical data. The same is

464   however not true for the molecular network, and its close approximation to empirical,

465   means the molecular could ideally be used instead to optimized the gravity model in

466   the absence of empirical data.

467   **Validity of inferences**

468   Static networks such as the one generated here suffer a weakness of representing

469   movement paths/routes that are not always present [30,39]. To evaluate the size of

470   this error in our networks, we calculated their temporal fidelity which suggest that the

471   first and last quarter represent unique characteristics given the lower fidelity. In this

472   regard the overall overestimation error was 4.5 and 5.8, interestingly the gravity

473   slightly out-performing the molecular network.

474

475   **Practical utility of molecular and gravity networks**

476   Practically, the gravity model could be a good alternative to extract contact

477   information in settings where surveillance level is limited or absent, and enable the

478   development of the early response strategies against future livestock disease

479   outbreaks. Where molecular data exists, it could complement empirical data to refine

480   such strategies, i.e. where empirical data is incomplete the molecular network can fill

481   in gaps based on pathogen mutation changes. Combined, these three layers of data

482   not only highlight the potential to cost effectively improve responses and

483   preparedness to livestock outbreaks but also provides a foundation for data driven

484 disease management in countries like Cameroon. We note with interest that all
485 generated networks show that animals move from the north regions to provide
486 protein to the large cities in the south, an observation previously reported in
487 Cameroon [7]. Furthermore, the methods show potential to identify important trade
488 and transboundary animal movement hubs such as Ngale, Tchollire and Touboro
489 [7,16,17]. We however observe limited movements in the northern part of Cameroon
490 especially in Garoua, likely because the region has approximately the same human
491 and animal population. This suggests, that in areas where human and animal
492 populations are approximately the same size gravity network performs poorly.
493 Crucially, the gravity model exploits livestock protein demand, therefore its network
494 topology inherently reflects livestock economic forces [13] so, overlaying epidemic
495 simulations on this provides a unique opportunity to combine trade forces with
496 epidemiological disease information. Infectious diseases cost African economies
497 billions of dollars in reduced productivity, lost revenues due to quarantine and
498 restricted market access [31,32]. By exploiting the contact structure inferred by our
499 generated networks, we demonstrate potential to utilise molecular and gravity
500 modelling to inform diseases preparedness and control strategies[6]. This is of
501 particular value especially because the molecular and census data used here is
502 considered obsolete and archived respectively in developing countries.
503 We recognise the potential bias that could arise as result of using a network work
504 with fewer nodes, this was on balance done to ensure the same node were
505 compared across all networks. However this bias mitigated by the by comparison
506 with random equivalent networks [7,29].
507
508 **Conclusion**
509
510 To conclude, we have adopted method which exploits freely available demographic
511 and archived molecular data to develop livestock network which could be used for
512 livestock disease control in developing countries. We show that the molecular
513 network which is a close approximation to the empirical network can be generated
514 from data considered obsolete in wealthier countries. The gravity model captures the
515 largest proportion of movements but likely overestimates movements and disease
516 epidemics. Better agreement is achieved across all three networks if less specific
517 epidemic characteristics such as the size of outbreak are investigated. Moreover,

518   these networks identify the same important animal movement and trade hotspots.

519   We therefore consider this proof-of-concept that archived census and molecular data

520   could be repurposed to inform livestock disease management in LMICs.

521   **Acknowledgements**

522

529

530

531

532   **References**

533   1.   Luke DA, Harris JK. Network Analysis in Public Health: History, Methods, and
534   Applications. Annu Rev Public Health. 2007.
535   doi:10.1146/annurev.publhealth.28.021406.144132

536   2.   Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et
537   al. GetReal in network meta-analysis: a review of the methodology. Res Synth
538   Methods. 2016. doi:10.1002/jrsm.1195

539   3.   Moslonka-Lefebvre M, Gilligan CA, Monod H, Belloc C, Ezanno P, Filipe JAN, et al.
540   Market analyses of livestock trade networks to inform the prevention of joint economic
541   and epidemiological risks. J R Soc Interface. 2016. doi:10.1098/rsif.2015.1099

542   4.   Green DM, Kiss IZ, Mitchell AP, Kao RR. Estimates for local and movement-based
543   transmission of bovine tuberculosis in British cattle. Proc Biol Sci. 2008;275: 1001–5.
544   Available:
545   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2366193&tool=pmcentrez&
546   rendertype=abstract

547   5.   Chaters GL, Johnson PCD, Cleaveland S, Crispell J, De Glanville WA, Doherty T, et
548   al. Analysing livestock network data for infectious disease control: An argument for
549   routine data collection in emerging economies. Philos Trans R Soc B Biol Sci.
550   2019;374. doi:10.1098/rstb.2018.0264

551   6.   Schrödle B, Held L, Rue H. Assessing the Impact of a Movement Network on the
552   Spatiotemporal Spread of Infectious Diseases. Biometrics. 2012. doi:10.1111/j.1541-
553   0420.2011.01717.x

554   7.   Motta P, Porphyre T, Handel I, Hamman SM, Ngu Ngwa V, Tanya V, et al.
555   Implications of the cattle trade network in Cameroon for regional disease prevention
556   and control. Sci Rep. 2017. doi:10.1038/srep43932

557   8.   Paolo M, Ian H, Saidou M H, Rob C, Victor N, Vincent T, et al. Implications of the
558   cattle trade network in Cameroon for regional disease prevention and control. Nat Sci
559   Reports. 2016.

560   9.   Dean AS, Fournié G, Kulo AE, Boukaya GA, Schelling E, Bonfoh B. Potential Risk of
561   Regional Disease Spread in West Africa through Cross-Border Cattle Trade. PLoS
562   One. 2013. doi:10.1371/journal.pone.0075570

563   10.   Eichengreen B, Irwin D a. The Role of History in Bilateral Trade Flows. The

564   Regionalization of the World Economy. 1998. doi:10.3386/w5565

565 11. Zipf GK. The P 1 P 2 D Hypothesis: On the Intercity Movement of Persons. Am Sociol
566   Rev. 1946. doi:10.2307/2087063

567 12. Apolloni A, Nicolas G, Coste C, El Mamy AB, Yahya B, El Arbi AS, et al. Towards the
568   description of livestock mobility in Sahelian Africa: Some results from a survey in
569   Mauritania. PLoS ONE. 2018. doi:10.1371/journal.pone.0191565

570 13. Nicolas G, Apolloni A, Coste C, Wint GRW, Lancelot R, Gilbert M. Predictive gravity
571   models of livestock mobility in Mauritania: The effects of supply, demand and cultural
572   factors. PLoS One. 2018. doi:10.1371/journal.pone.0199547

573 14. Simini F, González MC, Maritan A, Barabási AL. A universal model for mobility and
574   migration patterns. Nature. 2012. doi:10.1038/nature10856

575 15. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the
576   genomic era. Trends in Ecology and Evolution. 2015. doi:10.1016/j.tree.2015.03.009

577 16. Egbe NF, Muwonge A, Ndip L, Kelly RF, Sander M, Tanya V, et al. Abattoir-based
578   estimates of mycobacterial infections in Cameroon. Sci Rep. 2016;6.
579   doi:10.1038/srep24320

580 17. Egbe NF, Muwonge A, Ndip L, Kelly RF, Sander M, Tanya V, et al. Molecular
581   epidemiology of Mycobacterium bovis in Cameroon. Sci Rep. 2017;7.
582   doi:10.1038/s41598-017-04230-6

583 18. Ghebremariam MK, Hlokwe T, Rutten VPMG, Allepuz A, Cadmus S, Muwonge A, et
584   al. Genetic profiling of Mycobacterium bovis strains from slaughtered cattle in Eritrea.
585   PLoS Negl Trop Dis. 2018;12. doi:10.1371/journal.pntd.0006406

586 19. Muwonge A, Egbe F, Bronsvoort M, Areda DB, Hlokwe T, Michel A. Molecular
587   Epidemiology of Mycobacterium bovis in Africa. In: Dibaba AB, Kriek NPJ, Thoen CO,
588   editors. Tuberculosis in Animals: An African Perspective. Cham: Springer
589   International Publishing; 2019. pp. 127–169. doi:10.1007/978-3-030-18690-6_8

590 20. SWAC. Promoting and Supporting Change in Transhu- mant Pastoralism in the Sahel
591   and West Africa. Paris; 2007. Available:
592   http://www.oecd.org/swac/publications/38402714.pdf

593 21. Cour JM. The Sahel in West Africa: Countries in transition to a full market economy.
594   Glob Environ Chang. 2001;11: 31–47.

595 22. Cassidy JP. The pathogenesis and pathology of bovine tuberculosis with insights from
596   studies of tuberculosis in humans and laboratory animal models. Veterinary
597   Microbiology. 2006. pp. 151–161.

598 23. Neill SD, Pollock JM, Bryson DB, Hanna J. Pathogenesis of Mycobacterium bovis
599   infection in cattle. Vet Microbiol. 1994;40: 41–52. Available:
600   http://www.ncbi.nlm.nih.gov/htbin-
601   post/Entrez/query?db=m&form=6&dopt=r&uid=0008073627

602 24. Reyes JF, Tanaka MM. Mutation rates of spoligotypes and variable numbers of
603   tandem repeat loci in Mycobacterium tuberculosis. Infect Genet Evol. 2010;10: 1046–
604   51. doi:10.1016/j.meegid.2010.06.016

605 25. Hua H, Xie H, Tanin E. Is euclidean distance really that bad with road networks?
606   IWCTS 2018 - Proceedings of the 11th ACM SIGSPATIAL International Workshop on
607   Computational Transportation Science. 2018. doi:10.1145/3283207.3283215

608 26. Luoma M, Palomäki M. A New Theoretical Gravity Model and Its Application to a
609   Case with Drastically Changing Mass. Geogr Anal. 1983;15: 14–27.
610   doi:10.1111/j.1538-4632.1983.tb00756.x

611   27.   Bossenbroek JM, Kraft CE, Nekola JC. Prediction of long-distance dispersal using
612   gravity models: Zebra mussel invasion of inland lakes. Ecol Appl. 2001;11: 1778–
613   1788.

614   28.   Barabási A-L, Albert R. Emergence of Scaling in Random Networks. Science (80-).
615   1999;286: 509–512.

616   29.   Wasserman S, Faust K. Centrality and Prestige. Social Network Analysis. 2012.
617   doi:10.1017/cbo9780511815478.006

618   30.   Lentz HHK, Koher A, Hövel P, Gethmann J, Sauter-Louis C, Selhorst T, et al. Disease
619   spread through animal movements: A static and temporal network analysis of pig
620   trade in Germany. PLoS One. 2016;11. doi:10.1371/journal.pone.0155196

621   31.   Kobayashi M, Carpenter TE, Dickey BF, Howitt RE. A dynamic, optimal disease
622   control model for foot-and-mouth-disease:. II. Model results and policy implications.
623   Prev Vet Med. 2007;79: 274–286.

624   32.   Lesnoff M, Laval G, Bonnet P, Chalvet-Monfray K, Lancelot R, Thiaucourt F. A
625   mathematical model of the effects of chronic carriers on the within-herd spread of
626   contagious bovine pleuropneumonia in an African mixed crop-livestock system. Prev
627   Vet Med. 2004;62: 101–117.

628   33.   Charaudeau S, Pakdaman K, Boëlle PY. Commuter mobility and the spread of
629   infectious diseases: Application to influenza in France. PLoS One. 2014.
630   doi:10.1371/journal.pone.0083002

631   34.   Tanaka MM, Francis AR. Detecting emerging strains of tuberculosis by using
632   spoligotypes. Proc Natl Acad Sci U S A. 2006;103: 15266–71. Available:
633   http://www.pnas.org/content/103/41/15266.full

634   35.   Tanaka MM, Phong R, Francis AR. An evaluation of indices for quantifying
635   tuberculosis transmission using genotypes of pathogen isolates. BMC Infect Dis.
636   2006;6: 92. Available: http://www.ncbi.nlm.nih.gov/pubmed/16756684

637   36.   Reyes JF, Chan CHS, Tanaka MM. Impact of homoplasy on variable numbers of
638   tandem repeats and spoligotypes in Mycobacterium tuberculosis. Infect Genet Evol.
639   2012;12: 811–818.

640   37.   Awah-Ndukum J, Kudi a C, Bah GS, Bradley G, Tebug SF, Dickmu PL, et al. Bovine
641   tuberculosis in cattle in the highlands of cameroon: seroprevalence estimates and
642   rates of tuberculin skin test reactors at modified cut-offs. Vet Med Int. 2012;2012:
643   798502. Available:
644   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3329900&tool=pmcentrez&
645   rendertype=abstract

646   38.   Reuben J, A. DB, M. OA. Determinants of live animals and animal products trade
647   within the ECOWAS sub-region: A gravity model approach. J Dev Agric Econ. 2014.
648   doi:10.5897/JDAE2013.0485

649   39.   Lentz HHK, Selhorst T, Sokolov IM. Spread of infectious diseases in directed and
650   modular metapopulation networks. Phys Rev E - Stat Nonlinear, Soft Matter Phys.
651   2012;85. doi:10.1103/PhysRevE.85.066111

652

653   **Table legend**

654   Table 1: The comparison for network characteristics

Network Parameters	Empirical Network	Molecular Network	Gravity Network	Random Network*

Network Density	0.113	0.142	0.352	0.113(0.112-0.114)
Clustering coefficient	0.483	0.703	0.74	0.203(0.202-0.204)
Diameter	4.00	3.00	5.00	7.43(7.413-7.452)
Average shortest path length	2.80	3.01	1.87	2.61(2.14- 6.87)

655 *Random Network the parameters are estimated for 1000 randomly generated networks
656 based on the empirical network parameters
657
658
659
660
661
662 Table 2: Shows the assessment of temporal fidelity

Temporal empirical network	Static networks					
	Gravity		Molecular		Empirical	
	Fidelity	Error	Fidelity	Error	Fidelity	Error
Q1	0.201	4.975	0.111	9.009	0.767	1.303
Q2	0.238	4.202	0.222	4.504	0.883	1.132
Q3	0.238	4.202	0.222	4.504	0.883	1.132
Q4	0.238	4.202	0.148	6.756	0.883	1.132

663
664 **Figure Legend**
665
666 **Figure 1.** Map of Cameroon showing the cattle raring regions highlighted in blue, yellow,
667 green and red. We also show the locations pertinent to this work. The map is plotted in Qgis
668 version 3
669 **Figure 2.** Generated networks i.e. empirical, molecular gravity and the random network
670 plotted on a base map from Open Street Maps using *ggmaps* package in R.
671 **Figure 3.** Density plot of time to peak infection and the maximum number of nodes infected.
672 The blue and red colour represents the dynamics of the empirical and the respective network
673 topology. This plot is generated from the output of two hundred runs per node per network
674 topology i.e. 4000 data points per network.
675 **Figure 4. A-** Coefficient overlap between the networks. Here we compare the empirical and
676 its random equivalent to the molecular and empirical network dynamics. The coefficient of
677 overlap is fitted for a range of Rate of spread (β) and γ Recovery rates (p) values on the y
678 and x axis respectively. **B-** Coefficient overlap between the networks when we consider
679 percentage of infections that take off and **C** is when mean number of nodes infected at the
680 end of the infection. Here B and C represent the "non-specific epidemics parameters"
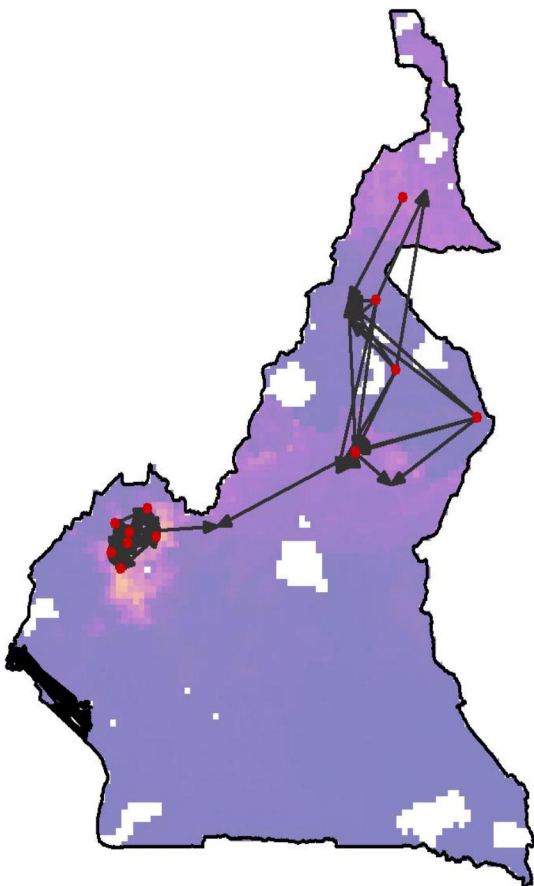
681   **Figure 5A.** National-wide cattle movement routes generated from the demographic data
682   using thresholds optimised for gravity model. **B**. Key actor analysis done to identify the
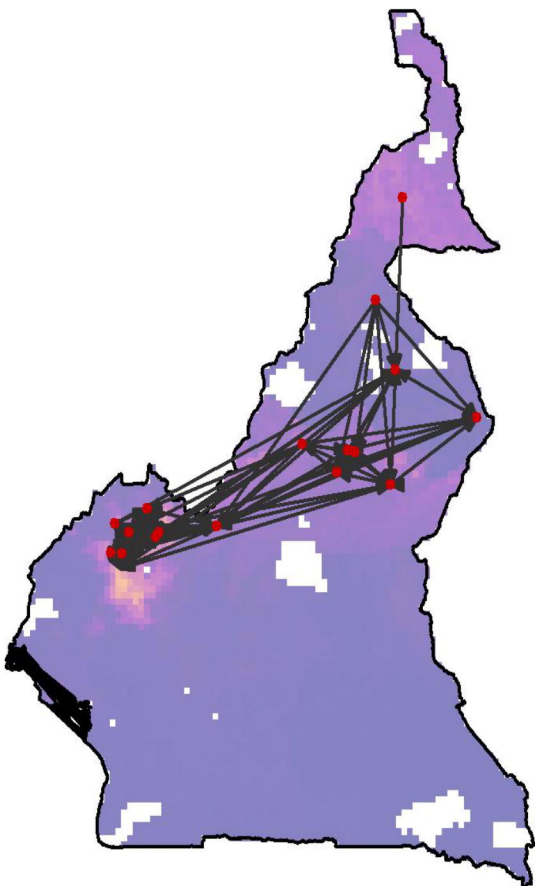683   important nodes for cattle movement
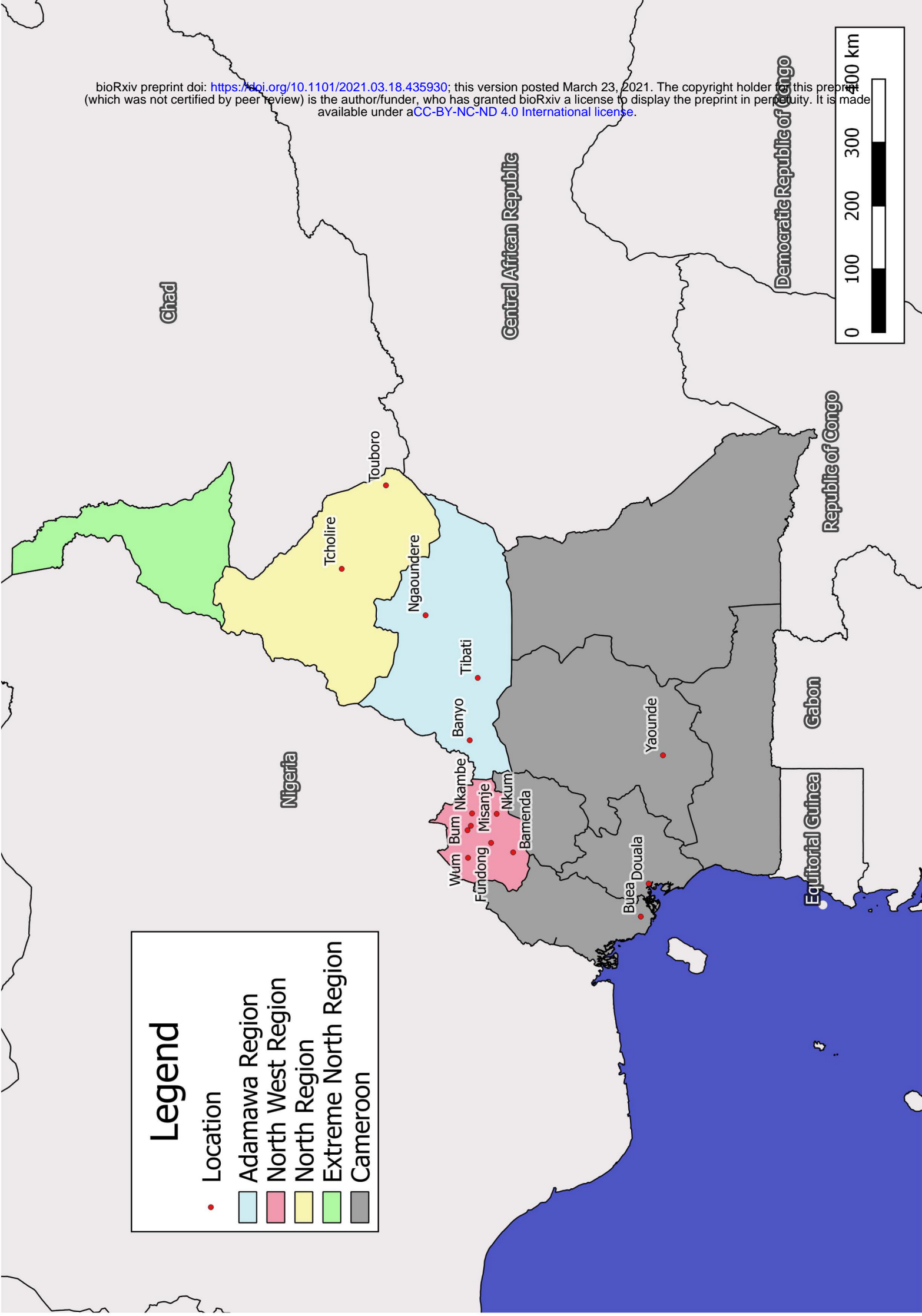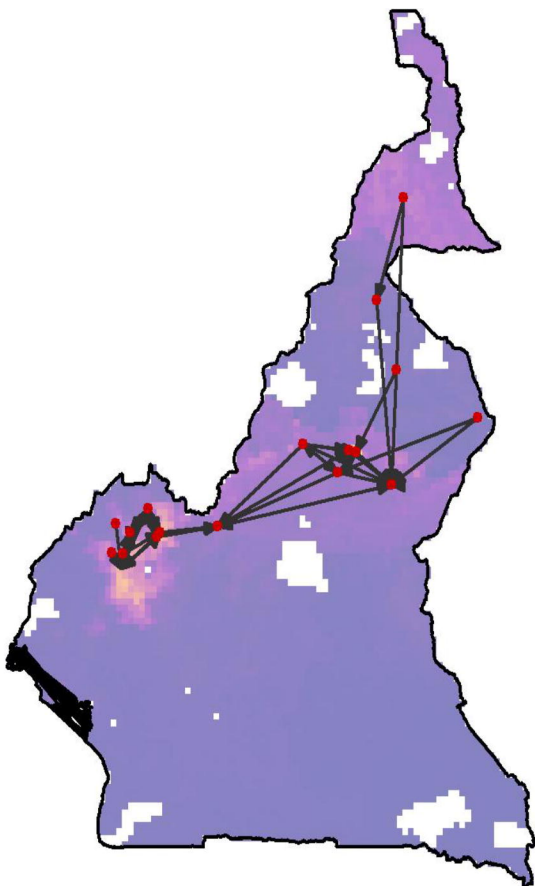
Chad

Central African Republic

Democratic Republic of Congo

Republic of Congo

100 km

0   100   200   300   400

Touboro

Tcholire

Ngaoundere

Banyo

Tibati

Nigeria

Nkambe

Bum   Misanje   Nkum

Wum   Fundong   Bamenda

Yaounde

Buea   Douala

Gabon

Equitorial Guinea

Legend

Location

Adamawa Region
North West Region
North Region
Extreme North Region
Cameroon

Emperical network toplogy

Cattle density

5000   10000

Molecular network toplogy

Cattle density

5000   10000

Gravity network toplogy

Cattle density

5000   10000

Random-Empirical network toplogy
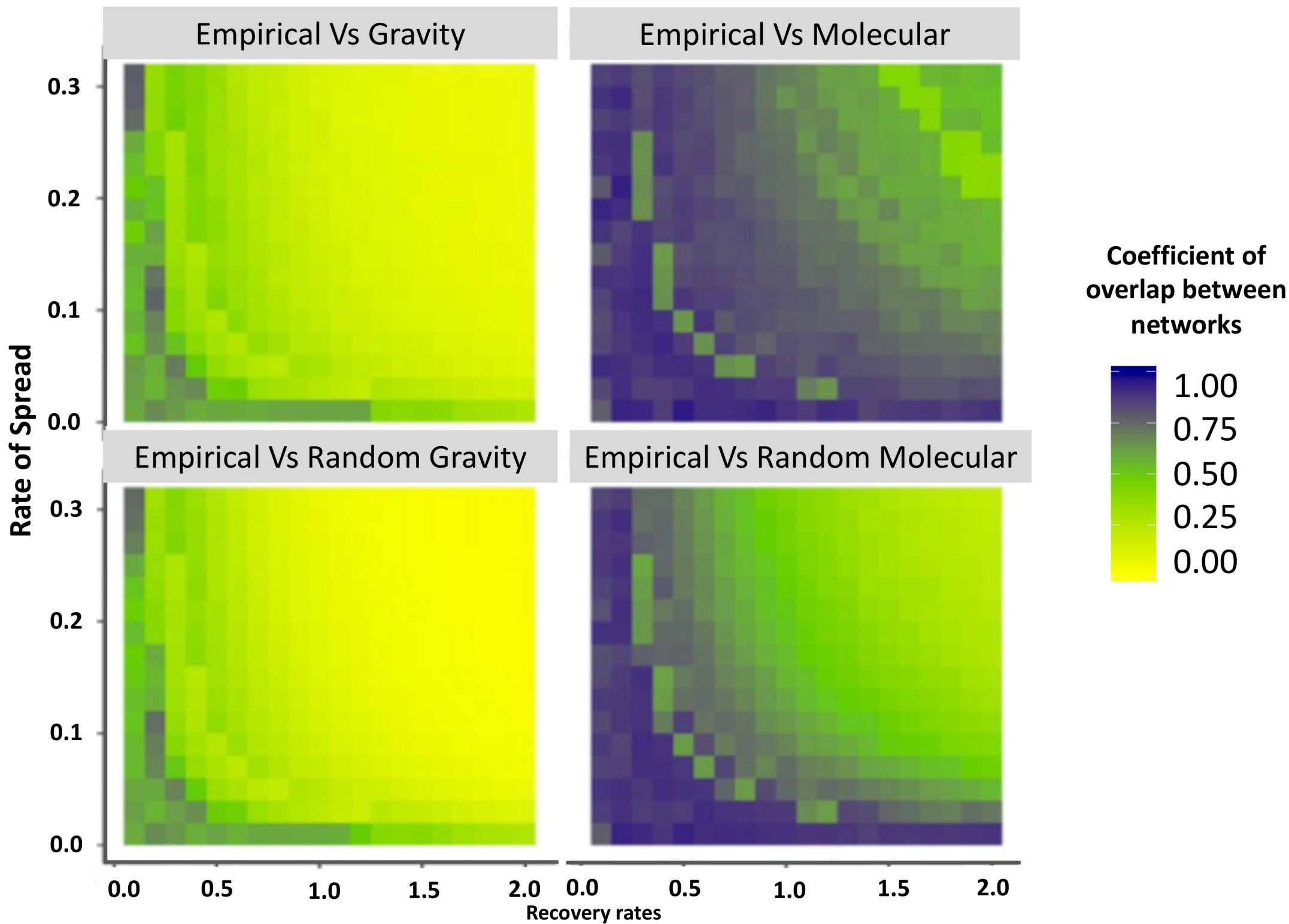

Cattle density
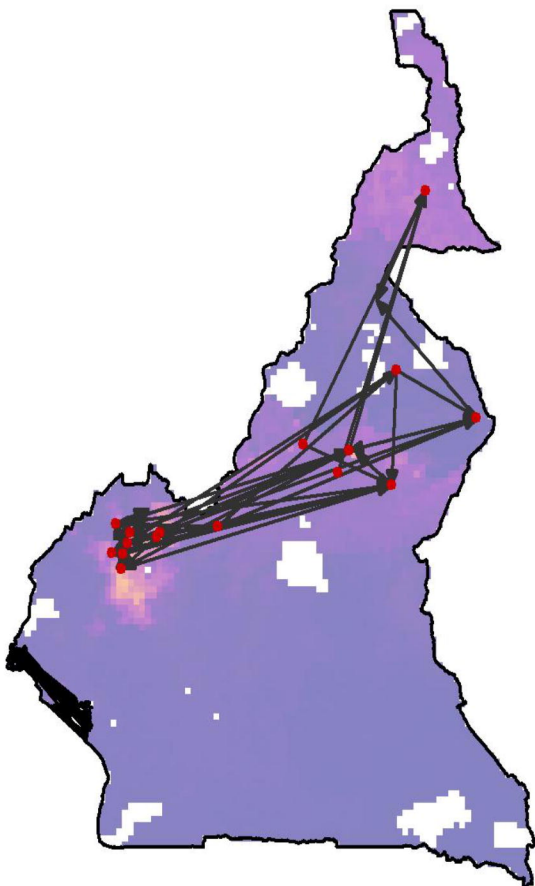
5000   10000

Random-Molecular network toplogy

Cattle density
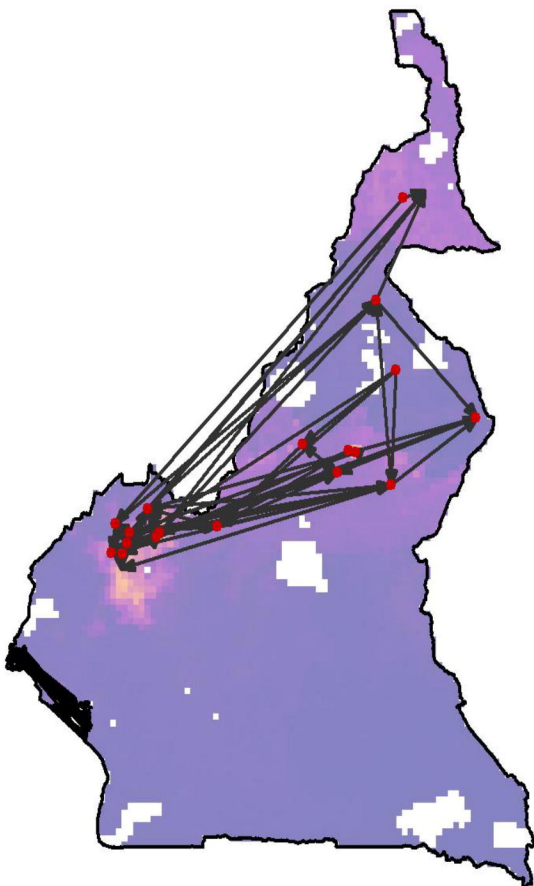
5000   10000

Random-Gravity network toplogy

Cattle density

5000   10000

Tibati

Tchollire

Banyo

Misaje

Touboro

Bamenda

Ngaoundere

Ngale

Nyambaka

Buea

Yaounde

Douala