Using GIS-linked Bayesian Belief Networks as a Tool for Modelling Urban Biodiversity

Supplementary materials

Extended method details: Biodiversity input data (supplement to section 2.4)

Plants, invertebrates and birds were sampled across the study area using a random stratified design and mapped as point data. A 500 m × 500 m grid was placed across the entirety of each urban area, and each grid square (hereafter 'tile') was classified into one of 25 urban form classes according to (a) percentage building cover (five evenly spaced categories, based on LiDAR and Ordnance Survey data), and (b) percentage cover of vegetation over 0.5 m tall (five evenly spaced categories, based on LiDAR). Five representatives of each urban form class were randomly selected, where available, yielding a total of 112 survey tiles (Figure 1). An additional four tiles were added for bird surveys only, from a range of urban forms.

Invertebrates were selected for study on the basis of representing a significant component of urban biodiversity, and were relatively feasible to sample in concert with aboveground plants. Identifying invertebrates to order level was the highest precision feasible across all groups. Litter organisms were a sub-group of total invertebrates and were singled out for more precise identification due to their association with structurally diverse habitats and the availability of in-house expertise. Coleoptera and Diptera identification also represent subsets of total invertebrates, but greater precision was possible and desirable for these groups given their ubiquity and diverse ecological functions, and available expertise. Plants represent a key component of urban greenspace and so were important to sample at a total level; neophytes and native plants were isolated for individual consideration due to the expectation that native richness may be more driven by underlying ecological conditions whereas neophyte richness may be more driven by human intervention. Lastly, birds were selected as they can be easily identified to species level in the field by trained surveyors, and are important ecological indicators as well as being important to human nature experience.

Bird surveys were conducted at four representative points in each tile at locations associated with areas of green space. Points were distributed across the tile, ≥ 200 m apart from each other and ≥ 100 m from tile edges, and where they could be accessed. Only three points were surveyed in a small number of tiles due to access issues or coverage by unsuitable habitat (e.g. open water). Assuming an effective survey radius of 200 m for singing birds from each point in a tile, survey coverage was close to complete in terms of area. Two 10-minute surveys were conducted at each point between 06:00 am and 10:00 am in May and June 2013 and 2014 following standard avian monitoring times in the UK (Harris et al., 2014). Observers recorded all birds seen or heard, including those in flight but below vegetation height or those feeding aerially within the survey site, within 200 m of the observation point. Overall richness values for modelling input were based on the maximum observed richness at each point in a given year, then averaged across the two years of observation (described in Grafius et al., 2017).

Bird surveys were designed to provide a standardized measure of community composition each point. Detectability varies between species, but surveys were timed so as to maximize detectability and to sample both earlier- and later-breeding species, while repeating surveys in each of two consecutive years reduces the probability of species being missed due to stochastic variation in behaviour or the influences of weather. It is possible that locally rare and inconspicuous species were missed, but such birds either cannot be numerically significant elements of the local community or are not closely associated with the habitat at a given survey point (e.g. wide-ranging species whose dependence on habitat cannot be captured effectively at the scale of a point with effective sampling over a 200m radius). Note also that our focus on overall relative species richness made detectability less of a perceived concern than it would have been if per-species abundance had been a research focus, because apparent abundance will be far more sensitive to habitat-specific variation in detectability.

Plants and invertebrates were surveyed between 2 July and 16 August 2013 within areas of continuous green space associated with the 112 survey tiles. Sites were identified by randomly selecting one of the four bird survey points in each tile, and all continuous green spaces connected to that point were considered as a site. Within each site, surveys were conducted in square 25 m^2 plots (although shape could vary for equivalent area), distributed across vegetation structural typologies (short grass, long grass, flowering herbaceous, shrubs, trees) in approximate proportion to their area present in the site. This resulted in 244 plots within 78 sites (average 3 plots $site^{-1}$; range 2-4). Plant surveys were conducted in three randomly located 1 m x 1 m quadrats in each plot. Cover was estimated using the Domin scale (Rodwell, 2006). Some taxa of planted specimens were only identified to genus level, due to the absence of key diagnostic features (e.g. flowers, fruits etc.), or the lack of critical keys or Floras. The entire 25 m^2 plot was surveyed when it was dominated by shrubberies or flowering herbaceous plantings, as these plots contain larger plants for which small quadrats are not appropriate. All plant data were summarised at the 25 m^2 plot level. Relevant species were categorised as native (excluding archaeophytes) or neophyte, following PLANTATT (Hill et al., 2004) and Stace (2010), supported by expert advice for species not included in these sources (OP, author).

Sweep net sampling was used to capture airborne insects along a 10 m transect in each plot (adjusted to fit plot dimensions, typically two parallel 5 m sections). Vacuum samples were taken from two circular, physically delimited areas 1.5 m tall by 0.5 m diameter to capture invertebrates on vegetation and in the leaf litter layer. All specimens were identified to order. Coleoptera and Diptera were identified to family and checked by expert recorders for each group. Isopoda, Diplopoda, Chilopoda, and Pseudoscorpiones (hereafter termed litter organisms) were identified to species by an expert recorder (JPR, author).

These sampling approaches yielded datasets on the richness of nine taxonomic groups: total invertebrates (to order level), litter organisms (to species level), Coleoptera (to family level), Diptera (to family level), total non-tree plants (to species level, including all vascular plants from the ground, field and shrub layers), native non-tree plants (to species level, as above), neophyte non-tree plants (to species level, as above), and birds (to species level).

Mapped modelling input parameters
Input parameters for the biodiversity BBN models included: landscape connectivity modelled for urban woodland birds using a circuit theory approach (Grafius et al., 2017), habitat patch area (ha), and LiDAR-derived vegetation height (m) (Table S1). Maps of these datasets are shown below (Figures S1-S3). The connectivity map was aggregated by spatial mean to 25 m resolution for the primary analysis described in the paper's main text.

Bayesian Belief Network (BBN)-based modelling of biodiversity at high resolution (5 m)
In the paper's main text we focus our analysis on 25 m resolution spatial data for two reasons: first, it more appropriately represents the scale of land use/land cover (LULC) data likely to be available to most researchers and urban planners, thus taking on a practical relevance; and second, because the landscape-scale perspective of biodiversity our research was concerned

with was believed to be more appropriately investigated at this coarser scale. However, the nature of the wider project that funded and encompassed this research ('Fragments, Functions, Flows and Urban Ecosystem Services' or F3UES) enabled access to unusually high-resolution data including LULC cover at 5 m resolution (2 m for LiDAR vegetation height). While not believed to be as appropriate for the primary focus of this research, we took the opportunity to generate and investigate BBN models to predict taxonomic richness at this finer scale. Below are included the resulting maps of predicted biodiversity (Figures S4-S11) as well as tables presenting the result map summary statistics (Table S2) and model error rates and parameter sensitivities (percent variance reduction statistics) (Table S3).

References:

Grafius, D.R., Corstanje, R., Siriwardena, G.M., Plummer, K.E., Harris, J.A., 2017. A bird's eye view: using circuit theory to study urban landscape connectivity for birds. Landsc. Ecol. 32, 1771–1787. https://doi.org/10.1007/s10980-017-0548-1

Grafius, D.R., Corstanje, R., Warren, P.H., Evans, K.L., Hancock, S., Harris, J.A., 2016. The impact of land use/land cover scale on modelling urban ecosystem services. Landsc. Ecol. 31, 1509–1522. https://doi.org/10.1007/s10980-015-0337-7

Harris, S.J., Risely, K., Massimino, D., Newson, S.E., Eaton, M.A., Musgrove, A.J., Noble, D.G., Procter, D., Baillie, S.R., 2014. The Breeding Bird Survey 2013 (BTO Research Report 658). Thetford.

Hill, M.O., Preston, C.D., Roy, D.B., 2004. PLANTATT - Atributes of British and Irish Plants: Status, Size, Life history, Geography and Habitats, PLANTATT. Centre for Ecology & Hydrology, Huntingdon, UK.

McRae, B.H., Dickson, B.G., Keitt, T.H., Shah, V.B., 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology 89, 2712–2724. https://doi.org/10.1890/07-1861.1

Rodwell, J.S., 2006. National Vegetation Classification: Users' Handbook. Joint Nature Conservation Committee, Peterborough, UK.

Stace, C., 2010. New Flora of the British Isles, Ed. 3. ed. Cambridge University Press, Cambridge.

Table S1: Summary table of datasets for landscape factor input data/biodiversity predictors. All were available at 5 m and 25 m resolution and used to conduct comparative analyses at these respective scales.

Dataset	Data type	Units	Source
Patch Area	Digital raster map	Hectares	Calculated from land cover raster; see main text
Connectivity	Digital raster map	Normalised cumulative current (see McRae et al., 2008)	(Grafius et al., 2017)
Vegetation Height	Digital raster map	Meters	(Grafius et al., 2016)

Table S2: Summary statistics of model prediction raster maps (5 m resolution) for invertebrate, plant and bird richness. 'Litter organisms' include species from Isopoda, Diplopoda, Chilopoda, and Pseudoscorpiones. Values are based on mean model results at each pixel.

	Mean	Std. Dev.	Min.	Max.
Invertebrate Order Richness	11.6	1.1	12	15
Litter Organisms Species Richness	1.8	1.1	1	7
Coleoptera Family Richness	4.1	0.5	4	6
Diptera Family Richness	10.5	1.4	5	15
Total Plant Species Richness	15.3	2.5	8	24
Native Plant Species Richness	11.5	1.7	7	18
Neophyte Plant Species Richness	3.2	2.2	1	11
Bird Species Richness	13.1	0.8	11	17

Table S3: Results of case testing (error rate) and sensitivity analysis (percent variance reduction) on Bayesian Belief Network models for invertebrate, litter organism, Coleoptera, Diptera, plant and bird richness at 5 m resolution. For each taxonomic group the landscape factor showing the greatest sensitivity is shown in bold.

	Error rate	Sensitivity (Percent Variance Reduction)		
		Vegetation Height	*Connectivity*	*Patch Area*
Invertebrate Order Richness	59.43%	**8.28**	4.80	1.57
Litter Organisms Sp. Richness	52.87%	**8.86**	2.88	1.61
Coleoptera Family Richness	64.34%	**3.24**	0.97	2.08
Diptera Family Richness	64.34%	**3.78**	2.57	2.92
Total Plant Sp. Richness	66.80%	**3.65**	3.32	1.31
Native Plant Sp. Richness	65.98%	3.12	2.40	**3.52**
Neophyte Plant Sp. Richness	45.08%	**9.00**	0.92	0.57
Bird Sp. Richness	74.06%	3.14	2.07	**3.20**

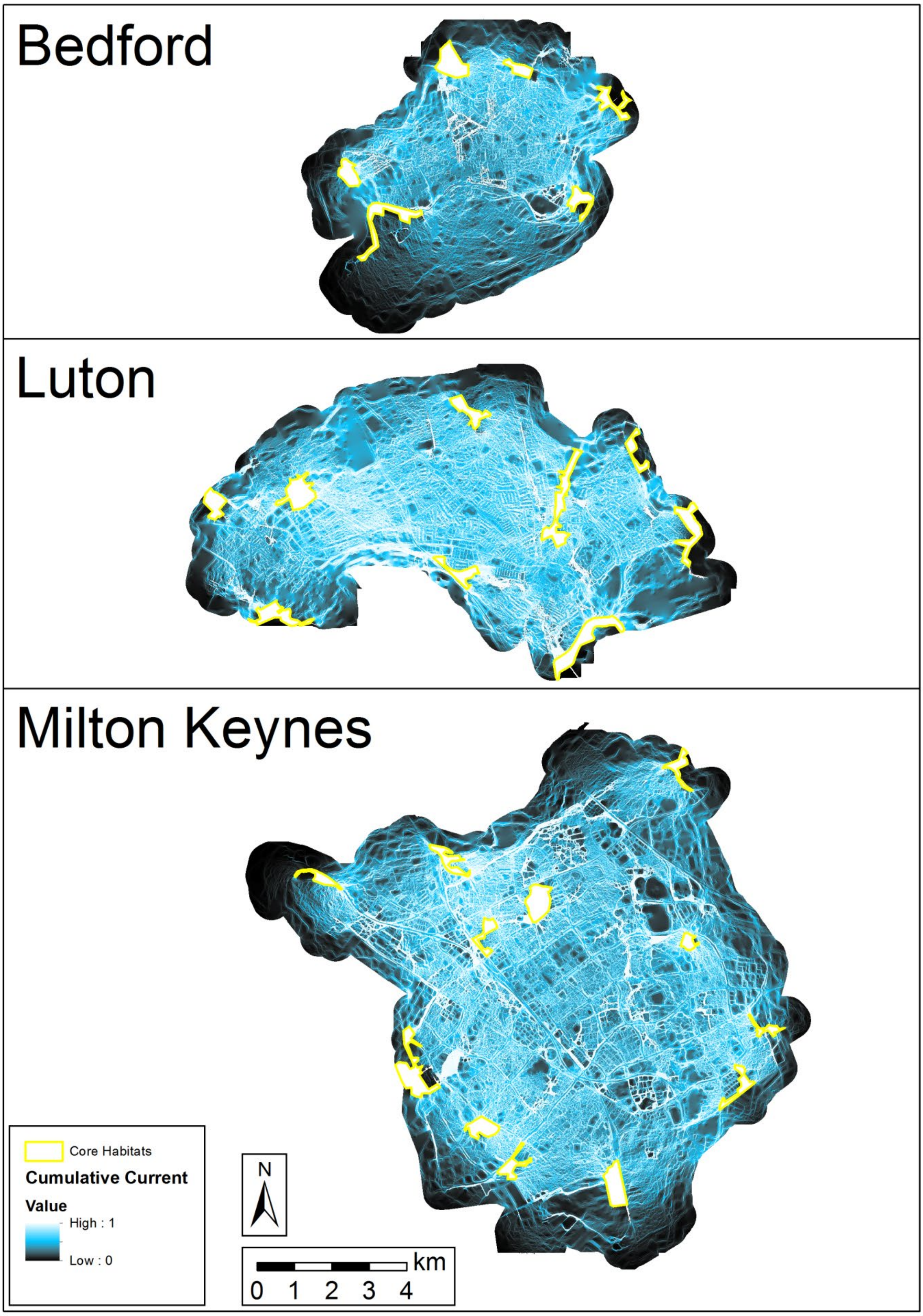

Figure S1:Modelled connectivity using a circuit theory approach for Bedford, Luton and Milton Keynes, UK (from Grafius et al., 2017).

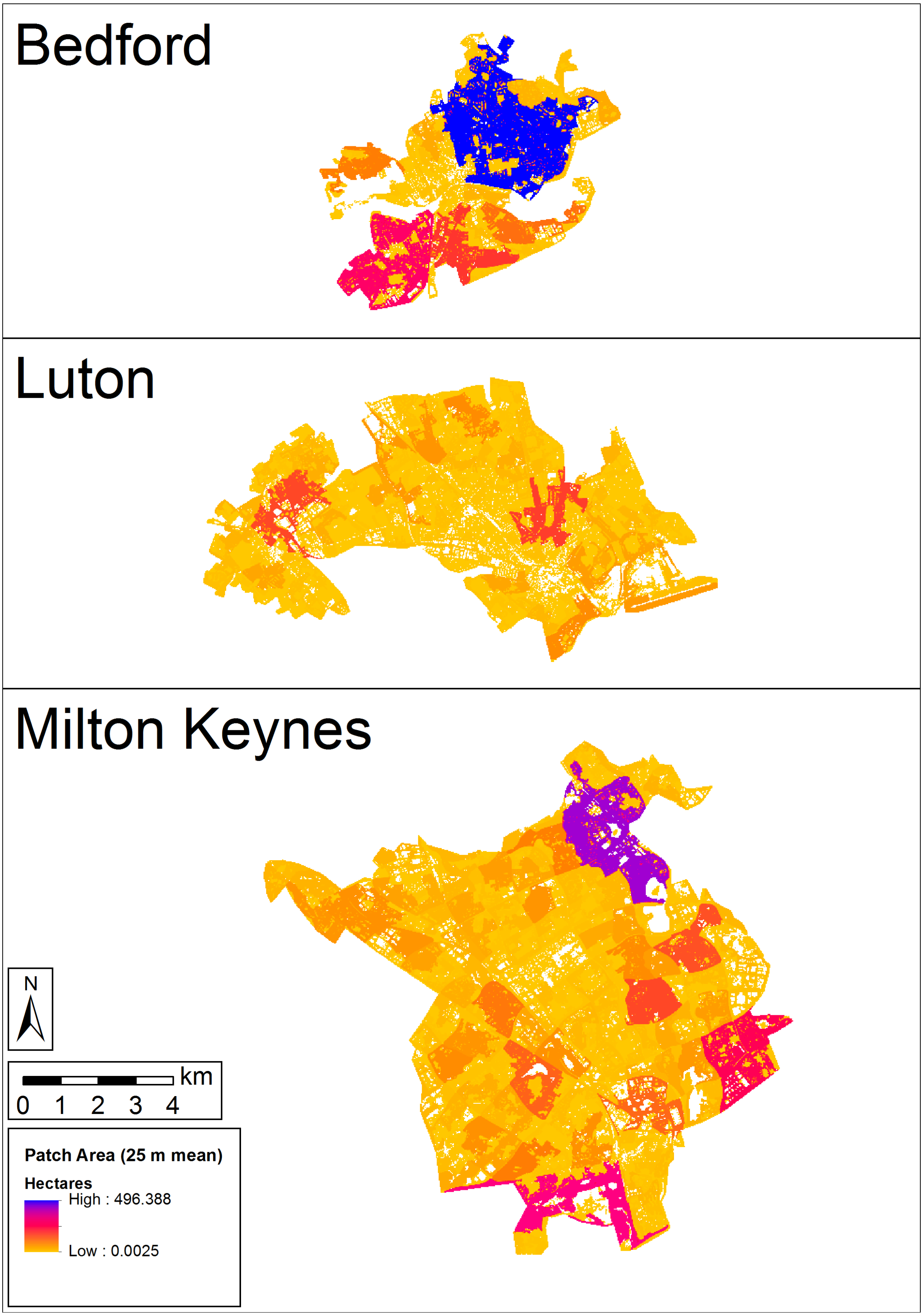

Figure S2: Mean habitat patch area based on 5 m land cover raster map, aggregated by areal mean value to 25 m, for Bedford, Luton and Milton Keynes, UK.

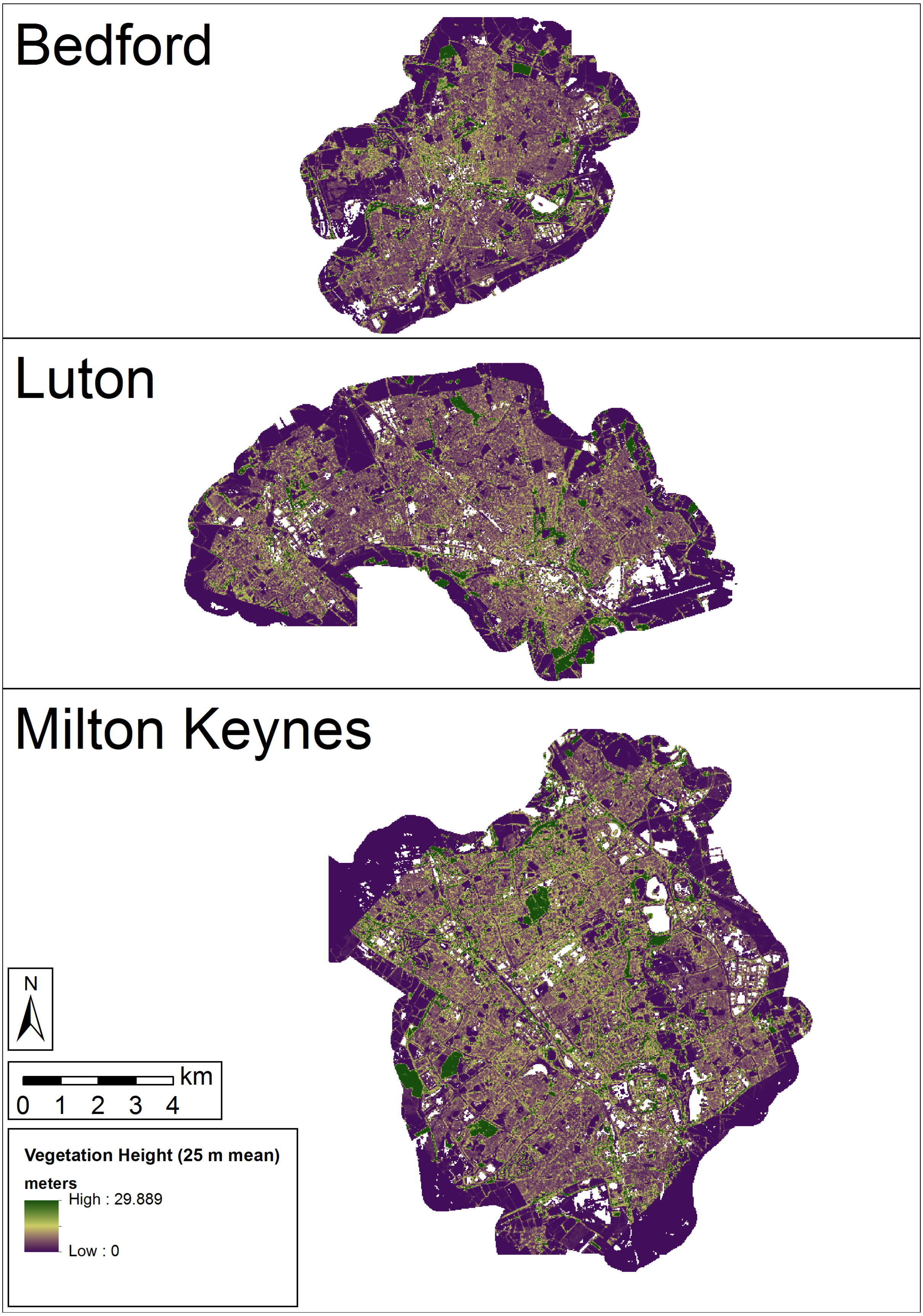

Figure S3: LiDAR-derived vegetation height (2 m aggregated to 25 m resolution by areal mean) for Bedford, Luton and Milton Keynes, UK.

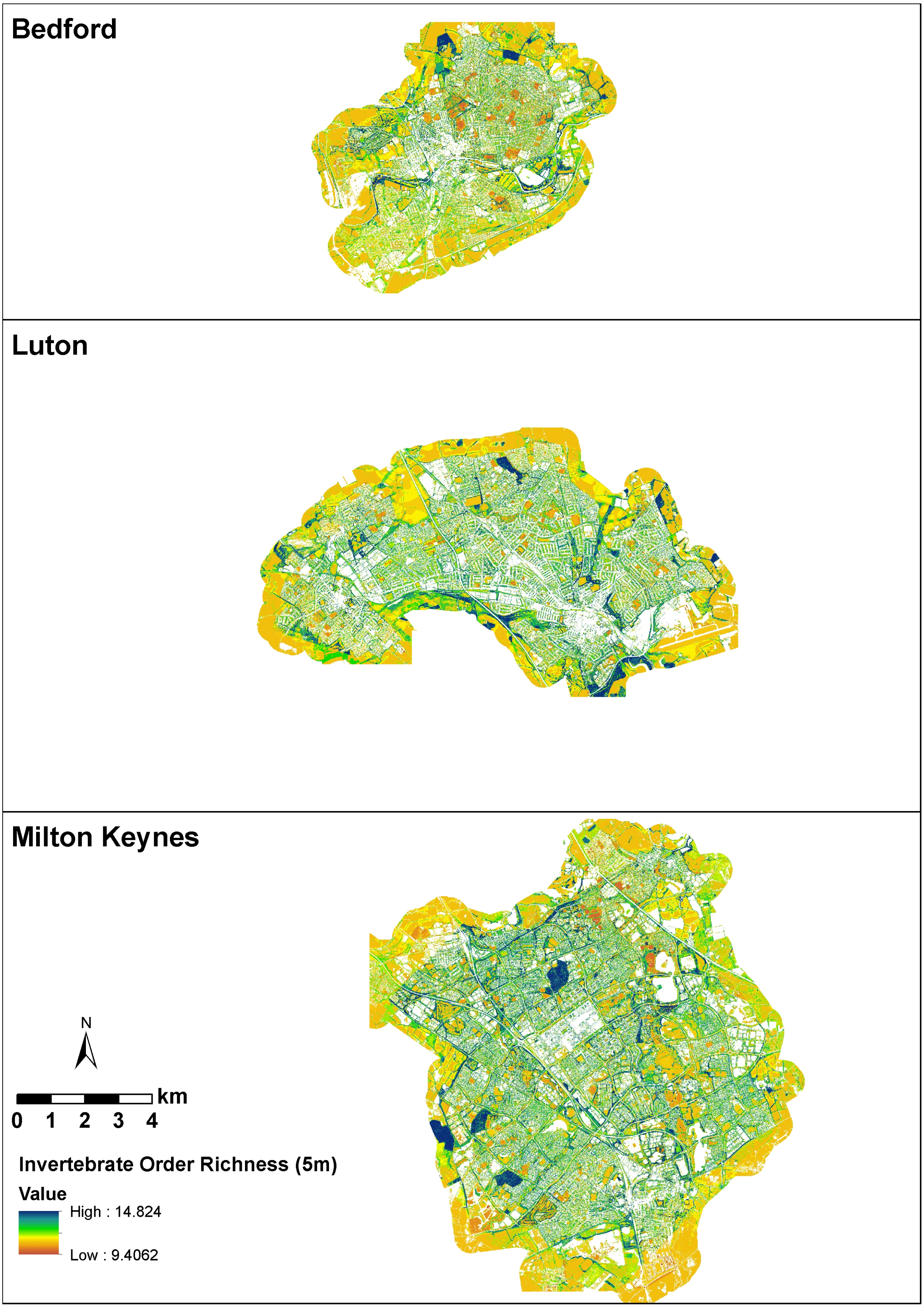

Figure S4: Modelled order-level invertebrate richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

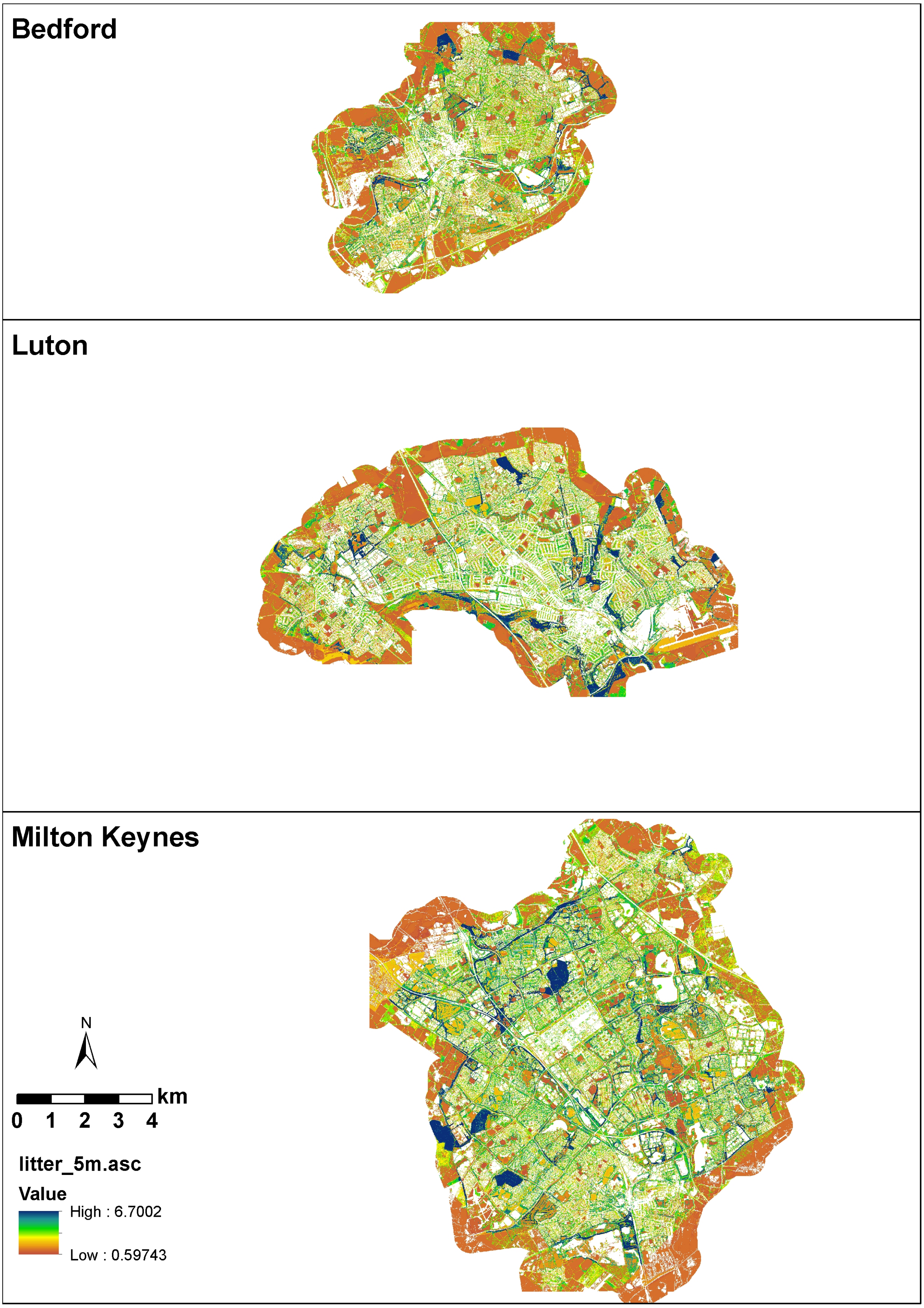

Figure S5: Modelled species-level low-mobility litter organism (Isopoda, Diplopoda, Chilopoda, and Pseudoscorpiones) richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

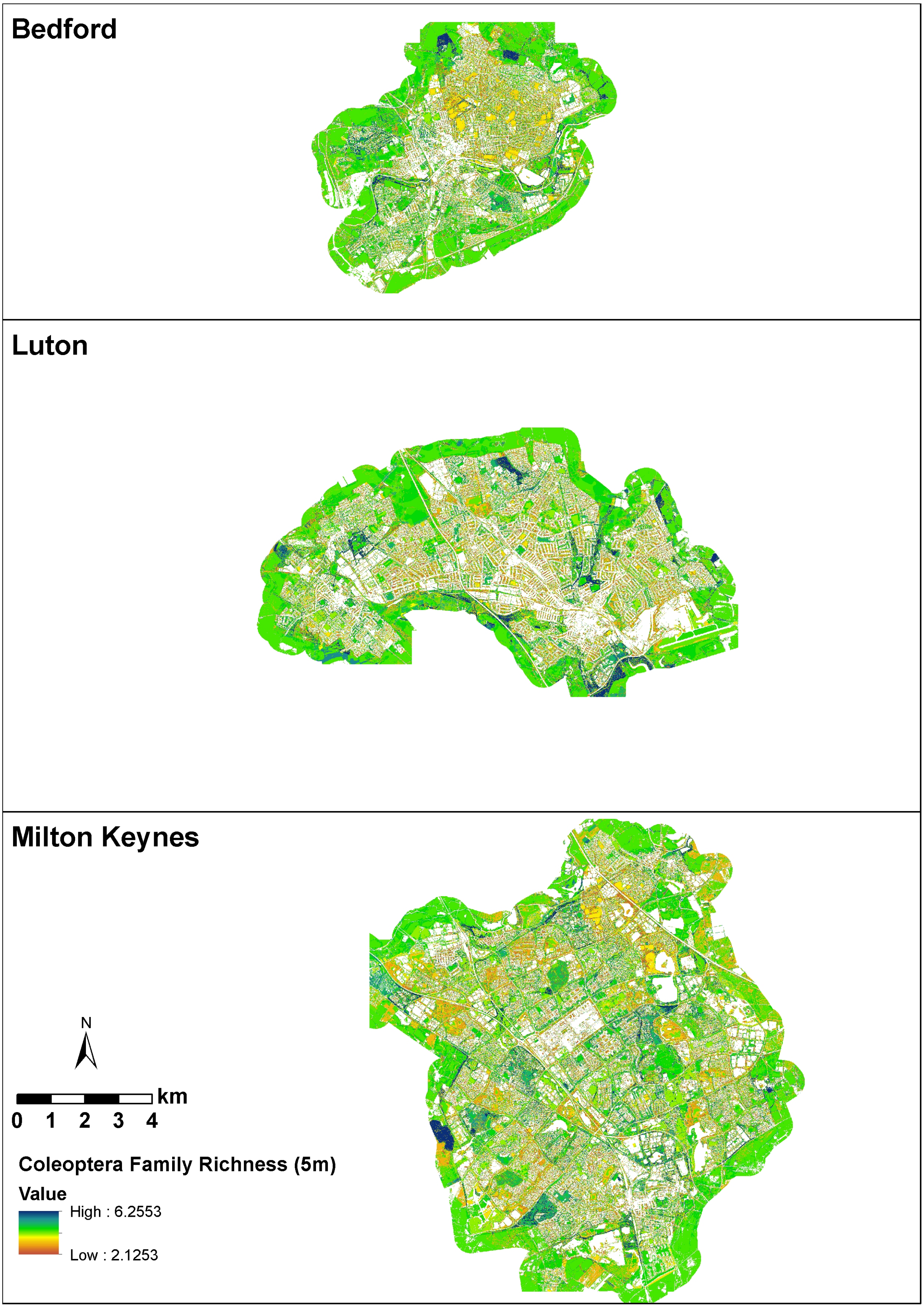

Figure S6: Modelled family-level Coleoptera richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

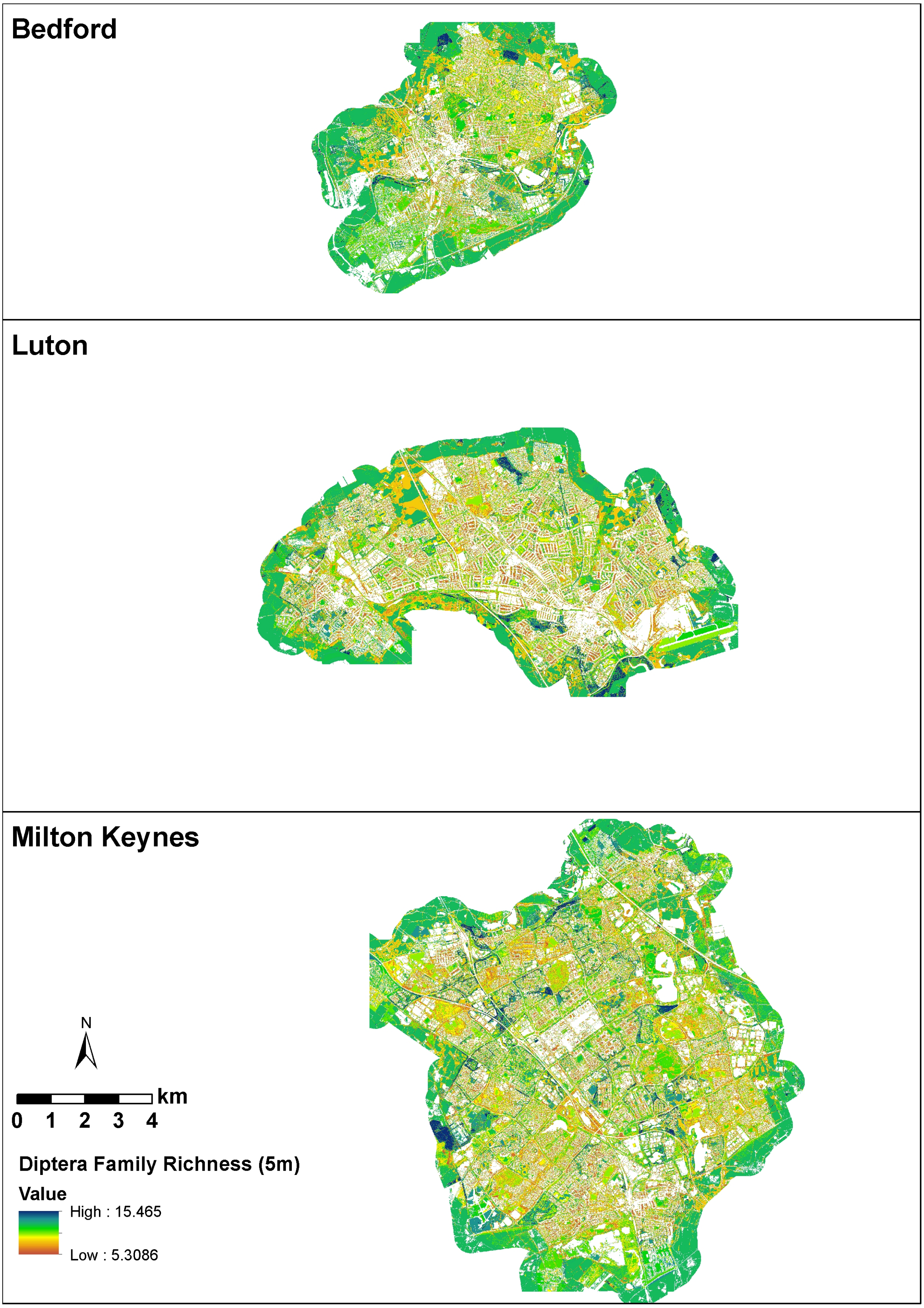

Figure S7: Modelled family-level Diptera richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

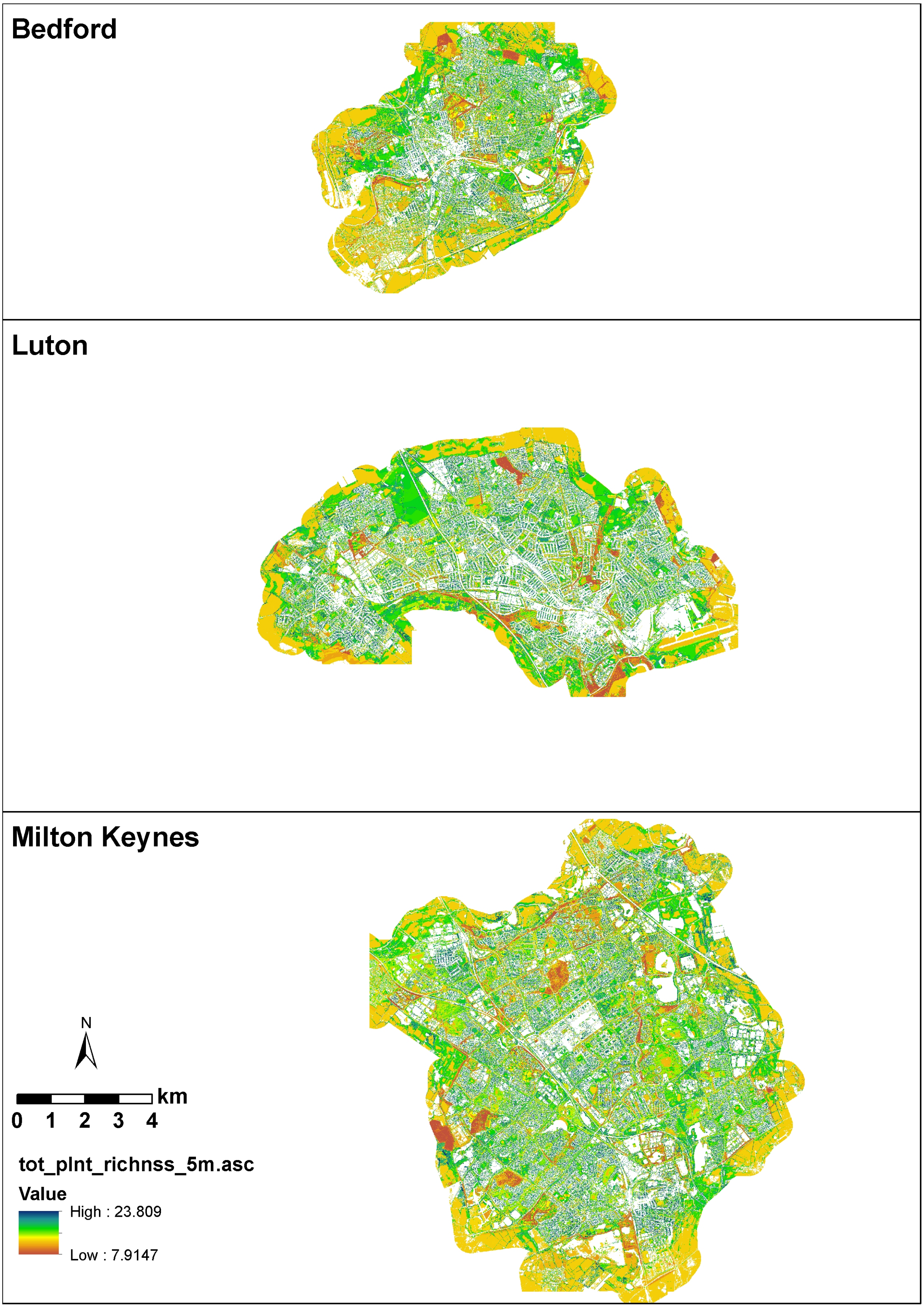

Figure S8: Modelled total non-tree plant species richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

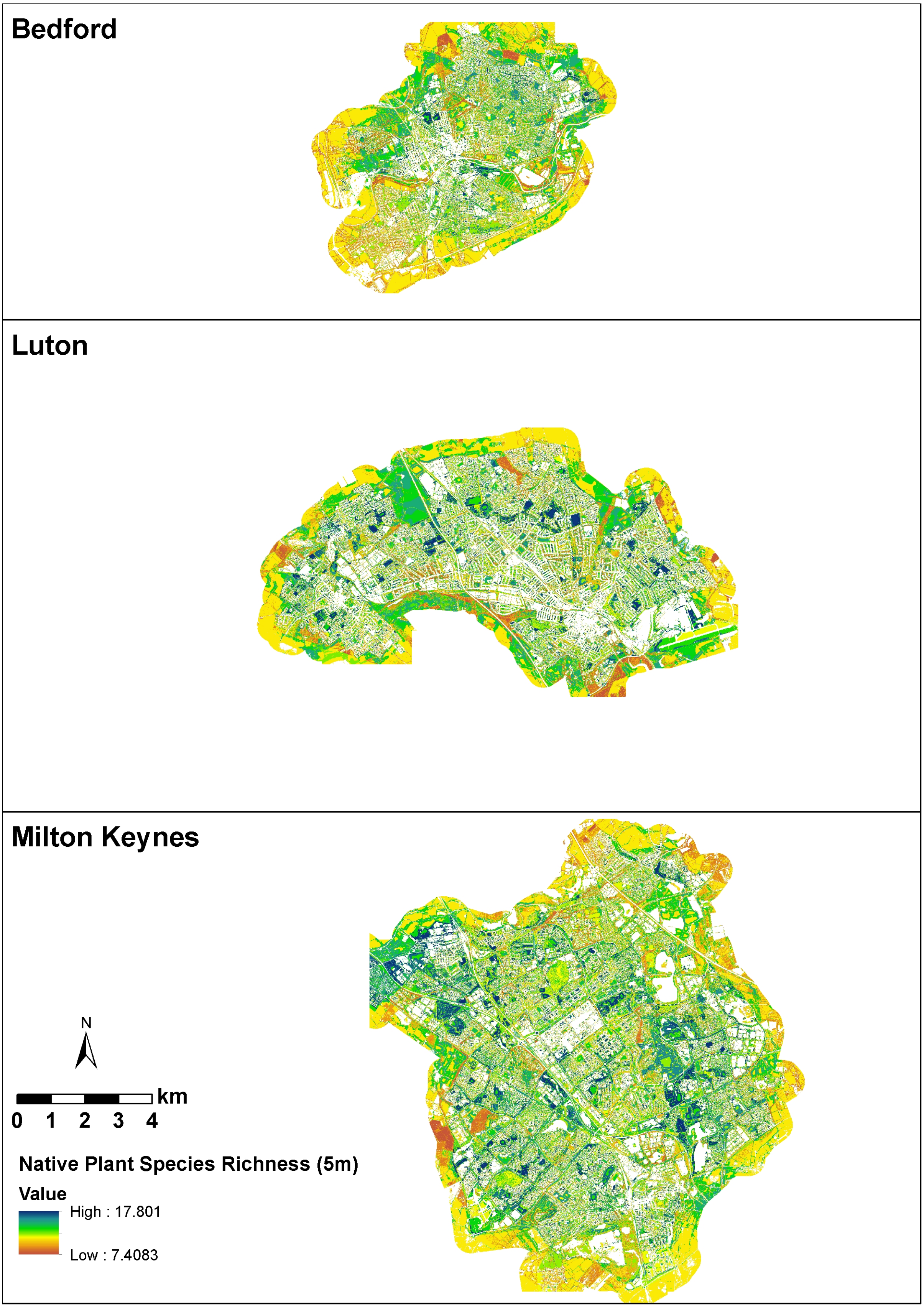

Figure S9: Modelled native non-tree plant species richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

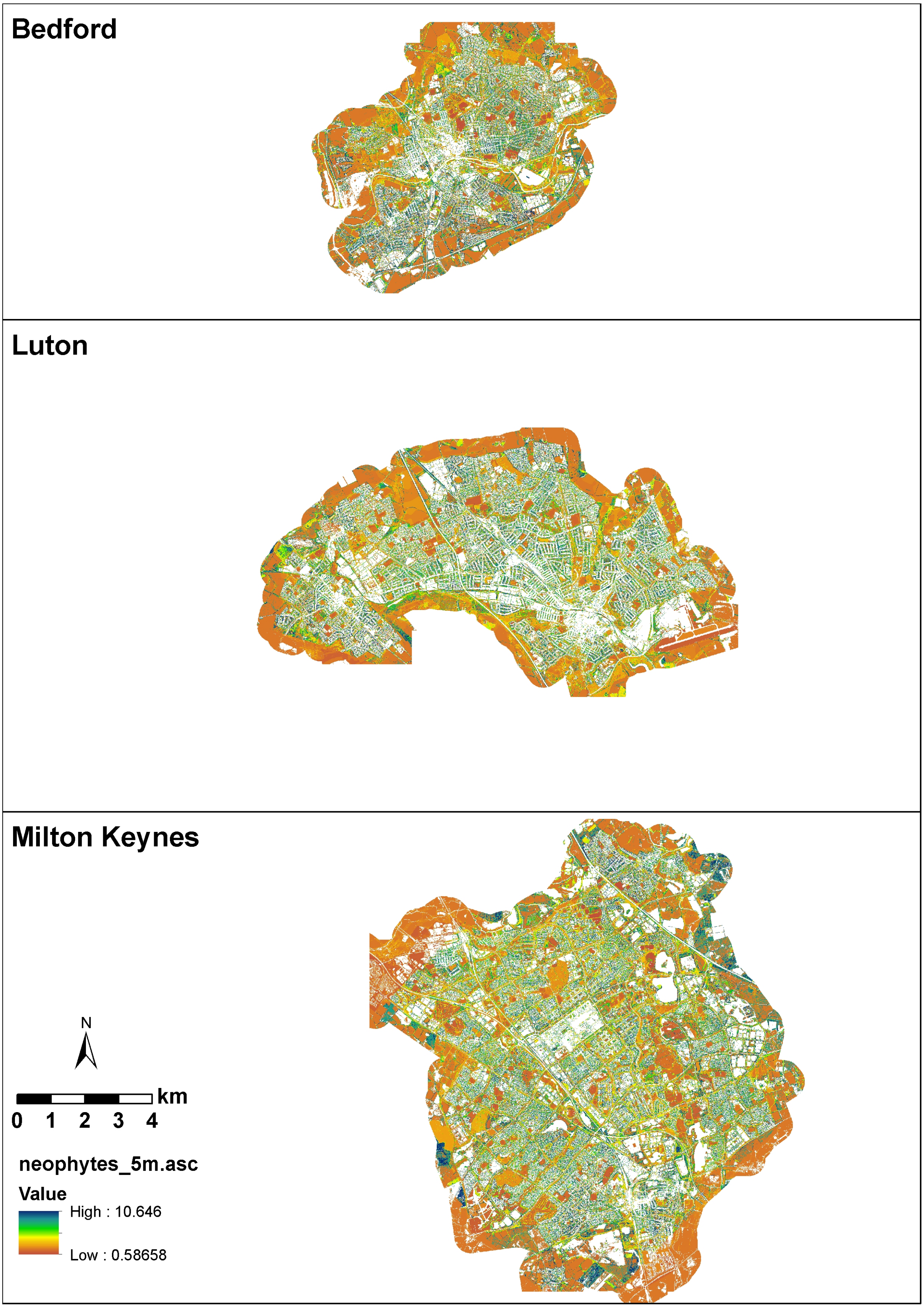

Figure S10: Modelled neophyte non-tree plant species richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.

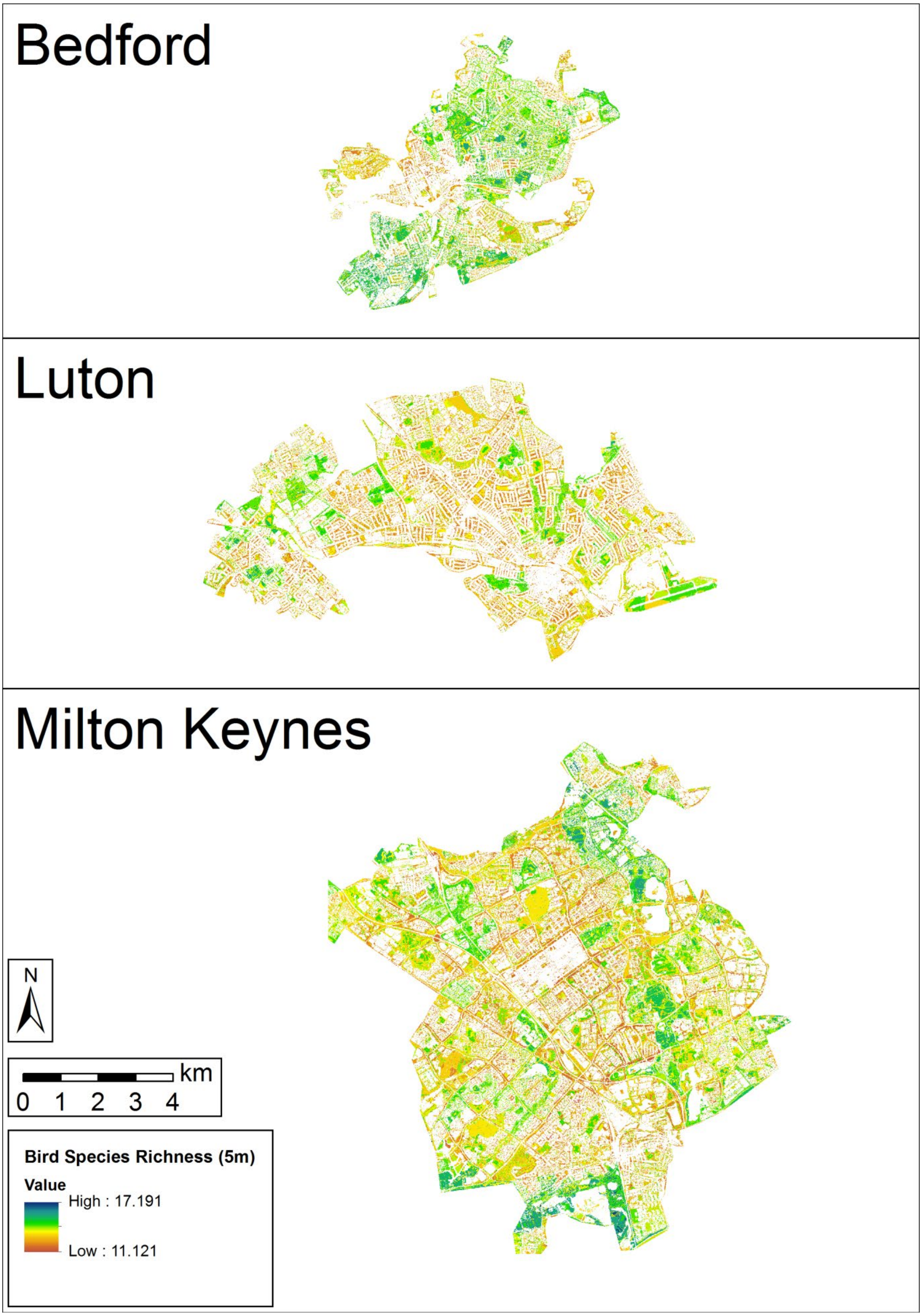

Figure S11: Modelled bird species richness at 5 m resolution for Bedford, Luton and Milton Keynes, UK.