On the Synergistic Use of Optical and SAR Time-Series Satellite Data for Small Mammal Disease Host Mapping

Christopher Marston [1,2,*] **and Patrick Giraudoux** [3]

[1] Department of Geography, Edge Hill University, St. Helens Road, Ormskirk, Lancashire L39 4QP, UK
[2] Centre for Ecology and Hydrology, Library Avenue, Lancaster Environment Centre, Bailrigg, Lancaster LA1 4AP, UK
[3] Department of Chrono-Environment, University of Bourgogne Franche-Comte/CNRS, La Bouloie, 25030 Besançon CEDEX, France; patrick.giraudoux@univ-fcomte.fr
[*] Correspondence: marstonc@edgehill.ac.uk; Tel.: +44-1695-654-396

check for updates

Abstract: (1) Background: *Echinococcus multilocularis* (Em), a highly pathogenic parasitic tapeworm, is responsible for a significant burden of human disease. In this study, optical and time-series Synthetic Aperture Radar (SAR) data is used synergistically to model key land cover characteristics driving the spatial distributions of two small mammal intermediate host species, *Ellobius tancrei* and *Microtus gregalis*, which facilitate Em transmission in a highly endemic area of Kyrgyzstan. (2) Methods: A series of land cover maps are derived from (a) single-date Landsat Operational Land Imager (OLI) imagery, (b) time-series Sentinel-1 SAR data, and (c) Landsat OLI and time-series Sentinel-1 SAR data in combination. Small mammal distributions are analyzed in relation to the surrounding land cover class coverage using random forests, before being applied predictively over broader areas. A comparison of models derived from the three land cover maps are made, assessing their potential for use in cloud-prone areas. (3) Results: Classification accuracies demonstrated the combined OLI-SAR classification to be of highest accuracy, with the single-date OLI and time-series SAR derived classifications of equivalent quality. Random forest analysis identified statistically significant positive relationships between *E. tancrei* density and agricultural land, and between *M. gregalis* density and water and bushes. Predictive application of random forest models identified hotspots of high relative density of *E. tancrei* and *M. gregalis* across the broader study area. (4) Conclusions: This offers valuable information to improve the targeting of limited-resource disease control activities to disrupt disease transmission in this area. Time-series SAR derived land cover maps are shown to be of equivalent quality to those generated from single-date optical imagery, which enables application of these methods in cloud-affected areas where, previously, this was not possible due to the sparsity of cloud-free optical imagery.

Keywords: *Echinococcus multilocularis*; *Ellobius tancrei*; land cover; *Microtus gregalis*; random forests; SAR; Sentinel; spatial epidemiology; time-series

1. Introduction

The parasitic tapeworm *Echinococcus multilocularis* (Em) is a highly effective pathogen of Human Alveolar Echinococcosis (HAE), which is a chronic debilitating disease with >90% mortality in untreated patients 10 years after diagnosis [1–3]. A neglected zoonotic disease, HAE is responsible for a significant burden of human disease across continental Asia, and is expanding in the prevalence and range in Europe, North America, and Asia [4–6]. In China and central Asia, HAE prevalence

in humans can exceed 10% locally [7–9] even though geographical patterns are highly variable, as elsewhere in the world [10]. The reasons for this variability are unknown.

　　The Em transmission cycle is based on predator–prey relationships between definitive hosts (such as the widely distributed red fox, corsac fox, Tibetan fox, and wolf), and small mammal (rodent or lagomorph) intermediate hosts (Figure 1). Tapeworm eggs are shed in definitive host faeces, contaminating the environment, with intermediate hosts infected by oral ingestion of eggs when feeding on vegetation. The transmission cycle is completed when definitive hosts themselves become infected by predating on infected, small mammals. Domestic dogs can also be infected through predation of small mammals, and, due to their close contact with people, are a major transmission source to humans [11] who are infected via accidental egg ingestion. Once infected, HAE is characterized by slow growing larval cysts or infiltrative metacestode lesions in the liver or other organs [12], with an asymptomatic period of several years usually followed by significant morbidity. Treatment is difficult and primarily based on surgery and/or high dose benzimidazole therapy. Disease control measures exist including anti-helminthic dog dosing. However, resources are limited and implementation lacks appropriate geographical targeting, which reduces the effectiveness of these measures.
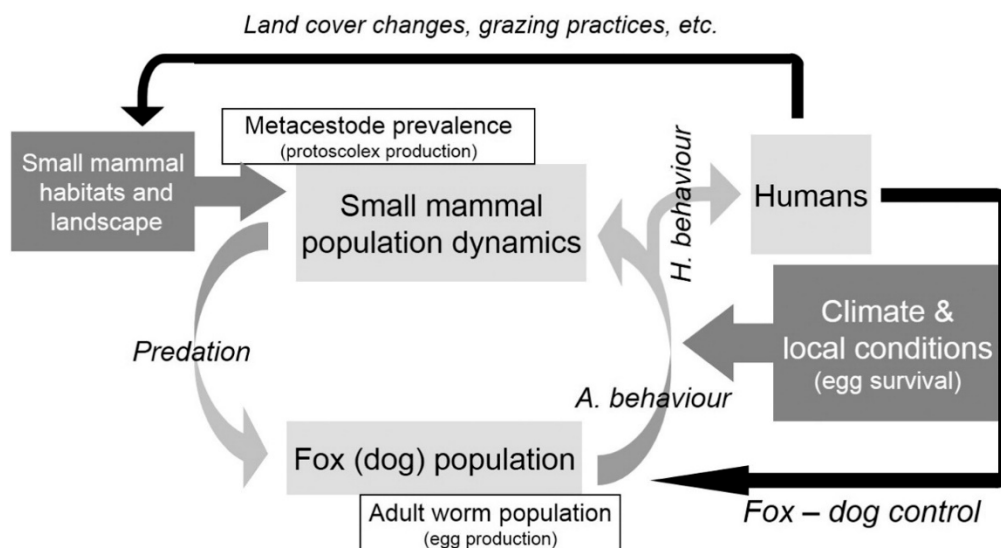

Figure 1. *E. multilocularis* transmission cycle. Adapted from Reference [13]. A. behaviour = animal behaviour. H. behaviour = human behaviour.

　　Domestic dog infection has been shown to be linked to surrounding high densities of small mammals [1]. Therefore, determining the environmental drivers influencing these small mammal distributions is critical to understanding Em transmission dynamics. Small mammal distributions are influenced by the availability of key favourable habitats [14,15], with landscape modification impacting these habitats, so that small mammal presence and transmission varies spatio-temporally [16,17]. Despite its importance, and due in part to its complexity, knowledge is lacking as to the specific landscape mechanisms that are driving these small mammal host distributions.

　　There is currently an urgent need for cost-effective methods to map the key landscape variables, which drive small mammal host distributions, as a proxy for identifying likely areas of active Em transmission. This cannot be achieved using conventional ground-based surveys, which are costly, time consuming, and spatially limited. Satellite remote sensing offers a solution. This enables cost-effective multi-scale and regional-level monitoring, which, alongside in-situ ecological surveys identifying small mammal distributions, can model and extrapolate landscape-small mammal relationships.

Remote sensing has been applied to the study of a number of small-mammal borne diseases including leptospirosis [18], junin virus [19], sin-nombre virus [20], and other hantaviruses [21,22], and has emerged as an important surveillance tool [23]. It has also been applied to studies of Em, which illustrates strong links between landscape composition and HAE prevalence in the South Gansu Province, China [8,24,25]. Links between Em intermediate host distributions and degraded grassland [14] and tree/shrub habitats [26] in China have also been demonstrated, which illustrate that landscape composition is a key spatial determinant of Em transmission [8,13,27].

Previous applications of remote sensing for the study of Em have utilised medium-resolution optical imagery, and, while this has advanced current knowledge of landscape-Em host dynamics, it does have limitations. Cloud cover inhibiting data collection is a persistent problem in many parts of the world, which significantly impacts data availability. Often, this results in limited optical satellite data availability for the study areas and time-periods pertinent to the studies being conducted. Where a single image is available, this can also be problematic. While a single image characterises the landscape at a snap-shot in time, that snap-shot may not necessarily be representative of the broader seasonal phenological variability in the landscape and vegetation characteristics of that location over the course of a year [28]. This can be problematic when investigating the relationships between the landscape and vegetation characteristics, and small mammal distributions for species where behaviour is highly responsive to the vegetation condition (and, consequentially, food and shelter availability) over relatively short time periods. This could result in inappropriate conclusions being drawn on the influences of the landscape on small mammal distributions. For example, in the case of *M. arvalis* and *A. terrestris* in France, absence is typically determined by bare ground in winter as a result of tilling, where populations cannot be sustained due to the lack of vegetation [16,29]. If single-date remote sensing data acquired in the spring or the summer is used to characterise vegetation in these locations, then these areas will, misleadingly, be characterised by cereal crops, which are considered a favourable habitat for these species. The only way to map the permanent presence of favourable habitats and vegetation conditions such as the permanent grassland shown to be key factors for species such as *M. arvalis* and *A. terrestris* [16,29] is to use a time-series of data. The importance of seasonal differences in vegetation condition for small mammal studies has also been illustrated by Reference [28], which determined relationships between *Ochotona* spp. small mammal presence and Enhanced Vegetation Index values to be the strongest in specific periods in the spring and autumn, and weaker in the summer and winter. Similarly, Reference [30] observed that correlations between vegetation index values and deer mouse (a primary host of sin nombre virus) density, and the number of infected deer mice, typically peaked in May or June. This again indicates that biomass levels at certain periods of year are key, and illustrates the importance of incorporating multi-temporal data capturing phenological variability, rather than just single-date imagery, for small mammal-landscape modelling.

Vegetation phenology is typically well defined [31], with the initial leaf emergence typically followed by rapid green-up prior to a stable period of maximum leaf area before senescence, leading to a low-leaf area period. Vegetation in different localities exhibits different and distinctive patterns in this phenological profile [32], with single date imagery alone unable to capture this distinctiveness. Whereas a time-series of optical imagery may typically be the first choice to monitor phenology, this may not be possible in many areas due to persistent cloud cover. In these situations, an alternative approach is to utilise a dense time-series of Synthetic Aperture Radar (SAR) data, using the SAR temporal backscatter profile to characterise vegetation type, via its changing physical structure due to seasonal phenology. Utilising this time-series of SAR imagery in combination with optical imagery where available, enables the generation of targeted land cover maps quantifying the spatial distribution of key land cover types influencing small mammal distributions. Multi-date SAR data has previously been utilised for other land cover mapping activities [33–36]. However, it has not been employed within an epidemiological context.

Optical and Synthetic Aperture Radar (SAR) remote sensing datasets characterise target features in different ways, but deliver complementary information that, if used in combination, generally leads to an increase in land cover mapping accuracy [37]. Multispectral optical remote sensing imagery delivers rich spectral information about the scattered energy (visible and infrared) from the target surface, which enables the discrimination of different land cover classes on the basis of spectral variations in the specific features in question [38]. In contrast, SAR characterises the physical structure of target features, with the differing spatial structure of these features scattering the SAR signal at differing levels of intensity and amplitude. The ground surface texture strongly influences SAR backscatter [39], and the varying characteristics of SAR backscatter can be used to derive information about that target feature. By using SAR-based texture (structural) and optical (spectral) information synergistically to exploit the different physical principles, which they characterise, improved classifications of land cover have been observed [37,40], compared to where either optical or SAR data have been used independently [38]. This has been illustrated for examples such as mapping of winter wheat in China [35], maize, soybeans, and sunflowers in Ukraine [41,42], forestry studies [43–46], and mapping land cover [47–52].

The Sentinel remote sensing satellite series launched as part of the European Space Agency's Copernicus program offers cost-free, broad-area, high temporal frequency coverage. Sentinel-1 offers cloud-penetrating SAR data, which guarantees data collection even in heavily cloud-affected areas. The availability of Sentinel satellite data at 10 m resolution, along with the six-day repeat frequency of Sentinel 1, provide yet more advantages, and enable mapping of key land cover types present in the study areas using the differing phenological characteristics of the vegetation communities and land cover types present [53,54]. By utilising both Sentinel-1 SAR and optical remote sensing data in combination (for example, from Sentinel-2 or Landsat OLI), misclassifications, which commonly occur using single-date imagery due to the spectral similarity of features, should be reduced by using the additional phenological backscatter profile data from the SAR time-series.

In this case, we address gaps in current knowledge of the Em transmission cycle by investigating small mammal-landscape links at spatial and temporal scales not previously achieved. We address the following problems: (1) lack of information of the small mammal distributions, and landscape drivers thereof, of the species involved in the Em transmission cycle, which limits efficient targeting of control measures, (2) cloud-cover inhibiting optical remote sensing-based landscape-small mammal modelling, and (3) inaccuracies resulting from the influence of vegetation phenological dynamics for landscape-small mammal modelling.

We propose to exploit and combine the technological advances of the Sentinel-1 satellites along with novel land cover classification methods to identify landscape drivers determining small mammal distributions, and construct predictive models identifying potential Em transmission foci. Key advancements of this research are, first, to increase our fundamental understanding of the distribution and ecology of the intermediate hosts involved in the Em transmission processes and apply this knowledge for improved disease control targeting, and second, utilise new remote sensing and modelling methodologies for characterisation of small mammal habitats. Specifically, we pose two research questions:

(1)　Can time-series SAR data generate land cover maps of equivalent quality as optical imagery?
(2)　What are the key land cover drivers impacting Em host *Ellobius tancrei* and *Microtus gregalis* distributions in the study area?

2. Materials and Methods

2.1. Study Site

The study area is located close to the town of Sary Mogol, Alay Valley, Kyrgyzstan (39.679°N latitude, 72.883°E longitude) (Figure 2). Located at altitudes between 2900 and 3200 m above sea level on the edge of the Tien Shan and Pamir mountains, the site is grassland dominated with HAE

prevalence rates of 7.1% reported [25]. The Em parasite found in both dogs and voles [55] indicates that a transmission cycle is and has previously been active here.
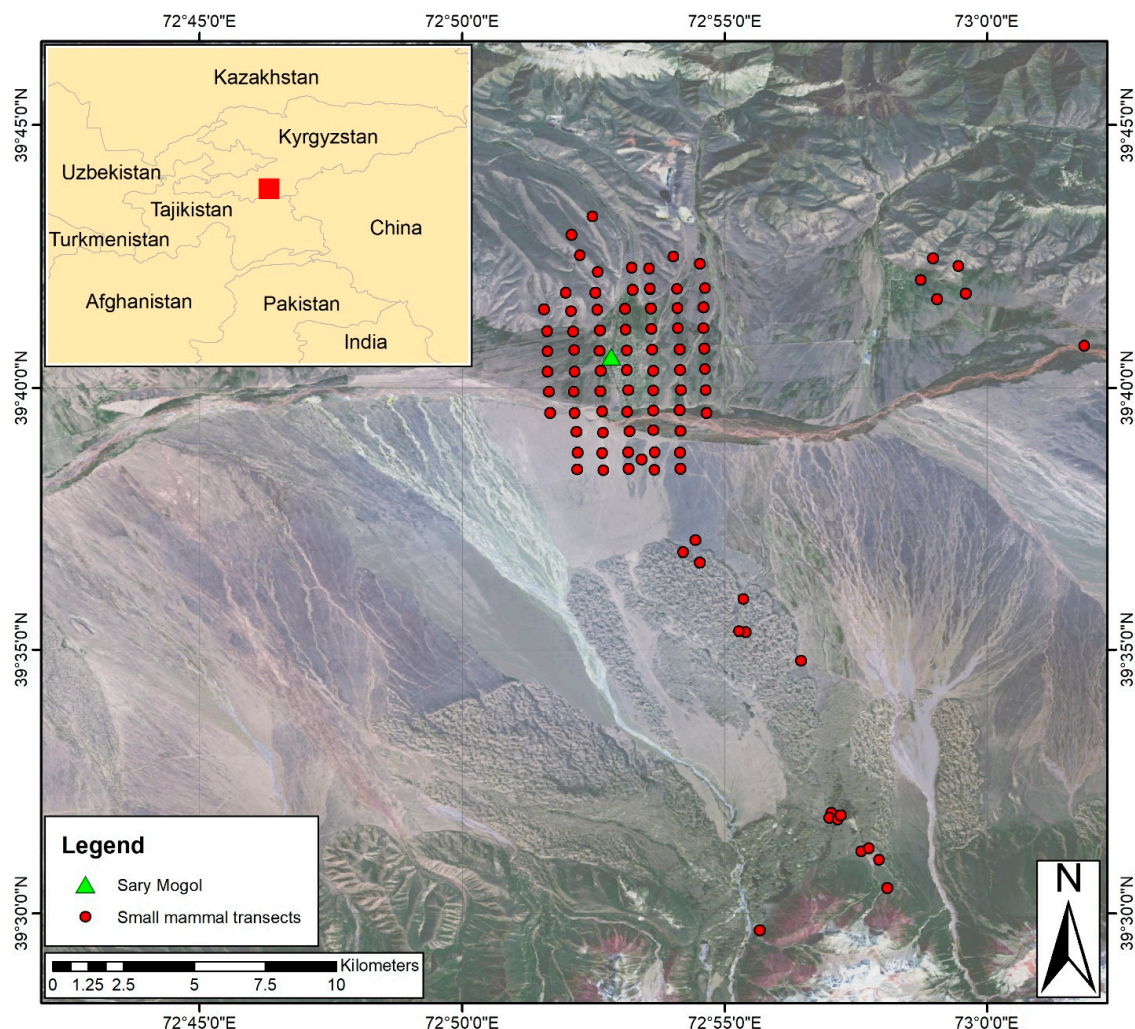

Figure 2. The Sary Mogol study area identifying small mammal survey locations.

Field surveys conducted in May 2012 and September 2014 comprised 95 transects (58 in 2012, 37 in 2014). Transect locations were separated at an average distance of 1.2 km from each other to ensure that spatial autocorrelation was not present [25]. Of these transects, 76 (80%) were completed in the winter pastures and 19 (20%) in more extensive expeditions to summer pastures to specific areas of interest identified in advance of satellite imagery. For each transect, 20 intervals of 10 paces were surveyed with indicators of small mammal activity recorded. Indicators identifiable to species or the genus level including visual sightings of small mammals, foraging corridors, ground holes, and small mammal faeces [1,56] were used as evidence of a small mammal presence using methods established in Reference [57]. Previous studies have performed transect surveys alongside trapping activities and shown good correspondence between activity indices and species with this method established and used widely (e.g., [56–59]). Relative density scores of small mammal presence (the number of intervals where indices of the small mammal presence were observed) were produced for each transect for each small mammal species present [25]. Transect routes were mapped with hand-held Global Positioning System (GPS) receivers with an accuracy of approximately 15 m. The small mammal surveys focused on two species known to be Em intermediate hosts, *E. tancrei* and *M. gregalis,* which are populations that dominate the small mammal community in this area. Previous studies have indicated that, in this region, the population density of *E. tancrei* increases with grassland productivity while *M. gregalis* is

often found along streams and in marshes [25]. They both exhibit low population densities in the early spring and population peaks in the early autumn, but, since they are habitat-specific, their relative distribution between habitats and seasons is the same over multiple years.

2.2. Satellite Data

A time-series of Sentinel-1 SAR data was acquired for a 12-month period (to capture the full annual phenological cycle), comprising 28 dates in total (Table 1), and was downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu). This 12-month period began on 18 October 2014 as this was the earliest date for which Sentinel-1 data was available after the satellite was launched and was as close as possible to the small mammal survey dates. The SAR data was utilised in both ascending and descending orbits, with the VV polarisation Ground Range Detected (GRD) data product from the Interferometric Wide Swath (IW) acquisition mode used. Pre-processing steps for the SAR data were performed, which included range doppler terrain correction (performed using the Shuttle Radar Topography Mission (SRTM) 3Sec digital elevation model), sigma calibration, and de-speckle procedures applied on an individual SAR image basis using a Lee Sigma filter with a 7×7 window size. The data was then sub-set to the study area extent using the SNAP software package. The time series of SAR images were then layer-stacked in order of the imagery acquisition date using ERDAS Imagine to generate a 28-band dataset, which formed the SAR only dataset for analysis.

Table 1. Sentinel-1 SAR and Landsat OLI image acquisition dates and orbital directions.

Number	Satellite	Date	Orbit
1	Sentinel-1A	18 October 2014	Descending
2	Sentinel-1A	24 October 2014	Ascending
3	Sentinel-1A	11 November 2014	Descending
4	Sentinel-1A	17 November 2014	Ascending
5	Sentinel-1A	5 December 2014	Descending
6	Sentinel-1A	11 December 2014	Ascending
7	Sentinel-1A	29 December 2014	Descending
8	Sentinel-1A	4 January 2015	Ascending
9	Sentinel-1A	22 January 2015	Descending
10	Sentinel-1A	15 February 2015	Descending
11	Sentinel-1A	21 February 2015	Ascending
12	Sentinel-1A	27 February 2015	Descending
13	Sentinel-1A	17 March 2015	Ascending
14	Sentinel-1A	23 March 2015	Descending
15	Sentinel-1A	10 April 2015	Ascending
16	Sentinel-1A	4 May 2015	Ascending
17	Sentinel-1A	3 June 2015	Descending
18	Sentinel-1A	21 June 2015	Ascending
19	Sentinel-1A	27 June 2015	Descending
20	Sentinel-1A	15 July 2015	Ascending
21	Sentinel-1A	21 July 2015	Descending
22	Sentinel-1A	8 August 2015	Ascending
23	Sentinel-1A	14 August 2015	Descending
24	Sentinel-1A	1 September 2015	Ascending
25	Sentinel-1A	7 September 2015	Descending
26	Sentinel-1A	25 September 2015	Ascending
27	Sentinel-1A	1 October 2015	Descending
28	Sentinel-1A	19 October 2015	Ascending
29	Landsat OLI	22 July 2014	

The time-period of this study pre-dated Sentinel-2 imagery availability. Therefore, a 30 m resolution Landsat OLI optical image (acquisition date 22 July 2014) was also acquired and was the closest available cloud and snow-cover free image of the study area to the 2014 small mammal survey data collection period. This was acquired from the EarthExplorer data access portal (https://earthexplorer.usgs.gov) and was downloaded as a level-2 atmospherically corrected surface reflectance data product. Landsat OLI bands 2 (blue, 0.452–0.512 µm), 3 (green, 0.533–0.590 µm), 4 (red, 0.636–0.673 µm), 5 (near infrared, 0.851–0.879 µm), 6 (shortwave infrared, 1.566–1.651 µm), and 7 (shortwave infrared, 2.107–2.294 µm) were used to generate the single-date (OLI only) dataset for analysis. The OLI only imagery and time-series SAR only 28-band dataset were layer-stacked together to generate the combined SAROLI (34-band) dataset using ERDAS Imagine. All data was projected to the Universal Transverse Mercator (UTM) WGS84 zone 43N.

2.3. Land Cover Classification

Three land cover classifications were then generated using identical training data to assess their relative abilities to discriminate the land cover classes present using (1) Landsat OLI optical image only, (2) Sentinel-1 SAR time-series data only, (3) Sentinel-1 SAR time-series data, and Landsat OLI optical image in combination. The land cover classifications were performed using a random forest classifier using the R statistical package, and used an eight-class classification nomenclature using the classes: built-up, bare, water, dry grassland, alpine grassland, steppe, bushes, and agriculture (hay and low productivity barley fields). Reference locations of known land cover types were used for training the classification and accuracy assessment, respectively, with field photographs illustrating the different land cover classes within the study area displayed in Figure 3. These data were generated from four sources: (1) field-collected land cover survey locations, (2) reference locations derived from field photos, (3) reference locations derived from very high-resolution Bing aerial and Google Earth imagery, and (4) expert knowledge and direct interpretation of reference locations of clear imagery features (for example, water). The use of higher resolution imagery accessible via public portals such as Google Earth for identification of reference data points is an established technique [38]. For each land cover class, the reference data points were allocated on an alternating basis as training or independent validation locations, which created two equal-sized datasets providing 352 locations for both the training and validation datasets (704 in total) distributed across all land cover classes (built up 7.1%, bare 14.2%, water 10.8%, dry grassland 14.2%, alpine grassland 14.2%, steppe 14.2%, bushes 11.1%, and agriculture 14.2%). Built-up, water and bush classes had lower availability of reference data due to their more restricted coverage in the study area when compared to the other land cover classes. For the training data, training locations were used to generate reference polygons where spectral homogeneity was allowed. Validation data was used to assess the accuracy of the classifications, and to statistically compare the three classifications using McNemar's Test [60]. This is a non-parametric test that assesses, in a statistically rigorous manner, the binary distinction between correct and incorrect class allocations of two classification outputs, and determines whether statistically significant differences in these outputs exist [61].
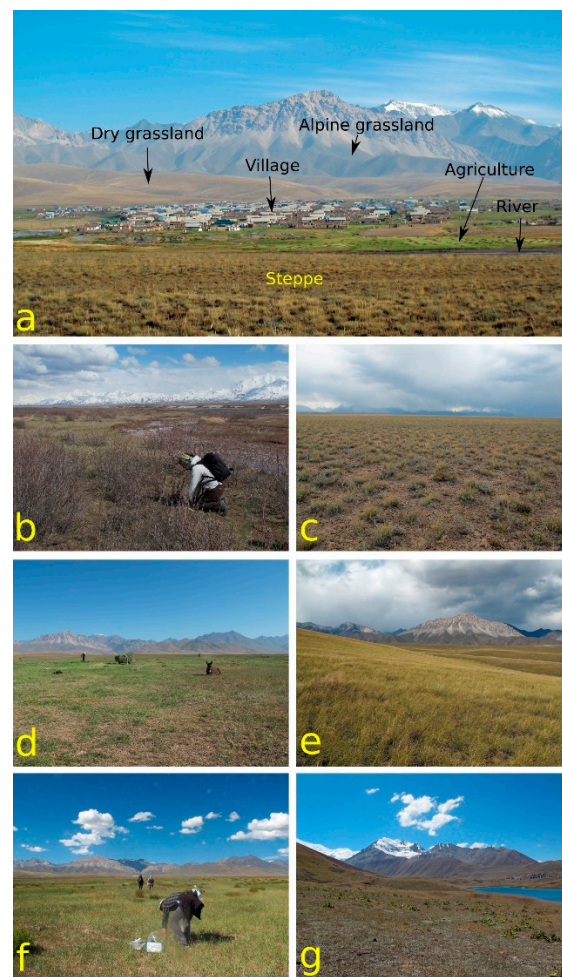
Figure 3. View of the main habitats identified. (**a**) General view of Sary Mogol village and surroundings. (**b**) Bushy areas along the river bed. (**c**) Steppe and (**d**) agriculture (here, hay field), (**e**) dry grassland, (**f**) dry grassland, and (**g**) Alpine grassland.

2.4. Land Cover Data Extraction

Since *E. multilocularis* transmission is largely dependent on small mammal host population distribution and densities through ecological processes linking the landscape to prey/predator relationships [8,27,62–64], determining the appropriate scale at which relevant information on small mammal host population densities can be captured is essential in determining risk areas for Em transmission. Since the home range size of *E. tancrei* and *M. gregalis* is not known *a-priori*, examining the relationships between small mammals and land cover presence over a pre-defined range without a specific basis for doing so could result in misleading results, which do not appropriately identify the relationships present. Instead, the relationships between small mammal relative density scores and surrounding land cover composition is investigated at multiple range sizes, with a series of nested circular buffers centered on the small mammal transect locations created using buffer radii from 50 m to 500 m increasing in 50 m increments. To minimise collinearity between nested land cover area measurements (variables calculated using smaller buffers partly measures the same area as the larger buffers) butretain the nested spatial structure, the buffers were converted to a series of concentric rings with the areas of smaller buffers removed from the larger buffer area within which they are nested. This created a new set of variables Z50 m . . . Z500 m following the methodology of Rhodes et al. [65] such that:

$$Z50\ m = X50\ m \tag{1}$$

$$Z100\ m = X100\ m - X50\ m \tag{2}$$

$$Z150\ m = X150\ m - X100\ m \tag{3}$$

$$Z200\ m = X200\ m - X150\ m \tag{4}$$

$$Z250\ m = X250\ m - X200\ m \tag{5}$$

$$Z300\ m = X300\ m - X250\ m \tag{6}$$

$$Z350\ m = X350\ m - X300\ m \tag{7}$$

$$Z400\ m = X400\ m - X350\ m \tag{8}$$

$$Z450\ m = X450\ m - X400\ m \tag{9}$$

$$Z500\ m = X500\ m - X450\ m \tag{10}$$

where X50 m . . . X500 m are the land cover class coverage data for the 50 m . . . 500 m buffer sizes, respectively, and the Z100 m . . . Z500 m provide the difference between the original variables and the variable nested within it. For each new variable (Z50 m . . . Z500 m), the area of each of the eight land cover classes present was calculated). This produced a total of 80 land cover variables with 10 area measurements each containing 8 land cover classes.

The nested land cover area variables were modelled using random forests in its regression form against the relative density scores for *E. tancrei* and *M. gregalis*, respectively, to identify the relative importance of the area coverage of the different land cover types at the different nested buffer sizes. Random forests (RF) are ideally suited for this analysis, and are an ensemble learning technique developed by Breiman [66] based on a large set of classification and regression trees. They are well-suited to complex, non-linear ecological datasets [67] handling large datasets with correlated predictor variables [68], are non-parametric [69], handle a variety of data types [70], make no assumption of independence concerning the data being analysed [71], and are robust to outliers, noise, and over-fitting [66]. Their use in ecology is becoming more widespread [72], including studies of landscape dynamic influences on parasite hosts [14,28].

Random forest analysis was performed six times, using the land cover maps derived from the combined SAROLI data, the SAR only data, and the OLI data only, for *E. tancrei* and *M. gregalis*, respectively. To achieve the final six random forest models, stepwise removal of the initial 80 nested land cover variables was performed, using the %IncMSE (percentage increase in mean squared error (MSE)) values produced for each variable. Variables with a negative %IncMSE value (i.e., their removal from the random forest would reduce the overall MSE) were removed with the RF then re-run with a reduced set of variables. This process was repeated iteratively, until all variables remaining in the random forest showed positive %IncMSE values (i.e., the MSE of the RF would increase if that variable was removed). No specific splitting rules or pruning methods were defined for the random forest analysis. Random forests also generate variable importance measures identifying the respective influence of the explanatory variables on the response variable (small mammal relative density) and allow the production of partial dependence plots for each variable, which enables further examination of the nature of the relationships present.

It is also possible to examine the statistical significance of the relative importance values of the predictor variables using a permutation-based random forest approach. This works by constructing a large number of random forest models from an identical dataset, and permuting the response variable to obtain the probability distributions of the relative importance measures of the respective predictor variables [73] under the null hypothesis of no relationship between the predictor and response variables. *p*-values for each predictor variable are also generated, which enables assessment of whether the observed relationships are statistically significant. The random forest analysis performed in this case built on methods previously applied in Reference [74] and was performed in R using the 'randomForests' package [75]. 2000 permutations were run for the random forest permutation analysis.

To enable the predictive application of the developed random forests across the study area, a regular point grid with 50 m spacing between points was created across the full extent of the study area. Around each point, nested buffers from 50 m to 500 m in 50 m increments were created, and the proportion of land cover class coverage calculated in an identical manner was conducted for the original survey transect locations. All points where the corresponding 500 m radius buffer extended beyond the classified area were disregarded from further analysis due to their incomplete coverage. This was performed for the SAROLI, SAR only, and OLI only generated land cover maps, respectively. The six random forest models were then applied predictively to generate a predictive relative density values for *E. tancrei* and *M. gregalis* for the land cover maps derived from the SAROLI, SAR-only, and OLI-only data set.

Lastly, to determine which buffer size is optimal for assessing the land cover class–small mammal relative density relationships, Pearson's Product Moment Correlations, p-values, and statistical significance are calculated individually for each land cover class at each nested buffer size for all three land cover classifications for both *E. tancrei* and *M. gregalis*.

3. Results

Three land cover classifications were produced using (1) Landsat OLI single-date optical imagery only, (2) a time-series of Sentinel-1 SAR data only, and (3) a combination of the single-date Landsat OLI image and time-series Sentinel-1 SAR data (Figure 4). Accuracy assessments of the three classifications was performed using identical validation data with classification accuracies of 88.92% (Landsat OLI only), 90.91% (SAR only), and 94.6% (combined Landsat OLI and SAR), respectively. Individual class user's and producer's accuracies are presented in Table 2, with full error matrices presented in Supplementary Information Tables S1–S3.

Table 2. Class-specific user's and producer's accuracy figures for the SAROLI, SAR-only, and OLI-only land cover classifications. UA = user's accuracy, PA = producer's accuracy.

Land Cover Class	SAR and OLI		SAR-Only		OLI-Only	
	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Built up	94.74	72.00	95.24	80.00	72.00	72.00
Bare	85.71	96.00	83.05	98.00	82.35	84.00
Water	100.00	97.37	100.00	76.32	88.10	97.37
Dry grassland	89.29	100.00	85.96	98.00	92.59	100.00
Alpine grassland	98.04	100.00	97.96	96.00	98.00	98.00
Steppe	97.87	92.00	86.27	88.00	97.83	90.00
Bushes	97.22	89.74	97.37	94.87	92.31	61.54
Agriculture	98.00	98.00	91.67	88.00	82.76	96.00

McNemar's test was performed between classification pairs (SAROLI and SAR-only, SAROLI and OLI-only, and SAR-only and OLI-only). Results examining the overall SAROLI and SAR-only classification accuracies produce a *p*-value of 0.037, which indicates that there is a statistically significant difference in classification accuracy between these two classifications at the 95% confidence interval. Likewise, a McNemar's test *p*-value of 0.003 for the SAROLI and OLI-only classifications also identifies a similar statistically significant difference in classification accuracy at the 95% confidence interval (refer to Supplementary Information Table S4). When taken in context with the respective classification accuracy figures, we can say that the SAROLI-based land cover classification offers a statistically significant improvement in classification accuracy over the SAR-only and OLI-only classifications. The McNemar's test results comparing SAR-only and OLI-only derived land cover classifications produced a *p*-value of 0.470, which confirm no statistically significant difference in classification accuracy between these two classifications. Therefore, where both SAR and optical remote sensing data are available, the best results are achieved by using both datasets in combination to generate land cover classification. Where optical remote sensing data is not available due to factors such as

persistent cloud cover, McNemar's test results show that SAR-only based classifications perform as well as single-date Landsat OLI derived classifications, which produce similar classification accuracies (SAR-only = 90.91%, OLI-only = 88.92%). This offers a viable alternative method of land cover characterisation based on vegetation phenology where cloud cover prevents the use of optical imagery.
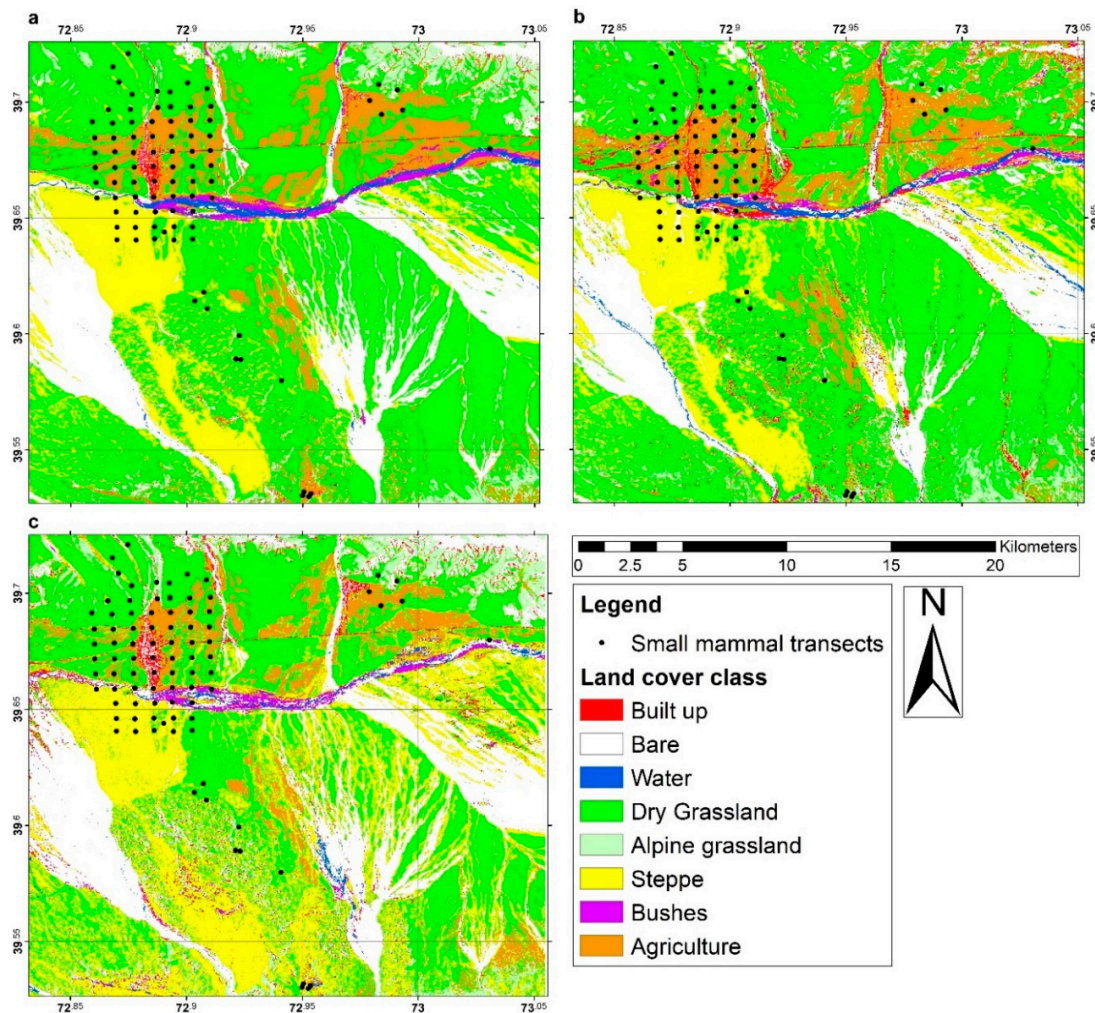

Figure 4. Land cover classification derived from (**a**) SAROLI; (**b**) OLI-only, and; (**c**) SAR-only data.

Random forest analysis was then performed to rank the relative importance values of the presence of the different land cover types, at the different nested buffer sizes, in relation to the relative density scores of *E. tancrei* and *M. gregalis*. The relative density values across all transects for *Ellobius tancrei* are mean = 0.75, median = 0, minimum = 0, maximum = 10, standard deviation = 1.91, and for *Microtus gregalis*, the density values are mean = 1.95, median = 0, minimum = 0, maximum = 19, and standard deviation = 3.27. The random forest analysis for *E. tancrei* consistently shows agriculture for multiple nested buffer sizes to be the most important variable in influencing the relative density scores, with these agriculture values (with the exception of 350 m and 450 m buffers for SAROLI, and 400 m, 450 m, and 500 m for SAR-only) shown to be statistically significant by the random forest permutation analysis (Table 3). Dry grassland (buffer sizes 350 to 500 m) were also ranked among the top 15 most important for SAROLI, built up (450 m, 500 m, 300 m, and 250 m), bare (500 m), and dry grassland (100 m) were ranked among the top 15 for SAR-only, and built-up (400 m, 500 m, 300 m, and 450 m) and alpine grassland (200 m) for OLI-only.

Table 3. Random forest permutation-based variable importance rankings in relation to *E. tancrei*, *p*-values, and statistical significance. For brevity, only the 15 highest ranked variables are presented. ** denotes statistical significance at the 0.01 significance level, * denotes statistical significance at the 0.05 significance level. AG = agriculture, DG = dry grassland, BU = built up, AP = alpine grassland, BA = bare. Random Forest %IncMSE figures are presented in parentheses.

	SAR and OLI		SAR-Only		OLI-Only	
Importance Ranking	**Variable**	***p*-Value**	**Variable**	***p*-Value**	**Variable**	***p*-Value**
1	AG 100 m **	0.001 (16.778)	AG 250 m **	<0.001 (17.924)	AG 150 m **	0.001 (17.187)
2	AG 150 m **	0.005 (14.300)	AG 100 m **	0.005 (15.603)	AG 300 m **	0.001 (16.200)
3	AG 200 m *	0.016 (13.478)	AG 300 m **	0.009 (15.562)	AG 200 m **	0.001 (16.137)
4	AG 300 m *	0.015 (13.308)	AG 350 m *	0.023 (12.557)	AG 250 m **	0.002 (13.392)
5	AG 250 m *	0.039 (11.696)	AG 150 m *	0.021 (11.416)	AG 100 m **	0.002 (13.242)
6	DG 350 m	0.067 (11.657)	BU 450 m *	0.018 (10.104)	AG 50 m **	0.004 (12.807)
7	AG 500 m *	0.042 (11.523)	AG 200 m *	0.041 (9.683)	AG 350 m **	0.006 (12.790)
8	AG 400 m *	0.036 (10.900)	AG 400 m	0.092 (9.239)	AG 450 m **	0.010 (12.771)
9	AG 450 m	0.053 (9.587)	BU 500 m *	0.046 (9.237)	AG 500 m *	0.013 (11.952)
10	DG 300 m	0.155 (9.370)	AG 500 m	0.082 (8.878)	BU 400 m **	0.007 (11.685)
11	AG 50 m *	0.033 (9.325)	AG 450 m	0.173 (8.818)	AG 400 m *	0.025 (11.420)
12	AG 350 m	0.096 (9.307)	BA 500 m	0.179 (7.659)	BU 500 m *	0.031 (10.545)
13	BU 500 m	0.063 (8.912)	BU 300 m	0.113 (7.163)	AP 200 m	0.057 (8.525)
14	DG 400 m	0.317 (8.663)	BU 250 m	0.128 (7.016)	BU 300 m	0.104 (7.551)
15	DG 500 m	0.369 (8.316)	DG 100 m	0.280 (6.938)	BU 450 m	0.123 (7.503)

Partial dependence plots were generated for the statistically significant variables in relation to *E. tancrei* relative density scores, as identified by the random forest permutation analysis to illustrate the nature of the relationships present. For brevity, only the SAROLI partial dependence plots are presented here, with the SAR-only and OLI-only plots displayed in Supplementary Information. For the SAROLI classification, Figure 5 consistently shows that *E. tancrei* relative density scores increase as the area of agricultural land present increases. This clearly illustrates that *E. tancrei* relative density scores are positively related to increasing areas of agricultural land. Similar patterns are observed for the SAR-only derived land cover classification of RF partial dependence plots (Supplementary Information Figure S1), and the OLI-only derived land cover classification partial dependence plots (Supplementary Information Figure S2), which demonstrate that this pattern is consistently observed for the land cover classifications derived using three different input datasets.
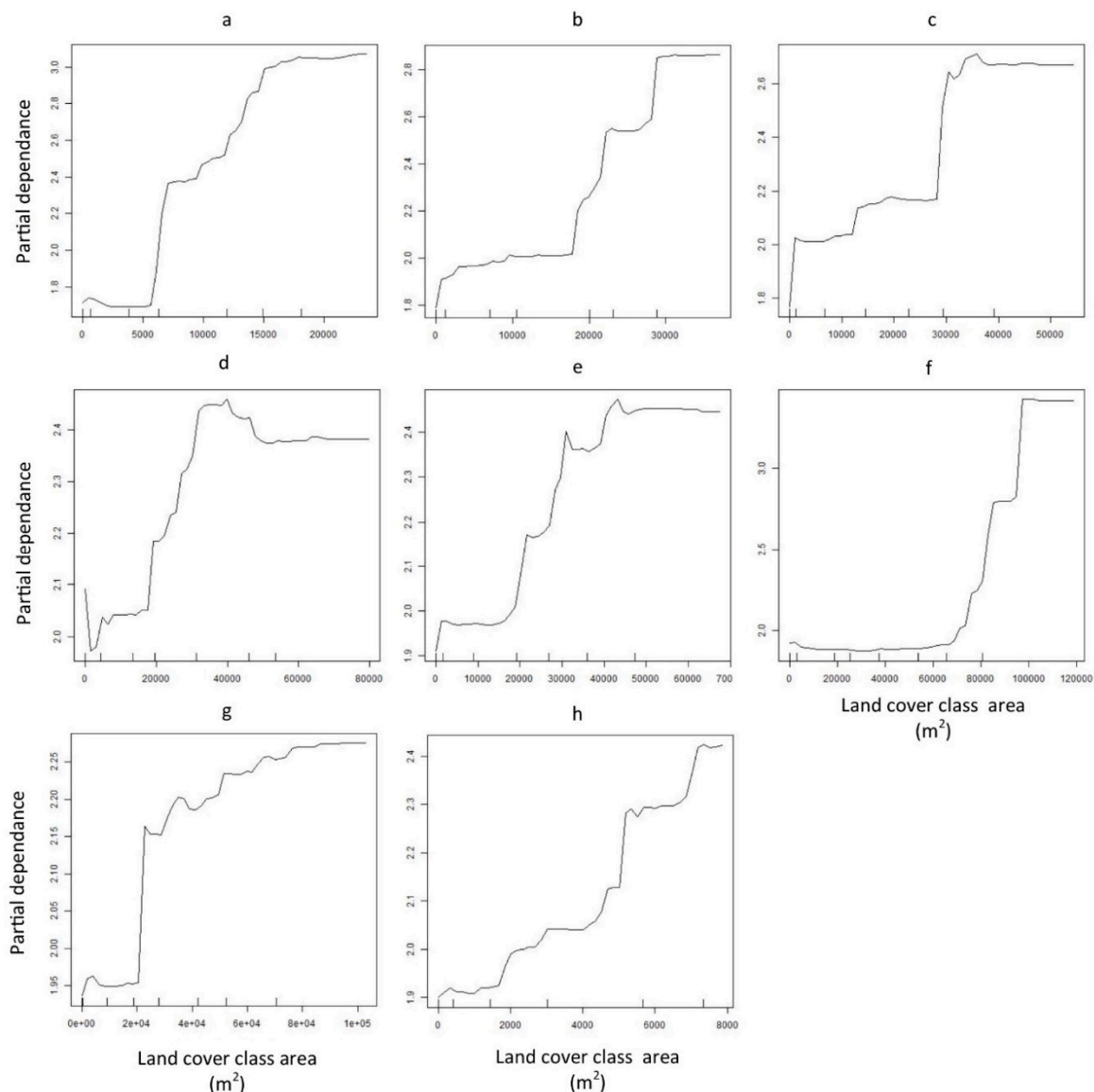
Figure 5. Random forest partial dependence plots for the statistically significant (as shown by random forest permutation analysis) nested land cover variables derived from the SAROLI land cover classification, and *E. tancrei* relative density. Presented in order of variable importance, (**a**) = agriculture 100 m, (**b**) = agriculture 150 m, (**c**) = agriculture 200 m, (**d**) = agriculture 300 m, (**e**) = agriculture 250 m, (**f**) = agriculture 500 m, (**g**) = agriculture 400 m, (**h**) = agriculture 50 m.

For *M. gregalis*, the nested land cover class variables ranked as being of highest importance were more variable (Table 4). However, SAROLI-only bushes (50 m, 100 m, and 300 m) were shown to be statistically significant. For SAR-only, only agriculture at 400 m, water at 200 m, and bushes at 50 m were statistically significant. For OLI-only, only bushes at 300 m was statistically significant. Partial dependence plots were again generated for the variables identified as being statistically significant in the random forest permutation analysis. For the SAROLI-derived land cover classification, the general trend was for increased *M. gregalis* relative density scores as bushes at 100 m and 300 m increased. This is indicative of a preference of *M. gregalis* to inhabit areas with higher levels of bushes present (Figure 6). This pattern was also reflected in the SAR-only derived land cover classification of partial dependence plots, with increasing *M. gregalis* relative density scores associated with increasing bushes at 50 m, and also with increasing water at 200 m (bushes are typically found in close proximity to water at this study site). Conversely, a negative relationship is observed with agriculture, with higher

M. gregalis relative density scores where agriculture is absent, and with these scores rapidly decreasing to low levels with increasing agriculture presence. For the OLI-only derived land cover classification partial dependence plots, only bushes at 300 m were shown to be statistically significant with a similar pattern of increasing *M. gregalis* relative density scores and with an increasing bush presence observed, as was seen with the SAROLI and SAR-only partial dependence plots.

Table 4. Random forest permutation-based variable importance rankings in relation to *M. gregalis*, *p*-values, and statistical significance. For brevity, only the 15 highest ranked variables are presented. ** denotes statistical significance at the 0.01 significance level, * denotes statistical significance at the 0.05 significance level. BS = bushes, AG = agriculture, ST = steppe, WA = water, DG = dry grassland, AP = alpine grassland, BU = built up. Random Forest %IncMSE figures are presented in parentheses.

Importance Ranking	SAR and OLI		SAR-Only		OLI-Only	
	Variable	*p*-Value	Variable	*p*-Value	Variable	*p*-Value
1	BS 50 m **	0.001 (13.673)	AG 450 m **	0.006 (22.610)	BS 300 m **	0.002 (14.649)
2	ST 350 m	0.217 (10.539)	WA 200 m *	0.038 (9.083)	ST 500 m	0.079 (13.446)
3	ST 400 m	0.278 (10.443)	ST 350 m	0.226 (7.801)	AG 500 m	0.201 (10.848)
4	ST 450 m	0.338 (8.961)	ST 300 m	0.238 (6.696)	ST 300 m	0.241 (8.914)
5	BS 100 m **	0.006 (8.761)	ST 500 m	0.295 (6.319)	DG 300 m	0.413 (8.151)
6	BS 300 m *	0.032 (8.452)	ST 450 m	0.355 (5.791)	DG 400 m	0.465 (7.978)
7	ST 500 m	0.416 (8.273)	WA 350 m	0.138 (5.558)	DG 450 m	0.466 (7.947)
8	AP 200 m	0.152 (8.192)	ST 400 m	0.446 (5.544)	ST 250 m	0.278 (7.849)
9	DG 200 m	0.722 (8.074)	AG 500 m	0.402 (5.428)	DG 500 m	0.478 (7.807)
10	BS 400 m	0.103 (7.846)	BS 300 m	0.200 (5.106)	DG 100 m	0.489 (7.744)
11	DG 300 m	0.772 (7.124)	ST 250 m	0.453 (4.761)	DG 350 m	0.636 (6.697)
12	ST 300 m	0.498 (7.123)	BU 500 m	0.257 (4.621)	AG 250 m	0.432 (6.246)
13	DG 50 m	0.549 (6.629)	BS 50 m *	0.035 (4.154)	AG 300 m	0.457 (6.242)
14	BS 350 m	0.121 (6.386)	WA 400 m	0.269 (4.121)	DG 200 m	0.649 (5.294)
15	DG 100 m	0.717 (6.359)	WA 250 m	0.175 (4.035)	AP 500 m	0.419 (4.764)

Predictive application of the random forest models illustrated the predicted relative density of *E. tancrei* and *M. gregalis* across the broader study area. Predicted *E. tancrei* distributions based on SAROLI, SAR-only, and OLI-only random forests show very similar predicted patterns of relative density (Figure 7). Predicted densities vary considerably across the study area with clusters of high predicted relative density shown to occur to the north and center of the study area. These patterns are consistent across all three *E. tancrei* land cover classifications, with high relative density clusters consistent with areas of agriculture presence. This is the case in close proximity to the village of Sary Mogol, which indicates that high densities of *E. tancrei* are likely to be present in close proximity to human settlement. Field surveyed relative density data for *E. tancrei* is overlaid on the predicted densities, which enables a comparison of predicted and observed densities. Although a small number of areas predicted low relative densities of *E. tancrei* where higher densities were observed during the field survey (towards the centre of the study area in Figure 7), the areas predicted to have high relative

densities by the random forests were where higher densities were observed in the field survey data. This is especially the case to the north of the study area.
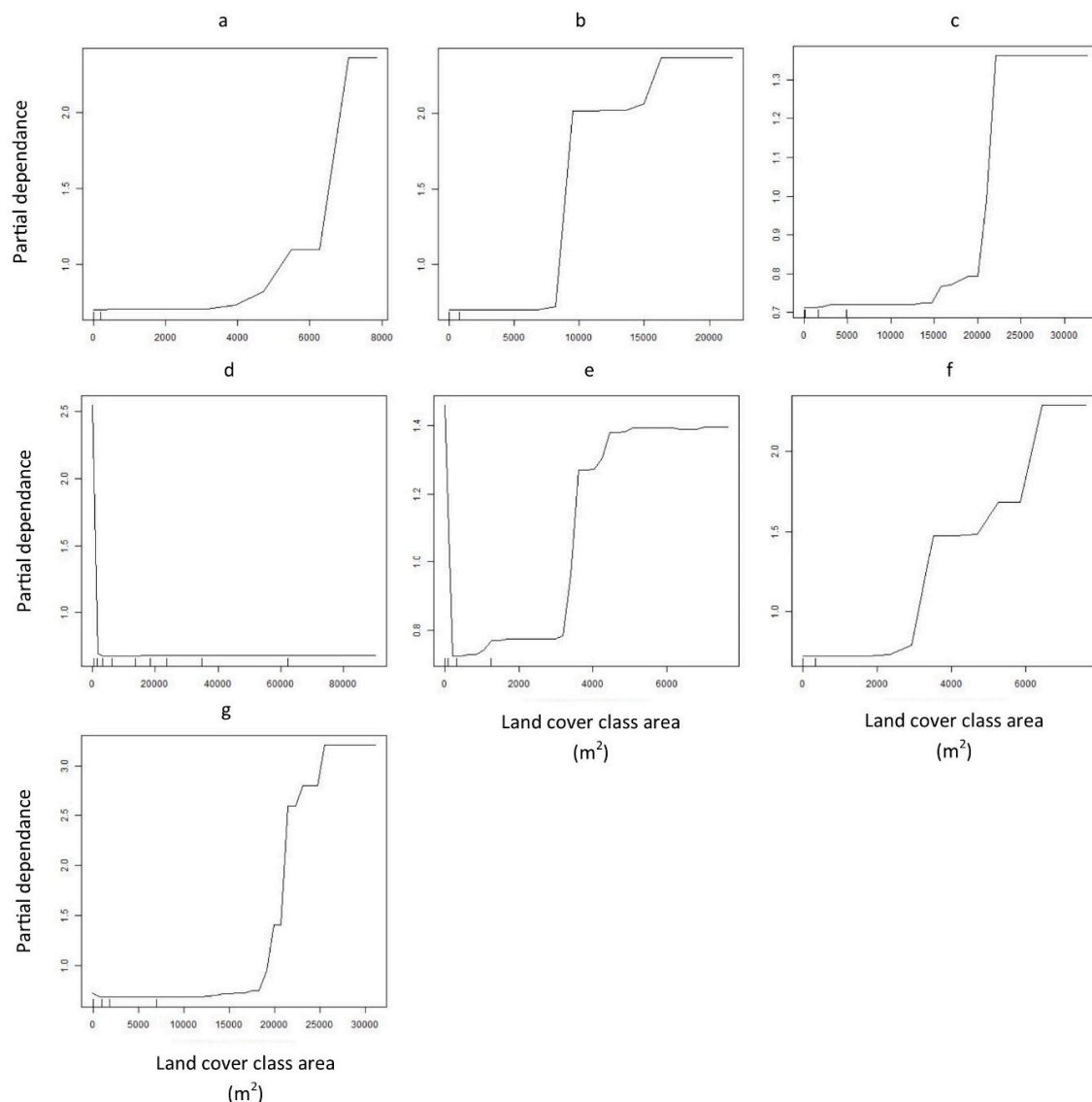

Figure 6. Random forest partial dependence plots for the statistically significant (as shown by random forest permutation analysis) nested land cover variables derived from the three land cover classifications, and *M. gregalis* relative density. Presented in order of variable importance for each classification, for the SAROLI classification, (**a**) = bushes 50 m, (**b**) = bushes 100 m, and (**c**) = bushes 300 m. For the SAR-only classification, (**d**) = agriculture 450 m, (**e**) = water 200 m, and (**f**) = bushes 50 m. For the OLI-only classification, (**g**) = bushes 300 m.

For *M. gregalis*, all three RF predictions consistently identified the same area along a river towards the north of the study area as being of high predicted relative density (Figure 8). This corresponds to the area where the principle area of bushes (shown to be significant by the random forest analysis) was present, and also close to water, which was identified as important in the SAR-only RF analysis. Generally, across the remainder of the study area, predicted *M. gregalis* relative density was low even though the SAR-only RF did predict areas of medium relative density to the west, south, and east of the study area that were not predicted by the SAROLI and SAR-only RF modelling. All three input classifications did, however, identify mostly consistent hot-spots of higher predicted relative density across the study area. When comparing the predicted and observed relative densities for *M. gregalis*

(Figure 8), the patterns were less consistent even though areas along the river where the highest relative densities were predicted also saw the highest observed densities during the field survey. Likewise, the area in the center of the study area was predicted to be of medium-to-high density and had also observed higher densities during the field survey.
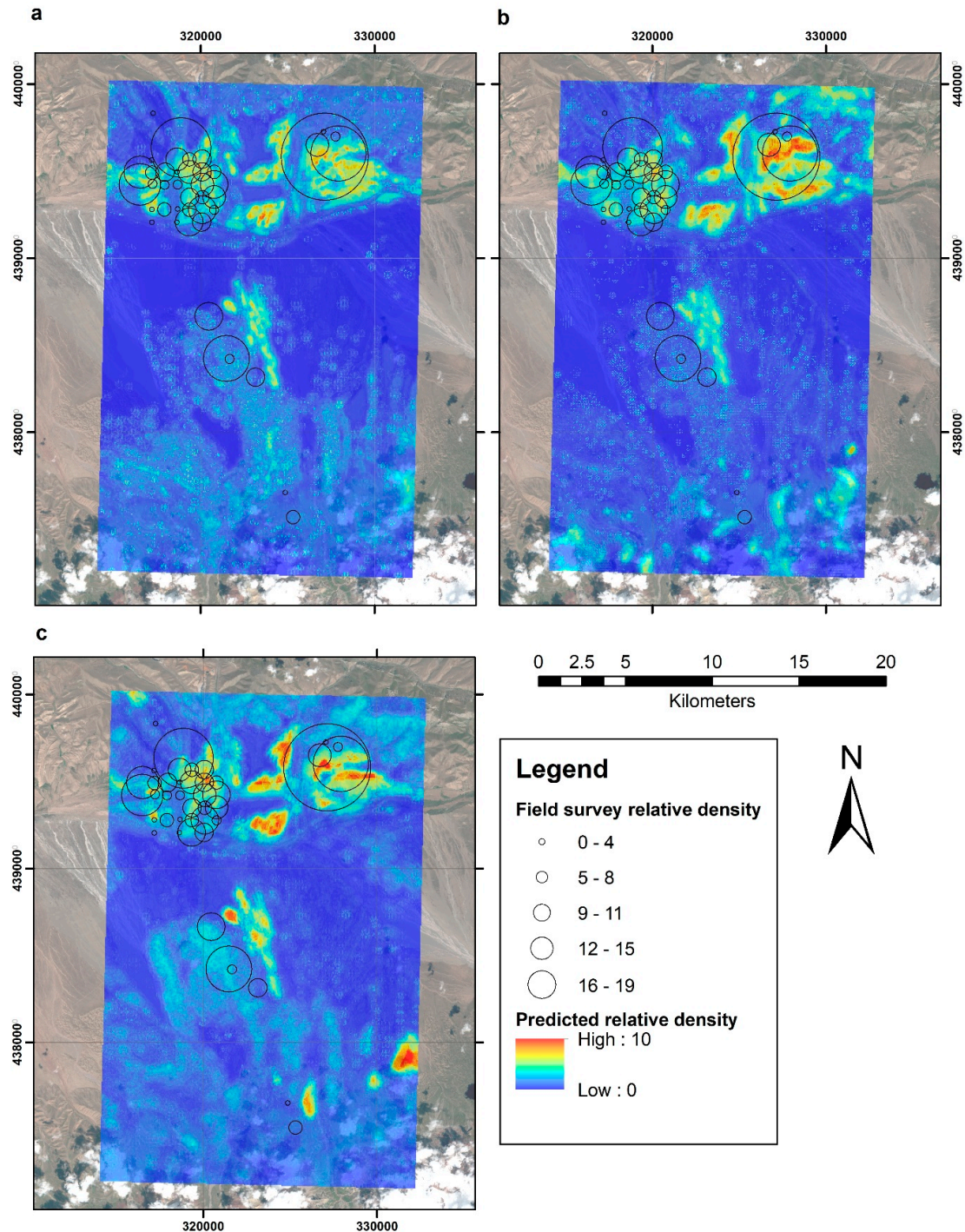

Figure 7. Predicted *E. tancrei* relative density scores based on (**a**) the SAROLI-derived land cover classification. (**b**) OLI-only derived land cover classification, and (**c**) SAR-only land cover classifications and random forest analysis. For context, predicted distributions are overlaid on a Landsat OLI image of the study area with small mammal transect locations displayed. The field survey *E. tancrei* relative density data is overlaid on the predicted distributions for comparison of modelled and observed relative densities.
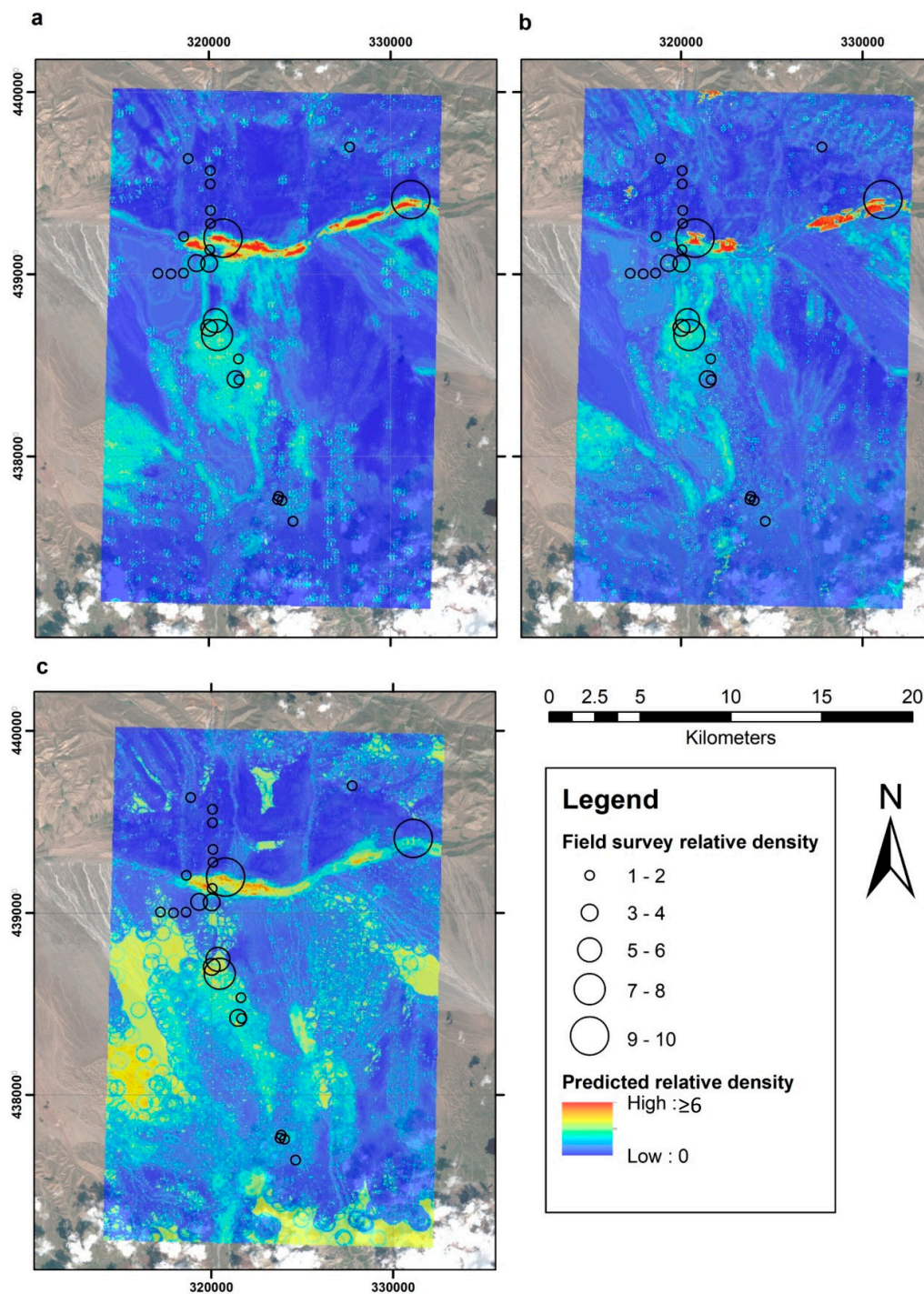
Figure 8. Predicted *M. gregalis* relative density scores based on (**a**) the SAROLI-derived land cover classification, (**b**) OLI-only derived land cover classification, and (**c**) SAR-only land cover classifications and random forest analysis. For context, predicted distributions are overlaid on a Landsat OLI image of the study area with small mammal transect locations displayed. Field survey *M. gregalis* relative density data is overlaid on the predicted distributions for a comparison of modelled and observed relative densities.

To determine the optimal buffer size for assessing the land cover class, small mammal relative density relationships, Pearson's Product Moment Correlations, *p*-values, and statistical significance are calculated individually for each land cover class at each nested buffer size for all three land cover classifications, for both *E. tancrei* and *M. gregalis*. Full results are presented in Supplementary

Information Tables S5–S7. When assessing the statistically significant correlations for *E. tancrei*, agriculture at the 150 m buffer radius size was shown to have the highest correlations with relative density for all three land cover classifications (SAROLI = 0.451, *p*-value ≤ 0.001, SAR-only = 0.479, *p*-value ≤ 0.001, OLI-only = 0.529, *p*-value ≤ 0.001), which indicates that a buffer size of 150 m can be considered optimal for detecting *E. tancrei* correlations. For *M. gregalis*, bushes at the 50 m buffer size showed the highest correlations with relative density for both the SAROLI (0.585, *p*-value ≤ 0.001) and the SAR-only (0.487, *p*-value ≤ 0.001) classifications. Although for the OLI-only classification, the bushes at the 300 m buffer size showed the highest correlation (0.504, *p*-value ≤ 0.001), the 50 m buffer size also exhibited a statistically significant correlation of 0.420 (*p*-value ≤ 0.001). Therefore, generally, a buffer size of 50 m can be considered optimum for assessing the correlations between *M. gregalis* relative density and land cover presence.

4. Discussion

This research examined a critical phase of the *Echinococcus multilocularis* (Em) transmission cycle, and adopted an analytical approach using RF to model and predict *E. tancrei* and *M. gregalis* small mammal presence in relation to landscape characteristics within a highly endemic Em area in Sary Mogol, Alay Valley, Kyrgyzstan. This approach successfully identified the key land cover types driving small mammal distributions and enabled the prediction of small mammal distributions based on land cover maps derived from combined SAR time-series and optical OLI, time-series SAR-only, and single-date optical OLI-only datasets.

Our first question 'Can time-series SAR data generate land cover maps of equivalent quality as optical imagery?' can be answered from our analysis. McNemar's test results showed no statistically significant difference between classification accuracies indicating that the performance of both time-series SAR data (90.91% accuracy) and single-date Landsat OLI imagery (88.92% accuracy) in generating land cover maps using identical training data are similar. However, in this study, the synergistic use of time-series SAR and Landsat OLI data in combination has produced statistically improved land cover classifications over each dataset used independently by using an identical training dataset. This shows that, where data availability allows, the combination of optical imagery (capturing spectral variability) and time-series SAR data (capturing structural variability and its dynamics resulting from seasonal vegetation phenology) is the best option for producing high accuracy land cover maps. Similar results have been observed in previous studies, which found higher classification accuracies when using SAR and optical data in combination [35,37,40,42,44,50] rather than using each dataset individually. These improved classification accuracies are due to the respective capabilities of optical and SAR datasets, which detect different physical properties of land cover features, provide complementary information across the electromagnetic spectrum, and compensate for limitations of using either sensor alone [40]. This is particularly relevant for the agriculture class identified as being of key importance, with its specific SAR backscatter profile exhibited over the growing cycle (based on sowing, growth, and harvesting of crops and ploughing of fields) differing from the typical phenological growth and senescence cycle of other vegetation types such as grassland. This provides additional information on which to classify this class.

Although the three land cover classifications generated are broadly similar, differences did exist in the distributions of the classified land cover types due to the way in which SAR and optical sensors characterise target features. The two key land cover classes of agriculture and bushes did vary in both producer's and user's accuracies for each of the SAROLI, SAR-only, and OLI-only classifications, although accuracy levels were generally high. For agriculture, producer's accuracies (98.0% (SAROLI), 88.0% (SAR-only), and 96% (OLI-only)) and user's accuracies (98.0% (SAROLI), 91.7% (SAR-only), and 82.8% (OLI-only)) showed that this class was classified well using all three input datasets. Likewise, bushes producer's accuracies (89.7% (SAROLI), 94.9% (SAR-only), and 61.5% (OLI-only)) and user's accuracies (97.2% (SAROLI), 97.4% (SAR-only), and 92.3% (OLI-only)) showed that this class was being classified well, with the exception of producer's accuracy for OLI, which performed less well.

Although there were some differences in the classification accuracies and spatial distributions of patches of these land cover classes, the random forest analysis consistently identified these classes of being of the highest importance in relation to *E. tancrei* and *M. gregalis* relative density.

Although there is some variation in the generated land cover maps and derived predicted *E. tancrei* and *M. gregalis* relative density maps, this study illustrates that where optical data is not available due to factors such as persistent cloud cover (a common problem in many parts of the world including Em endemic areas), time-series SAR data is shown to also produce high quality land cover classifications. Whereas, in this example, SAR-only could not achieve as high a classification accuracy as SAROLI, it did achieve classification accuracy similar to that of the single-date OLI data derived classification. This offers new opportunities for land cover mapping in cloud-affected areas, with the ability of SAR data to penetrate cloud offering guaranteed data available from Sentinel-1. This provides an improved potential for characterising phenological variability to strengthen land cover classifications. This variability cannot be captured from single-date imagery, which potentially results in misleading conclusions being drawn on landscape modelling results derived from snap-shot imagery.

Our second question 'What are the key land cover drivers impacting Em host *Ellobius tancrei* and *Microtus gregalis* distributions in the study area?' can again be answered by our analysis. *E. tancrei* distributions are consistently shown by all three land cover classifications to be directly related to agriculture, with increased *E. tancrei* relative density values demonstrated to be related to an increasing presence of this class. This relationship was shown to be statistically significant by the random forest permutation analysis, with agriculture consistently identified as variables of high importance at various nested buffer sizes by the random forest importance plots. When applied predictively across the study area, all three *E. tancrei* predicted models identified the same areas as being of high relative density. The models identified these areas as likely hot-spots for active Em transmission foci due to the denser presence of intermediate hosts (Figure 7). Since the predicted random forest relative density maps all identified similar areas as the highest *E. tancrei* relative density, this further supports the suitability of applying SAR-only datasets for generating land cover classifications for small mammal distribution modelling in this region. This shows that similar results can be achieved by using optical remote sensing imagery, while overcoming the challenges of persistent cloud cover and phenological vegetation dynamics.

For *Microtus gregalis*, the results were more varied. However, increasing the presence of bushes was consistently identified as a key variable resulting in higher *M. gregalis* relative density, which is both statistically significant and ranked as of high importance by the random forest analysis. Water was also identified as statistically significant in influencing *M. gregalis* relative density, which is consistent with the relationship observed with bushes since most areas of higher presence of bushes were found along a river towards the north of the image. This was reflected in the predicted random forest *M. gregalis* relative density map, which consistently identified areas along the river toward the north of the study area as being areas of higher relative density than the remainder of the study area (Figure 8).

This study also determined, based on the highest statistically significant correlations between relative density and the land cover class presence, the optimum buffer sizes for assessing the relationships between *E. tancrei* and *M. gregalis* and the landscape. Results showed that a 150 m buffer size was optimal for assessing *E. tancrei* relative density and land cover presence. A 50 m buffer size is generally optimal for *M. gregalis*. This is important for future research conducted on the spatial distributions of these species, and the design and improvement of sampling strategies for small mammal distribution data collection. As shown in the results, the influence of different land cover types is also dependent on buffer size. Previously, in this region, there was no prior knowledge about small mammal home ranges or the relevant resolutions at which landscape analysis should be performed to capture relevant information for modelling small mammal distributions. Previous articles have supported the idea that small mammal population distributions and dynamics are responsive to landscape configuration, and that these are scale-dependent [17,76]. Identifying the appropriate scales on which to conduct these activities is critical for planning sampling strategies for further studies and

to implement small mammal monitoring since the resolution chosen must be optimal. The multiple nested buffer methods presented here are appropriate for determining the optimal buffer size to investigate small mammal species distribution in a given area. This is a critical point for designing the resolution of risk maps by taking into account small mammal host species distribution and their optimal habitats. However, our results should be considered as preliminary since grassland small mammals can present large inter-annual variations of population density [77]. This might lead to variations in landscape patch occupancy that cannot be detected in short-term studies and calls for further research.

While the results presented illustrate the potential of these methods, it is also necessary to consider the limitations of this approach for small mammal distribution mapping. Although the acquisition dates of the satellite data used was selected to provide a full annual time-series as close as possible to the field survey dates, there was a temporal mismatch between field data collection and imagery acquisition dates. This was unavoidable in this case as the Sentinel-1 satellite was launched in April 2014. However, it is acknowledged that, as some small mammal species experience population cycles, population densities may differ between the field survey and imagery acquisition periods. These small mammals follow "classical" population dynamics with seasonal variations. Due to mortality and no reproduction (bad conditions in winter), a lower density of old animals having overwintered is observed in early spring (April–May), with the population then increasing until September-October due to breeding. Despite these seasonal changes, the distribution of each species is habitat-specific and spill-over to unfavourable habitats is negligible. This means the general pattern of their relative distribution between habitats is constant whatever the season. This is supported by the random forest analysis, which consistently identified the same land cover classes as key in influencing small mammal distributions, although the relative density of the two-species studied may have differed during the intervening period.

Misclassification is also observed within the three land cover classifications and is quantified in the accuracy assessment performed. Although the overall accuracy of the land cover classifications was good at 94.6% (combined Landsat OLI and SAR), 90.91% (SAR-only) and 88.92% (OLI-only), respectively, class accuracies did differ. Agriculture class accuracies did vary between the three land cover classifications with user's accuracy (UA) = 98.00%, producer's accuracy (PA) = 98.00% for SAROLI, UA = 91.67%, PA = 88.00% for SAR-only, UA = 82.76%, and PA = 96.00% for OLI-only classifications. Although generally high, there were differences in classification accuracy and this will, in turn, impact the resultant modelling and predicted small mammal relative density maps, which explains why there are some differences in the three predicted models for each species. Similarly, bushes also showed variability in class accuracy with UA= 97.22% and PA = 89.74% for SAROLI, UA = 97.37%, PA = 94.87% for SAR-only, and UA = 92.31%, and PA = 61.54% for OLI-only. It is likely that it is the misclassification of bushes that produced the inconsistent predicted relative density of *M. gregalis* toward the left of the image in Figure 8c and illustrates that, although the modelling approaches employed here have considerable potential, they are still limited by the quality of the land cover classifications on which the modelling is built. Future work should also investigate the influence of frequency of SAR data acquisition over the annual growing cycle (and how this affects its ability to capture phenological change), and ascending or descending orbital direction to establish their impacts of this classification accuracy.

Additionally, due to the sparsity of historical small mammal survey data it was not possible to validate the predictive distribution maps. This is a current weakness, and although the land cover classifications generated were accuracy assessed, it is important for future application of these methods that validation of the predicted models is also performed including additional data collection if required. However, when the relative density scores from the field survey data were overlaid on the predicted relative densities for visual comparison, the general pattern was that higher observed densities were generally in the areas where the random forest predicted densities were also high although there are some exceptions to this. A comparison can still be drawn between the predictive

models generated from the three land cover maps using identical small mammal distribution data, with the primary outcome of this research being the comparison of the use of optical and radar data to map habitats favourable to the small mammal disease transmission vectors, and to propose an alternative to optical images for regions where cloud cover is persistently high.

5. Conclusions

The results of this study, both in confirming the ability of SAR-only time series remote sensing data to generate land cover classifications of equivalent quality as single-date optical imagery, and the ability to use random forests to identify key small mammal-landscape relationships and predictively apply these models over broader study areas, holds significant potential for Em disease control activities. These techniques, along with the availability of cost-free high temporal frequency and broad geographical coverage of Sentinel-1 SAR data, can be used to identify areas of high relative density of the small mammal species, which act as intermediate hosts in the Em transmission cycle. In turn, this could aid the improved geographical targeting of pre-emptive disease control measures, including targeted treatment of dogs with anti-helminthic drugs to disrupt the Em transmission cycle in that region. Therefore, this would reduce Em infection risk in local human populations. While this study shows that the best results can be achieved by using time-series SAR data and optical imagery in combination, for the first time, we offer a platform to conduct predictive modelling of small mammal distributions based on land cover distributions without the need for optical imagery. This overcomes a major challenge in many Em endemic regions, where persistent cloud cover inhibits the acquisition of good quality optical imagery. By providing new opportunities to incorporate remote sensing as a core element in disease control strategies in Em endemic areas where this was not previously possible, improved geographical targeting of limited disease control resources could contribute strongly to the disruption of Em transmission, and alleviate societal disease burdens in these regions.

References

1. Wang, Q.; Raoul, F.; Budke, C.; Craig, P.S.; Xiao, Y.F.; Vuitton, D.A.; Campos-Ponce, M.; Qiu, D.C.; Pleydell, D.R.J.; Giraudoux, P. Grass height and transmission ecology of *Echinococcus multilocularis* in Tibetan communities, China. *Chin. Med. J.* **2010**, *123*, 61–67. [PubMed]

2.	Cheng, Z.; Zhu, S.; Wang, L.; Liu, F.; Tian, H.; Pengsakul, T.; Wang, Y. Identification and characterisation of Emp53, the homologue of human tumor suppressor p53, from Echinococcus multilocularis: Its role in apoptosis and the oxidative stress response. *Int. J. Parasitol.* **2015**, *45*, 517–526. [CrossRef] [PubMed]

3.	McManus, D.P.; Zhang, W.; Li, J.; Bartley, P.B. Echinococcosis. *Lancet* **2003**, *362*, 1295–1304. [CrossRef]

4.	Torgerson, P.R.; Keller, K.; Magnotta, M.; Ragland, N. The Global Burden of Alveolar Echinococcosis. *PLoS Negl. Trop. Dis.* **2010**, *4*, e722. [CrossRef] [PubMed]

5.	Thompson, A.; Deplazes, P.; Lymbery, A. *Echinococcus and Echinococcosis*; Academic Press: London, UK, 2017.

6.	Massolo, A.; Liccioli, S.; Budke, C.; Klein, C. Echinococcus multilocularis in North America: The great unknown. *Parasite* **2014**, *21*, 73. [CrossRef] [PubMed]

7.	Craig, P.S.; Giraudoux, P.; Shi, D.; Bartholomot, B.; Barnish, G.; Delattre, P.; Quéré, J.P.; Harraga, S.; Bao, G.; Wang, Y.H.; et al. An epidemiological and ecological study of human alveolar echinococcosis transmission in south Gansu, China. *Acta Trop.* **2000**, *77*, 167–177. [CrossRef]

8.	Giraudoux, P.; Raoul, F.; Pleydell, D.R.J.; Li, T.; Han, X.; Qui, J.; Xie, Y.; Wang, H.; Ito, A.; Craig, P.S. Drivers of Echinococcus multilocularis transmission in China: Small mammal diversity, landscape or climate? *PLoS Negl. Trop. Dis.* **2013**, *7*, 1–12. [CrossRef]

9.	Raimkylov, K.M.; Kuttubaev, O.T.; Toigombaeva, V.S. Epidemiological analysis of the distribution of cystic and alveolar echinococcosis in Osh Oblast in the Kyrgyz Republic 2000–2013. *J. Helminthol.* **2015**, *89*, 651–654. [CrossRef]

10.	Said-Ali, Z.; Grenouillet, F.; Knapp, J.; Bresson-Hadni, S.; Vuitton, D.A.; Raoul, F.; Richou, C.; Millon, L.; Giraudoux, P. The FrancEchino Network. Detecting nested clusters of human alveolar echinococcosis. *Parasitology* **2013**, *140*, 1693–1700. [CrossRef]

11.	Wang, Q.; Yu, W.; Shang, J.; Huang, L.; Mastin, A.; Renqingpengcuo, H.Y.; Zhang, G.; He, W.; Giraudoux, P.; Wu, W.; et al. Seasonal pattern of Echinococcus re-infection in owned dogs in Tibetan communities of Sichuan, China and its implications for control. *Infect. Dis. Poverty* **2016**, *5*, 60. [CrossRef]

12.	Nunnari, G.; Pinzone, M.R.; Gruttadauria, S.; Celesia, B.M.; Madeddu, G.; Malaguarnera, G.; Pavone, P.; Cappellani, A.; Cacopardo, B. Hepatic echinococcosis: Clinical and therapeutic aspects. *World J. Gastroenterol.* **2012**, *18*, 1448–1458. [CrossRef] [PubMed]

13.	Giraudoux, P.; Craig, P.S.; Delattre, P.; Bao, G.; Bartholomot, B.; Harraga, S.; Quéré, J.P.; Raoul, F.; Wang, Y.; Shi, D.Z.; et al. Interactions between landscape changes and host communities can regulate Echinococcus multilocularis transmission. *Parasitology* **2003**, *127*, 121–131. [CrossRef]

14.	Marston, C.G.; Danson, F.M.; Armitage, R.P.; Giraudoux, P.; Pleydell, D.R.J.; Wang, Q.; Qiu, J.; Craig, P.S. A random forest approach to describing Echinococcus multilocularis reservoir *Ochotona* spp. presence in relation to landscape characteristics in western China. *Appl. Geogr.* **2014**, *55*, 176–183.

15.	Raoul, F.; Pleydell, D.R.J.; Quéré, J.P.; Vaniscotte, A.; Rieffel, D.; Takahashi, K.; Bernard, N.; Wang, J.L.; Dobigny, T.; Galbreath, K.E.; et al. Small-mammal assemblage response to deforestation and afforestation in central China. *Mammalia* **2008**, *72*, 320–332. [CrossRef]

16.	Giraudoux, P.; Delattre, P.; Habert, M.; Quéré, J.P.; Deblay, S.; Defaut, R.; Duhamel, R.; Moissenet, M.F.; Salvi, D.; Truchetet, D. Population dynamics of fossorial water vole (Arvicola terrestris scherman): A land use and landscape perspective. *Agric. Ecosyst. Environ.* **1997**, *66*, 47–60. [CrossRef]

17.	Lidicker, W.Z. *Landscape Approaches in Mammalian Ecology and Conservation*; University of Minnesota Press: Minneapolis, MN, USA, 1995.

18.	Herbreteau, V.; Demoraes, F.; Khaungaew, W.; Hugot, J.P.; Gonzalez, J.P.; Kittayapong, P.; Souris, M. Use of geographic information system and remote sensing for assessing environment influence on leptospirosis incidence, Phrae province, Thailand. *Int. J. Geoinform.* **2006**, *2*, 43–49.

19.	Porcasi, X.; Calderón, G.; Lamfri, M.; Gardenal, N.; Polop, J.; Sabattini, M.; Scavuzzo, C.M. The use of satellite data in modeling population dynamics and prevalence of infection in the rodent reservoir of Junin virus. *Ecol. Model.* **2005**, *185*, 437–449. [CrossRef]

20.	Boone, J.D.; McGwire, K.C.; Otteson, E.W.; DeBaca, R.S.; Kuhn, E.A.; Villard, P.; Brussard, P.F.; St Jeor, S.C. Remote sensing and geographic information systems: Charting Sin Nombre virus infections in deer mice. *Emerg. Infect. Dis.* **2000**, *6*, 248–258. [CrossRef]

21.	Glass, G.E.; Cheek, J.E.; Patz, J.A.; Shields, T.M.; Doyle, T.J.; Thoroughman, D.A.; Hunt, D.K.; Enscore, R.E.; Gage, K.L.; Irland, C.; et al. Using remotely sensed data to identify areas of risk for hantavirus pulmonary syndrome. *Emerg. Infect. Dis.* **2000**, *63*, 238–247. [CrossRef]

22. Goodin, D.G.; Koch, D.E.; Owen, R.D.; Chu, Y.; Hutchinson, J.S.; Jonsson, C.B. Land cover associated with hantavirus presence in Paraguay. *Glob. Ecol. Biogeogr.* **2006**, *15*, 519–527. [CrossRef]

23. Wayant, N.M.; Maldonado, D.; Rojas de Arias, A.; Cousiño, B.; Goodin, D.G. Correlation between normalized difference vegetation index and malaria in a subtropical rain forest undergoing rapid anthropogenic alteration. *Geospat. Health* **2010**, *4*, 179–190. [CrossRef] [PubMed]

24. Danson, F.M.; Craig, P.S.; Man, W.; Shi, D.Z.; Giraudoux, P. Landscape dynamics and risk modelling of human alveolar echinococcosis. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 359–366. [CrossRef]

25. Giraudoux, P.; Raoul, F.; Afonso, E.; Zaidinov, I.; Yang, Y.; Li, L.; Li, T.; Quere, J.-P.; Feng, X.; Wang, Q.; et al. Transmission ecosystems of *Echinococcus multilocularis* in China and Central Asia. *Parasitology* **2013**, *140*, 1655–1666. [CrossRef] [PubMed]

26. Danson, F.M.; Graham, A.J.; Pleydell, D.R.J.; Campos-Ponce, M.; Giraudoux, P.; Craig, P.S. Multi-scale spatial analysis of human alveolar echinococcosis risk in China. *Parasitology* **2003**, *127*, S133–S141. [CrossRef] [PubMed]

27. Pleydell, D.R.J.; Yang, Y.R.; Danson, F.M.; Raoul, F.; Craig, P.S.; McManus, D.P.; Vuitton, D.A.; Wang, Q.; Giraudoux, P. Landscape composition and spatial prediction of alveolar echinococcosis in Southern Ningxia, China. *PLoS Negl. Trop. Dis.* **2008**, *2*, e287. [CrossRef] [PubMed]

28. Marston, C.G.; Giraudoux, P.; Armitage, R.P.; Danson, F.M.; Reynolds, S.; Wang, Q.; Qiu, J.; Craig, P.S. Vegetation phenology and habitat discrimination: Impacts for *E. multilocularis* transmission host modelling. *Remote Sens Environ.* **2016**, *176*, 320–327. [CrossRef]

29. Delattre, P.; Giraudoux, P.; Baudry, J.; Truchetet, D.; Musard, P.; Toussaint, M.; Stahl, P.; Poule, M.L.; Artois, M.; Damange, J.P.; et al. Land use patterns and types of common vole (Microtus arvalis) population kinetics. *Agric. Ecosyst. Environ.* **1992**, *39*, 153–169. [CrossRef]

30. Cao, L.; Cova, T.J.; Dennison, P.E.; Dearing, M.D. Using MODIS satellite imagery to predict hantavirus risk. *Glob. Ecol. Biogeogr.* **2011**, *20*, 620–629. [CrossRef]

31. Zhang, X.; Friedl, M.A.; Schaaf, C.B.; Strahler, A.H.; Hodges, J.C.F.; Gao, F.; Reed, B.C.; Huete, A. Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* **2003**, *84*, 471–475. [CrossRef]

32. Yu, X.; Zhuang, D.; Chen, H.; Hou, X. Forest classification based on MODIS time series and vegetation phenology. In Proceedings of the International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; pp. 2369–2372.

33. Whelan, T.; Siqueira, P. Time-series classification of Sentinel-1 agricultural data over North Dakota. *Remote Sens. Lett.* **2018**, *9*, 411–420. [CrossRef]

34. Kontgis, C.; Warren, M.S.; Skillman, S.W.; Chartrand, R.; Moody, D.I. Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics. In Proceedings of the 9th International Workshop on the Analysis of Multi Temporal Remote Sensing Images (MultiTemp), Brugge, Belgium, 27–29 June 2017.

35. Zhou, T.; Pan, J.; Zhang, P.; Wei, S.; Han, T. Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors* **2017**, *17*, 1210. [CrossRef] [PubMed]

36. Balzter, H.; Cole, B.; Thiel, C.; Schmullius, C. Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens.* **2015**, *7*, 14876–14898. [CrossRef]

37. Clerici, N.; Calderón, C.A.V.; Posada, J.M. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: A case study in the lower Magdalena region, Colombia. *J. Maps* **2017**, *13*, 718–726. [CrossRef]

38. Ali, M.Z.; Qazi, W.; Aslam, N. A comparative study of ALOS-2 PALSAR and landsat-8 imagery for land cover classification using maximum likelihood classifier. *Egypt. J. Remote Sens. Space Sci.* **2018**, *21*, S29–S35. [CrossRef]

39. Herold, M.; Woodcock, C.; Di Gregorio, A.; Mayaux, P.; Belward, A.; Latham, J.; Schmullius, C.C. A joint initiative for harmonization and validation of land cover datasets. *IEEE Trans. Geosci. Remote* **2006**, *44*, 1719–1727. [CrossRef]

40. De Alban, J.D.T.; Connette, G.M.; Oswald, P.; Webb, E.L. Combined Landsat and L-Band SAR Data Improves Land Cover Classification and Change Detection in Dynamic Tropical Landscapes. *Remote Sens.* **2018**, *10*, 306. [CrossRef]

41. Kussul, N.; Skakun, S.; Shelestov, A.; Kravchenko, O.; Kussul, O. Crop Classification in Ukraine Using Satellite Optical and Sar Images. *Int. J. Inf. Models Anal.* **2013**, *2*, 118–122.

42. Skakun, S.; Kussul, N.; Shelestov, A.Y.; Lavreniuk, M.; Kussul, O. Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine. *IEEE J.-STARS* **2015**, *9*, 1–8. [CrossRef]

43. Kou, W.; Xiao, X.; Dong, J.; Gan, S.; Zhai, D.; Zhang, G.; Qin, Y.; Li, L. Mapping deciduous rubber plantation areas and stand ages with PALSAR and Landsat images. *Remote Sens.* **2015**, *7*, 1048–1073. [CrossRef]

44. Qin, Y.; Xiao, X.; Dong, J.; Zhang, G.; Roy, P.S.; Joshi, P.K.; Gilani, H.; Murthy, M.S.R.; Jin, C.; Wang, J.; et al. Mapping forests in monsoon Asia with ALOS PALSAR 50-m mosaic images and MODIS imagery in 2010. *Sci. Rep.* **2016**, *6*, 570. [CrossRef]

45. Reiche, J.; Souza, C.M.; Hoekman, D.H.; Verbesselt, J.; Persaud, H.; Herold, M. Feature level fusion of multi-temporal ALOS PALSAR and Landsat data for mapping and monitoring of tropical deforestation and forest degradation. *IEEE J.-STARS* **2013**, *6*, 2159–2173. [CrossRef]

46. Reiche, J.; Verbesselt, J.; Hoekman, D.; Herold, M. Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sens Environ.* **2015**, *156*, 276–293. [CrossRef]

47. Erasmi, S.; Twele, A. Regional land over mapping in the humid tropics using combined optical and SAR satellite data—A case study from Central Sulawesi, Indonesia. *Int. J. Remote Sens.* **2009**, *30*, 2465–2478. [CrossRef]

48. Gessner, U.; Machwitz, M.; Esch, T.; Tillack, A.; Naeimi, V.; Kuenzer, C.; Dech, S. Multi-sensor mapping of West African land cover using MODIS, ASAR and TanDEM-X/TerraSAR-X data. *Remote Sens. Environ.* **2015**, *164*, 282–297. [CrossRef]

49. Jhonnerie, R.; Siregar, V.P.; Nababan, B.; Prasetyo, L.B.; Wouthuyzen, S. Random Forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries. *Procedia Environ. Sci.* **2015**, *24*, 215–221. [CrossRef]

50. Torbick, N.; Ledoux, L.; Salas, W.; Zhao, M. Regional mapping of plantation extent using multisensor imagery. *Remote Sens.* **2016**, *8*, 236. [CrossRef]

51. Vaglio Laurin, G.; Liesenberg, V.; Chen, Q.; Guerriero, L.; Del Frate, F.; Bartolini, A.; Coomes, D.; Wilebore, B.; Lindsell, J.; Valentini, R. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 7–16. [CrossRef]

52. Wijaya, A.; Gloaguen, R. Fusion of ALOS PALSAR and Landsat ETM data for land cover classification and biomass modeling using non-linear methods. In Proceedings of the 2009 International Geoscience and Remote Sensing Symposium (IGARSS), Cape Town, South Africa, 12–17 July 2009; Volume 3, pp. 581–584.

53. Rüetschi, M.; Schaepman, M.E.; Small, D. Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland. *Remote Sens.* **2018**, *10*, 55. [CrossRef]

54. Minh, H.L.; Truong, V.; Duong, N.D.; Anh, T.T. Identification of land cover features phenology using multi-temporal Sentinel-1 data: A case study in Hanoi, Vietnam. In Proceedings of the 37th Asian Conference on Remote Sensing, Colombo, Sri Lanka, 17–21 October 2016.

55. Afonso, E.; Knapp, J.; Tete, N.; Umhang, G.; Rieffel, D.; van Kesteren, F.; Ziadinov, I.; Craig, P.S.; Torgerson, P.R.; Giraudoux, P. Echinococcus multilocularis in Kyrgyzstan: Similarity in the Asian EmsB genotypic profiles from village populations of Eastern mole voles (Ellobius tancrei) and dogs in the Alay valley. *J. Helminthol.* **2015**, *89*, 664–670. [CrossRef]

56. Raoul, F.; Quere, J.P.; Rieffel, D.; Bernard, N.; Takahashi, K.; Scheifler, R.; Wang, Q.; Qiu, J.; Yang, W.; Craig, P.S.; et al. Distribution of small mammals in a pastoral landscape of the Tibetan plateau (Western Sichuan, China) and relationship with grazing practices. *Mammalia* **2006**, *42*, 214–225.

57. Giraudoux, P.; Quere, J.P.; Delattre, P.; Bao, G.; Wang, X.; Shi, D.; Vuitton, D.; Craig, P.S. Distribution of small mammals along a deforestation gradient in south Gansu, China. *Acta Theriol.* **1998**, *43*, 349–362. [CrossRef]

58. Delattre, P.; Giraudoux, P.; Damange, J.P.; Quere, J.P. Recherche d'un indicateur de la cinétique démographique des populations du Campagnol des champs (Microtus arvalis). *Rev. Ecol.* **1990**, *45*, 375–384.

59. Giraudoux, P.; Pradier, B.; Delattre, P.; Deblay, S.; Salvi, D.; Defaut, R. Estimation of water vole abundance by using surface indices. *Acta Theriol.* **1995**, *40*, 77–96. [CrossRef]

60. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [CrossRef] [PubMed]

61. Momeni, R.; Aplin, P.; Boyd, D.S. Mapping complex urban land cover from spaceborne imagery: The influence of spatial resolution, spectral band set and classification approach. *Remote Sens.* **2016**, *8*, 88. [CrossRef]

62. Giraudoux, P.; Delattre, P.; Takahashi, K.; Raoul, F.; Quere, J.P.; Craig, P.S.; Vuitton, D. Transmission ecology of Echinococcus multilocularis in wildlife: What can be learned from comparative studies and multiscale approaches? In *Cestode Zoonoses: Echinococcosis and Cysticercosis: An Emergent and Global Problem*, 341st ed.; Craig, P., Pawlowski, Z., Eds.; IOS Press: Amsterdam, The Netherlands, 2002; pp. 251–266.

63. Pleydell, D.R.; Raoul, F.; Tourneux, F.; Danson, F.M.; Graham, A.J.; Craig, P.S.; Giraudoux, P. Modelling the spatial distribution of Echinococcus multilocularis infection in foxes. *Acta Trop.* **2004**, *91*, 253–265. [CrossRef] [PubMed]

64. Liccioli, S.; Giraudoux, P.; Deplazes, P.; Massolo, A. Wilderness in the "city" revisited: Different urbes shape transmission of Echinococcus multilocularis by altering predator and prey communities. *Trends Parasitol.* **2015**, *31*, 297–305. [CrossRef]

65. Rhodes, J.R.; McAlpine, C.A.; Zuur, A.F.; Smith, G.M.; Ieno, E.N. *GLMM Applied on the Spatial Distribution of Koalas in a Fragmented Landscape. Mixed Effects Models and Extensions in Ecology with R (pp. 469e492)*; Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., Eds.; Springer: New York, NY, USA, 2009.

66. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

67. Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef]

68. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random Forests: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Model.* **2003**, *43*, 1947–1958. [CrossRef]

69. Strobl, C.; Boulesteix, A.L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinf.* **2008**, *9*, 307. [CrossRef] [PubMed]

70. Duro, D.C.; Franklin, S.E.; Dube, M.G. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *Int. J. Remote Sens.* **2012**, *33*, 4502–4526. [CrossRef]

71. Perdiguero-Alonso, D.; Montero, F.E.; Kostadinova, A.; Raga, J.A.; Barrett, J. Random forests, a novel approach for discrimination of fish populations using parasites as biological tags. *Int. J. Parasitol.* **2008**, *38*, 1425–1434. [CrossRef] [PubMed]

72. Marston, C.G.; Wilkinson, D.M.; Reynolds, S.C.; Louys, J.; O'Regan, H.J. Water availability is a principle driver of large-scale land cover spatial heterogeneity in sub-Saharan savannahs. *Landsc. Ecol.* **2018**, 1–15. [CrossRef]

73. Ryo, M.; Rillig, M.C. Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere* **2017**, *8*, e01976. [CrossRef]

74. Veldhuis, M.P.; Rozen-Rechels, D.; le Roux, E.; Cromsigt, J.P.G.M.; Berg, M.P.; Olff, H. Determinants of patchiness of woody vegetation in an African savanna. *J. Veg. Sci.* **2016**, *28*, 93–104. [CrossRef]

75. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.

76. Giraudoux, P.; Pleydell, D.; Raoul, F.; Quere, J.P.; Qian, W.; Yang, Y.; Vuitton, D.A.; Qiu, J.M.; Yang, W.; Craig, P.S. Echinococcus multilocularis: Why are multidisciplinary and multiscale approaches essential in infectious disease ecology? *Trop Med. Health* **2007**, *55*, S237–S246. [CrossRef]

77. Krebs, C.J. *Population Fluctuations in Rodents*; The University of Chicago Press: Chicago, IL, USA; London, UK, 2013.