Modelling of Dubas Bug Habitat and Population Density in Oman, Based on Associations with Environmental, Climatological and Human Factors: A Review

Abstract

To understand the distribution and prevalence of Dubas Bug, highly comprehensive and sophisticated information regarding the environmental, climatological, and agricultural practices is required in order to understand the current biogeographical patterns and to predict future distributions. The analytical techniques available in modern spatial analysis packages, such as Geographical Information Systems (GIS), and Remote Sensing (RS) are required to detect, develop and model spatial links and correlations between presence, absence and density of Dubas bug in terms of climatic, environmental and human factors. The main objective of this review is to demonstrate how mapping, modelling suitability, movement, and interact with Dubas bug stress levels, based on spatial information technologies (SITs). SITs are approaches that can be used to visualize and analyse insect infestations and other entomological and agricultural data, describing factors that assist the management and control of Dubas bug infestation. The Humid-Thermal index also can be used to predict the occurrence or non-occurrence of the Dubas Bug. Solar radiation models and other techniques incorporating climate, socio-economic, environmental, and culture factors and Dubas Bug distribution patterns, should be possible to investigate the high-risk area levels on a spatial scale of infestation. This review describes the most important geospatial statistics, spatial interpolation, temporal risk indices and other methods of analysis. An exhaustive search of related literature showed there are few studies linking location-based infestation levels of Dubas bugs with climate, environmental and human culture related variables in the Middle East. This review also highlights the accumulated knowledge and addresses the gaps in this area of research. It makes recommendations for future studies, and gives suggestions for monitoring and surveillance sites and designing local, regional- level integrated pest management policing of palm tree and other affected cultivated crops.

Key words; Spatial Information Technologies, GIS, Remote Sensing, Dubas Bug

1. Introduction

The date palm, *Phoenix dactylifera* L, is generally considered an important economic yield in the Sultanate of Oman (Mani, Handoo, & Livingston, 2005) as well as in most other Arab countries (El-Juhany, 2010). However, a variety of insect pests cause major damage to this crop resulting in plant deaths, reduced yields and production losses (Abdullah, Lorca, & Jansson, 2010; Blumberg, 2008; El-Shafie, 2012; Howard, 2001; Rivera et al., 2015). One such species, the Dubas bug (*Ommatissus Lybicus* Bergevin, Homoptera: Tropiduchidae), has been identified as being of major negative economic importance, quantitatively and qualitatively affecting growth yield palm in the Sultanate of Oman (Al-Yahyai, 2006). Indeed the Dubas bug has been identified as one of the reasons for the decline in date production in Oman (Al-Yahyai & Al-Khanjari, 2008; Al-Zadjali, Abd-Allah, & El-Haidari, 2006). It is also seen as a major pest of date palms in many countries in

38 the Middle East, North Africa and southeast Russia (Klein & Venezian, 1985). It is said to have
39 originated in the Tigris-Euphrates River Valley from where it has spread to other zones in recent
40 decades (Blumberg, 2008).

41 The Dubas bug is a sap feeding insect and both the adults and nymphs suck the sap from the date
42 palm, causing chlorosis, and repeated heavy infestations which cause the weakening and death of
43 date palms (Al-Khatri, 2004; Howard, 2001; Hussain, 1963; Mahmoudi, Sahragard, Pezhman, &
44 Ghadamyari, 2015; A. M. Mokhtar, Nabhani, & Saif, 2010; A Payandeh & Dehghan, 2013). There
45 is also an indirect effect whereby the honeydew produced by this species can lead to the growth of
46 a black sooty mould on the foliage and consequently a reduction in the photosynthetic rates of date
47 palms (Blumberg, 2008; A. Mokhtar & Al-Mjeini, 1999). The female of this species ranges in size
48 from 5-6 mm, while the males range from 3-3.5 mm. Their colour is a yellowish-green and the
49 main distinguishing feature between males and females is the presence of spots on females, while
50 the males have a more tapered abdomen and larger wings relative to the abdomen. The nymphs
51 have 5 instars and each instar has white waxy filaments (Hussain, 1963).

52 Numerous studies have investigated the biology of this species (Hussain, 1963; Klein & Venezian,
53 1985; Shah, Naeem, Nasir, Irfan-ul-Haq, & Hafeez, 2012). Lybicus bug species have two
54 generations annually. The first generation is in the summer with young nymphs appearing in mid
55 to late April. After two months young nymphs mature, and eggs of the second generation appear
56 in late September (Blumberg, 2008; Hussain, 1963). Females lay between 100 and 130 eggs
57 (Elwan & Al-Tamimi, 1999; A. M. Mokhtar et al., 2010). The Dubas female lays its eggs by
58 inserting them into holes in the tissue of the date palm frond at the end of each generation. For
59 about three months the eggs remain dormant and when they hatch the resulting nymphs remain
60 living on the fronds of the same tree feeding on the sap and defecating huge amounts of honeydew
61 which covers the palm fronds and leads to the growth of black sooty mould (A. M. Mokhtar et al.,
62 2010). In extreme cases the mould can block the stomata of the leaves causing partial or complete
63 suffocation of the date palm, which in turn reduces its crop. The honeydew secretion also makes
64 the dates unpalatable (Aminaee, Zare, & Assari, 2010; El Bouhssini, Brownbridge, & Gassouma;
65 El-Juhany, 2010). In summer the eggs can hatch within 18-21 days but in winter may take up to
66 170 days (Blumberg, 2008). The developmental time of Lybicus eggs was studied under three
67 temperatures, 17.6°C, 27.5°C and 32.4°C. The results showed that 27.5°C appeared to be the
68 optimum temperature for the biological activities of this species (Al-Khatri, 2011). At 35°C the
69 biological processes of the bug were disrupted, leading to high mortality, particularly in females
70 (A Payandeh & Dehghan, 2013). Investigations into the population, and spatial distribution
71 fluctuation of two generations of Dubas bug in the Bam region of Iran showed that this pest had
72 an aggregated spatial distribution pattern (Arezoo Payandeh, Kamali, & Fathipour, 2010).
73 Seasonal activity showed that nymphs were dynamic from April to May at first generation and
74 August to October at second generation. In the first and second generations the adults were active
75 from May to June and September to November, respectively (Arezoo Payandeh et al., 2010).
76 Blumberg (2008), reported that temperatures below 0°C adversely affect the ability of adults to

77 survive. The Dubas bug avoids direct sunlight (Klein & Venezian, 1985; Shah, Mohsin, Hafeez,
78 Naeem, & Haq, 2013), and it is spread via transfer of infested offshoots and by wind or wind dust
79 (Blumberg, 2008; Shah et al., 2013).

80 Some research has also been conducted on the natural biological control of the Dubas bug, such
81 as predators and parasites. Early results show a variety of natural predators that could be used as
82 biological control agents, among these being *Aprostocetus sp., Oligosita sp.*and *Runcinia sp.*
83 (Biennial Agricultural Report 2006). Further, Hussain (1963) reported that the eggs of the Dubas
84 bug are parasitized by a small Chalcidoid, while the nymphs and adults were preyed upon by and
85 the larvae of the lace wing (*Chrysopa carnea Steph.*). Hussain also reported three adult species of
86 Coccinellids that prey on nymphs and adults of the Dubas bug. However, further study is needed
87 to determine the distributions of these natural enemies in Oman as well as their effectiveness in
88 controlling Dubas bug populations. Some measure of success was also achieved with pathogenic
89 bacteria as microbiological control agents (Samail Y05), although their toxicological aspects need
90 further research to assess the safety of their implementation on a large scale (Biennial Agricultural
91 Report 2006).

92 Given the significant economic impact of this pest, research into effective management strategies
93 demands high priority. Several insecticides have been evaluated for Dubas bug control in Oman
94 with SUMI-ALPHA® 5 EC being effective as a ground spray and KARATE® 2 ULV, TREBON®
95 30 ULV and SUMICOMBI® 50 ULV achieving some measure of success as aerial sprays (Biennial
96 Agricultural Report 2006). KARATE-ZEON was also found to be very effective as it gave 100%
97 reduction in numbers of Lybicus bug instars and adults, between three and seven days after
98 application. However, the use of this particular pesticide is restricted due to its side effects such as
99 irritation (Annual Report 2008). In Israel systemic carbamates such as aldicarb and butocarboxim
100 have been successful, while in Iraq dichlorvos (DDVP) injected directly into the infected palms
101 were also successful in suppressing the pest population (Blumberg 2008). Nonetheless, this
102 method of control is expensive with negative environmental impacts on non-target species
103 (particularly natural enemies of Dubas bug), as well as human health. Research has shown that
104 some pesticide residues persist in the fruit up to sixty days after application (Al-Samarrie & Akela,
105 2011; Khan, Azam, & Razvi, 2001). In addition, chemical control has achieved limited success
106 and this pest continues to pose a major challenge to Omani agriculture.

107 Areas of future research have arisen from this literature review. There are further opportunities for
108 research into the biology and ecology of this species to gain a thorough understanding of the life
109 cycle of this pest, as well as its distribution. The climatic factors that influence its survival and
110 distribution also merit further study, because such information may be useful in determining
111 current and future potential distributions, particularly in the light of climate change. In a review of
112 the effects of climate change on pest populations, Cammell and Knight (1992) suggested that
113 increases in mean global temperatures, as well as changes in rainfall and wind patterns, could have
114 profound impacts on the population dynamics, abundance and distribution of insect pests of
115 agricultural crops. More recent research has supported this finding (Bale et al., 2002; Cannon,

116 1998; Tobin, Parry, & Aukema, 2014; Worner & Gevrey, 2006). A key issue in ecology and
117 conservation is the mapping of pest species distributions, as this information can be used to devise
118 more effective management strategies for their control. Limited research has been carried out into
119 the current (Al-Zadjali et al., 2006) and potential distribution of the Dubas bug.

120 Spatial analysis techniques, remote sensing and species distribution modelling are useful tools for
121 carrying out such analyses. Such studies require information on the current distribution of Dubas
122 bug, which can be obtained by detailed sampling as well as literature searches and searches on
123 databases such as the Global Biodiversity Information Facility (GBIF). Once sites of known
124 occurrence are identified, the environmental conditions at these sites can be used to make
125 projections to other regions to identify potentially suitable areas that can be colonized by the
126 species of interest. An assessment, as to whether specific areas will continue to be climatically
127 suitable under future climate scenarios, can also be made. Once suitable climatic variables have
128 been identified, other environmental variables such as soil types, slope, aspect, geology and
129 elevation can be incorporated in a Geographical Information System (GIS) to evaluate which
130 variables, or combinations of variables, play a role in enhancing the development, survival and
131 spread of the Dubas bug.

132 Remote Sensing can also be used to study the health of date palms. High spatial resolution imagery
133 such as IKONOS and Quickbird are useful in the early detection of date palms infected with the
134 Dubas bug. Such analyses can be carried out on a broad scale and are less costly and more effective
135 using remotely sensed imagery than extensive field based surveys. Early detection plays a crucial
136 role in the management and control of the infestations and further research into techniques for
137 early detection merits urgent attention.

2. The Use of Remote Sensing for Detecting Insect Infestation, and Extracting Vegetation Damage

Dubas Bug impacts are severe in Oman, Southern Iraq and Southern Iran, while not so severe in Saudi Arabia, Egypt, Tunisia, Libya, United Arab Emirates, Algeria and Morocco. In Oman, the infestation seems more severe in mountain wadi biomes and lees in open plains and coastal areas away from the mountains. Preliminary data collected by Ali Al-Whahabi (personal communication) shows different levels of infestation in different localities in Oman. Hussain (1963) and Mahmoudi et al. (2015) have stated that heavy infestations mostly occur along valleys.

Insect destruction disturbs the growth, affecting the quality and production of the palm fruit over large area. Fronds in the palm are important indicators of tree health and defoliated trees have reduced likelihood of survival as they are prone to attack via insects and diseases. The most common method of mapping the insect infestations has been established in North America. This method consists of sketch mapping based on aerial photography (Eklundh, Johansson, & Solberg, 2009). De Beurs and Townsend (2008) have stated that sketch mapping is prone to subjectivity, is time consuming and is rarely comprehensive countrywide. Other remote sensing techniques are used to detect and control vegetation from satellite at local and global scale. Remote Sensing tools and data have been used to develop early detection techniques for the insect infestations on a broad scale. Satellite images can be applied to map the spatial distribution of insect infestations, and possibly do this on temporal scale as well to indicate the directions and speed of the spread. The data captured, especially the spatial distribution and spread images, can be used as inputs to the GIS-based predictive model. Hyperspectral and multispectral (airborne) data has been used to detect a variety of vegetation or tree problems caused by insects. For examples, D. G. Leckie, Cloney, and Joyce (2005) used multispectral imagery to detect and map crown discoloration caused by jack pin budworm with 2.5 resolutions. D. Leckie, Jay, Gougeon, Sturrock, and Paradine (2004) used multispectral data for detecting and assessing of trees with *Phellinus weirii* (Laminated root rot). Hyperspectral data was used to map the high-risk areas of infestations caused by insects in Malaysia (Shafri & Hamdan, 2009). A benefit of using multispectral information is that it enables the detection of field heterogeneities of crop vigour. However, a limitation of using multispectral data is that multispectral images are not particularly suitable for early detection of crop infestation (Franke & Menz, 2007). For this purpose, hyperspectral data is more suitable and expansive compared with multispectral data (Shafri & Hamdan, 2009) and there is a need to use this data type for identifying palm tree–level pre-visual reductions in chlorophyll contact, essential for early stress detection.

Satellite imagery data also has been used for mapping and control of insect stress levels. The capability for detecting damage from sensors depends on the type or degree of frond damage, which may range from discoloration to total defoliation and tree death. It also depends on the stress level and spatial distribution of the damage. The approach is typically based on differences in the red, near-infrared (NIR), and short- wave infrared (SWIR) wavelengths, and arithmetic incorporation of these bands. Remote sensing is a powerful technique for mapping insect damage at different resolutions. However, the scale at which remote sensing studies

measure attributes of vegetation conditions varies with the particular sensor used. Sensors vary in their spatial, spectral, radiometric and temporal resolutions of data collection, and consequently in their ability to record and monitor attributes of vegetation (Lawley, Lewis, Clarke, & Ostendorf, 2015). Ground resolution varies from 0.5m to 10m for high spatial resolution sensors such as IKONOS and 8 bands WorldView, to 30m for medium resolution sensors from Landsat, and 250-1000 m for low resolution from MODIS and NOAA AVHRR. The size of the ground resolution cells determines the scale of the individual features that can be detected, and each image pixel may contain a combined reflectance from more than one landscape view.

Wulder, Dymond, White, Leckie, and Carroll (2006) used Landsat for detecting mountain pin beetle data damage at the land level with a 75% overall accuracy. Landsat images of insect damage can be used to provide a historical record and to evaluate past forest disturbances at medium resolution (Williams & Birdsey, 2003). However, to develop nationwide systems for monitoring and early detection of insect destruction, it is necessary to study the performance of high temporal resolution systems such as NOAA (AVHRR), SPOT vegetation, Envisat (MERIS) and EOS (MODIS). These satellites have limited spatial resolution but have been used to serve as real methods for viewing extensive spread damage and providing early symptoms of attacks over large areas. The capacity of high temporal resolution remotely sensed data may be applied to monitor the seasonal vegetation variations. For example, Goward, Tucker, and Dye (1985) used NOVAA AVHRR to study North America vegetation patterns. Fraser, Abuelgasim, and Latifovic (2005) used SPOT vegetation at (1 km) resolution for mapping defoliation and mortality of coniferous forests due to eastern hemlock looper. The most common index of vegetation from MODIS is the normalised different vegetation index (NDVI). NDVI, which calculates the association between near infrared (NIR) and red (R), can be used to measure photosynthetic activity of vegetation and is computed by:

$$NDVI = \frac{NIR - R}{NIR + R} \qquad (1)$$

where a result close to 0 denotes no vegetation, while a result close to 1 illustrates the highest possible density of healthy vegetation. The effectiveness of photographic output is influenced by several climatological and environmental conditions such as humidity, temperature, rainfall, geology, soil types, elevation and soil contamination. All these factors can account for observed differences in vegetation heath, density, and evaporation rate. NDVI was used by Ji, Xie, Li, Li, and Zhang (2004) to classify locusts in the forest outback in cultivation lands. However, effective real data in terms of vegetation is not only obtained from visible and near-infrared channels, but also from thermal channels. Essential levels of detail on the extent and severity of the Dubas Bug stress level are required for a variety of farms, in terms of development, management, and modelling activities. To provide the necessary level of detail, data variations needs require appropriate survey methods. Moreover, the timing of surveys must be coordinated relative to outbreak of attack in tree-crown foliage, which can vary substantially. High resolution RS imagery (IKONOS or Quickbird) of agriculture may be used to provide resources management with important data regarding insect infestations. Images taken over several years can be used to study the spatial dynamics of diversity of farms pests, particularly

those which have special relevance to farm canopies. Indeed, several attempts at detecting damage caused by insects have proven successful in research that utilizes high resolution remote sensing data and the analytical tools of GIS for estimating the susceptibility of farms to various insects, but this is not extensive (Dennison & Roberts, 2003; Gimbarzevsky, Dawson, & Van Sickle, 1992).

3. Use of Spatial Interpolation Analysis, Time Risk Indices, and Predictor Methods on Insects Infestations.

3.1 Interpolation Analyst

To provide data for populations of insect pests at given sites, many processes are required such as data collection, using traps, longitude, latitude and the subsequent analysis (S. Kumar, Graham, West, & Evangelista, 2014). However, the spatial interpolation method can be used to predict unknown values within the site by using geo-referenced known point locations and the associated table population data (Beckler, Wade French, & Chandler, 2005). However, the number and distribution of some point can affect the accuracy of spatial interpolation. To be effective for estimation, control points should be well distributed within the study area. Nonetheless, this ideal situation is rare in real-world applications because a study area frequently contains some data-poor areas. This review illustrates the types of spatial interpolation and which elements can be used to attain a valid result.

3.1.1 Density Estimation

Simple and kernel density estimation methods have been used to measure cell densities in a raster by using a sample of known points. The simple method is a counting method, while the kernel method is based on probability functions and offers options in terms of how density estimation is made. Moreover, to use the simple density method is necessary to place a raster on a point distribution, tabulate points that fall within each cell, total the point value, and estimate the cell's density by dividing the total point value by the cell size. The kernel density estimation associates each known point with a kernel function for the purpose of estimation (Bailey & Gatrell, 1995; Scott, 2015; Silverman, 1986). A kernel density estimation method looks like a "bump" centring at a known point and tapering off to 0 over a defined window area (Silverman, 1986). The kernel density estimator at point x is then the sum of bumps placed at the known point x_i within the bandwidth:

$$\hat{f}(x) = \frac{1}{nh^d} k(\frac{1}{n}(x - x_i)) \qquad (2)$$

where $K(\ )$ is the kernel function, h is the bandwidth, n is the number of known points within the bandwidth, and d is the data dimensionality. The kernel density estimation method (KDE) typically produces a smoother surface than the simple density estimation method. KDE has been applied to a wide variety of fields such as forest resources (Lundquist & Beatty, 1999), natural hazards (Van Der Veen & Logtmeijer, 2005). Carvalho, Faro, and Subramanyam (2013) used density estimations to study insect population distributions in a large rice mill in Portugal. Through spatial interpolation, an estimation of the precipitation values at a location

can be made with no recorded data, by using known precipitation readings at nearby weather stations. Thus, spatial interpolation works with rasters for creating surface data that can be used for analysing and modelling. The density surface method can be used to identify Dubas bugs infestations and damaged palm tree hot spots and cold spots, direction (e.g., north Oman). Therefore, KDE has been applied to develop GIS layers such as density and distribution of Dubas bug and it may be possible to link these layers with environmental, meteorological and human factors in more advanced analyses.

3.1.2 Inverse Distance Weighted (IDW), Radial Basis Function (RBF) and Kriging

ArcGIS® ESRI Software can be used to import the geo-references of insect infestation locations, capture data, and spatial distribution analysis. For example, the Average Nearest Neighbour analysis tool in ArcGIS software can be used to assess if the pattern of distribution of insect infestation locations at each site could potentially be clustered since they were distributed in an irregular fashion. Inverse Distance Weighted (IDW), Radial Basis Function (RBF), and Kriging interpolation methods have been used to produce contour maps of stored – product pests. IDW is an exact method that imposes the condition that the estimated values of a point are influenced more by nearby known points by those at a greater distance. The most important characteristic of IDW method is that all predicted values are within the range of maximum and minimum values of the known points. The general equation for the IDW method is:

$$Z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \tag{3}$$

where z_0 is estimate value at point 0, z_i is the z value at known point i, d_i is the distance between point 0, s is the number of known points used in estimation, and k is the specified points. The five RBF methods are: thin–plate spline, spline with tension, completely regularized spline, multiquadric function, and inverse multiquadric function. Each RBF interpolation has a parameter that controls the smoothness of the generated surface. The difference between the surfaces is usually small, while each combination of the RDF method and a parameter value can create a new surface. RDF is only considered informative when working with large datasets and attributes of little variation. Splines have been used for interpolating mean rainfall surface (Hutchinson, 1995), and land demand surface (Wickham, O'neill, & Jones, 2000).

Kriging varies from other local interpolation methods, because it can be used to evaluate the quality of prediction and estimated prediction errors. In GIS environmental Kriging has become a popular method adopted in a wide variety of disciplines (e, g. DEM). The semivariance method has been used in kriging method to measure the spatially correlated component, a component that is called spatial autocorrelation or spatial dependence. It can also be used to estimate the degree of variance between multiple pairs of measurements and provide information on the scale and pattern of the spatial variance (Curran, 1988; Journel & Huijbregts, 1978). According to Krige (1976) the most common form of Kriging is Ordinary

kriging. It is a spatial modelling technique that provides optimal and unbiased estimates of unknown values from sample data (Curran & Atkinson, 1998). The Semivariogram is computed by:

$$y(h) = \frac{1}{2}[z(x_i) - z(x_j)]^2 \tag{4}$$

where *y(h)* is the semivariance between known points, x_i and x_j, separated by the distance *h*; and *z* is attribute data. There is also method call a semivariogram cloud of kriging interpolation which can be used for investigating the spatial variability of the phenomena under study (Gringarten & Deutsch, 2001). However, a semivariogram cloud is too difficult to use and manage because it has pairs of known points. In this case, the method called binning is typically used in kriging to average semivariance data by distance and direction. Geostatistics interpolation is widely used in spatial association of insects by home and overseas academics (Kemp, 1987; Liebhold, Halverson, & Elmes, 1990; Park & Obrycki, 2004; Pearce & Zalucki, 2006; Z. Wang, Li, Shang, & Cheng, 2001; Zhou, Wang, Zhao, & Zhang, 2012). Some studies have used geostatistics interpolation with kriging to analyse the spatial pattern of insects. Indeed this method can analyse the spatial structural dynamics of insect habitat under different environmental conditions (R. Zhang, Leng, Zhu, & Wang, 2007). For example, Zhou et al. (2012) was used Semivariogram kriging to analyse spatial patterns of grasshopper in vegetation in China. All these examples provide valuable methods for ecological, entomological and geographical studies in the monitoring and management of Dubas bug.

4. Using Spatial Statistics for investigating the Distribution of Insect Infestations

4.1. Spatial Statistics Methods

4.1.1 Analysis Pattern

4.1.1.1 Moran's I Autocorrelation

Some types of spatial autocorrelation analysis can be used to describe these mapped populations in spatial terms. The procedure determines the spatial distribution or dispersion pattern (i.e. random, clusters and clumped) of the subject by measuring the relationship population density (attributes of objectives) (i.e. infestations sites) and spatial factors with distance between objectives (Griffith, 1987). The most common measure of spatial autocorrelation is Moran's I. Moran's I spatial autocorrelation can be computed by:

$$1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(x_i - x^-)(x - x^-)}{s^2 \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}} \tag{5}$$

where x_i is a value at point *i*, *x* is the value at point neighbour *j*, w_{ij} is a coefficient, n is the number of points, and s^2 is the variance of *x* values with a mean of. It can be used to measure the spatial autocorrelation of insect infestations based simultaneously on the infestation

locations and values,. The values Moran's I tack on are anchored at the expected value *E(I)* for a random pattern: *E(I)* approaches when the number of points n is large. Moran's I is close to *E(I)* if the pattern is random. It is greater than *E(I)* if adjacent points tend to have similar values. Moran's method is similar to nearest neighbour analysis and can be used to compute the Z score associated with Moran's I. The Z score shows the likelihood that the point pattern could be the result of random chance. Moran's I can also be applied to polygons and remains the same for computing the index value, but the coefficient is the based on relationship between polygons. However, recent developments in spatial statistics have included the Local Indicator of Spatial Association (LISA) (Anselin, 1995) . A local version of Moran's I, LISA calculates for each feature (point or polygon) an index value and a Z score. A high negative Z score suggests that the feature is adjacent to features of similar values, either above or below the value.

Many studies have used Moran's I. For example, Beckler et al. (2005) used Moran' I coefficient to measure corn rootworm spatial distribution and determine the degree of autocorrelation for the interpolated species abundance map. In British Columbia, Canada (Christopher Bone, Wulder, White, Robertson, & Nelson, 2013) used local Morn's I to estimate the risk of mountain pin beetle attack on pines. It was also used to identify the degree spatial cluster of Lyme disease cases, ticks and forest disease (Kitron, 1998). Khormi and Kumar (2012) used Moran's I to investigate the averages of dengue fever and dengue hemorrhagic fever case prevalence in the northern parts of Jeddah, Saudi Aribia. However, Moran's I, either general or local, can only detect the presence of the clustering of similar values. It cannot tell whether the clustering is made of high or low values. Such types of method need more prevalent knowledge of how spatial patterns will change from the past to the future. There is also Geary's index statistic method which was developed by economist and statistician Robert Geary in the 1950s and is termed Geary's contiguity ratio. He first applied it to identify agricultural and demographic patterns in his native Ireland. Geary's index can be used to measure the similarity of nearby features, or similar values occurring together. However, like Moran's I function cannot discern from the statistics whether clusters are composed of high or low values.

4.1.1.2 G-Statistic for Measuring High/Clustering

G-statistic can be used to separate clusters of high from clusters of low values (Getis & Ord, 1992). The overall G-statistic method is based on a stated distance d. The weight can be based on some G (d) to evaluate its statistical significance. However, it is similar to Moran's I and a local version of the G-statistic is also available (Anselin & Getis, 2010; Ord & Getis, 1995). The local G-statistic, denoted by $G_i(d)$, is often described as a tool for "hot spot" analysis. Moreover, a cluster of high positive Z scores suggests the presence of a cluster of hot spot values. On the other hand, a cluster of high positive Z scores suggests the presence of a cluster of cold spot values. The G-statistic allows the use of a distance, but no discernible increase in clustering of high or low values exists. The general G-statistic based on a specified distance, *d*, is defined as:

$$G(d) = \frac{\sum\sum w_{ij}(d)x_i x_j}{\sum\sum x_i xj}, i \neq j \tag{6}$$

where x_i is the value at location i, x_j is the value at location j, if j is within d of i, and w_{ij} (d) is the spatial weight which can be based on some weighted distance or 1 and 0. The expected value of G (d) is:

$$E(G) = \frac{\sum\sum w_{ij}(d)}{n(n-1)} \tag{7}$$

where E (D) is typically a very small value when n is huge. A high G (d) value suggests a clustering of high values, and a low G.

4.1.2 Multi-distance, Ripley's K Function

Ripley's K function, or multi-distance spatial analysis cluster, can be used to identify clustering or dispersion of features over a distance. Also it is a standard technique that can be used to analyse the spatial distribution and structure of plant species (Hao, Zhang, Song, Ye, & Li, 2007; Wiegand & A Moloney, 2004). Spatial autocorrelation analysing has been used to investigate the temporal changes of spatial distributions to study rangeland grasshoppers (Goovaerts & Jacquez, 2005). Further, it is useful for quantifying the spatial dependency over distance classes (Overmars, De Koning, & Veldkamp, 2003). Spatial autocorrelation is also important for validating the use of standard statistical tests. According to M. Goodchild (2009), statistical interpretation typically applies to controlled experiments, which are seldom used in geographic research. If the data exhibits significant spatial autocorrelation, it should encourage the research to integrate spatial dependence into analysis.

4.1.3 Mapping Clusters

4.1.3.1 Hotspot Analysis Gi*

The other version, called Gi*, the value of target feature is included. Gi* mapping clusters can be used to identify the hot spot or cold spot of diseases, insect infestations, and migration of features. Gi* uses a neighbourhood, based either on adjacent features or on a set distance. However, using a distance-based neighbourhood, the distance specified is based on the knowledge of the feature and its behaviour. To calculate Gi*, the GIS sums the values of the neighbour and divides by the sum of the values of all features in the study area. Different studies with different applications have used Gi* hotspot worldwide (Kounatidis et al., 2008; Pinault & Hunter, 2011; Schools et al., 2012). The hotspot Analysis Gi* can be computed by:

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \tag{8}$$

where Gi* represents a feature i, at a distance d. The value of each neighbour x_i is multiplied by the weighted target- neighbour pair w_{ij}. A Gi* value close to 0 indicates that there is no concentration of either low or high values nearby the target feature. This happens when the surrounding values are close to the mean, or when the target features is surrounded by a mix of low and high values. Other methods and technologies are significant and can be used to give

scientific and sufficient results. Table 1 illustrates the differences between some methods for measuring the pattern of feature values.

Table 1. Comparing methods for measuring the pattern of feature values (*Source: (Mitchell, 2005)*)

Displays	Statistic	Advantages	Limitations
Similarity of nearby features.	Moran's I & Geary's	Delivers a single statistic and summarising the pattern.	Does not indicate if clustering is for high or low values
Concentration of high/low values.	General G	Indicates whether high or low values are clustered.	no visible rise in clustering of low or high values occurs.
Counting number of features within defined distances.	*K*-function	Calculating concentration of features at a range of scales or distances simultaneously.	Patterns are suspect at a distance because of edge effects

5. Examples of Use Modelling Spatial Relationships between Insect Infestations and Environmental and Climate Factors

Inverse Distance Weighted and Radial Basis Function, with the exception of kriging, are techniques that do not sufficiently address the problem of spatial dependence and spatial heterogeneity. However, kriging with its own assumptions, might be an effective method in an area in which there is a richness of available data, and is valuable for analysing the placement of new gauges (Tsintikidis, Georgakakos, Sperfslage, Smith, & Carpenter, 2002). Spatial autoregressive methods deliver a versatile and valid technique for guesstimating applications. Integration of these methods into ArcGIS environmental models makes it comparatively easy to apply the methods to a variety of spatial phenomena. GIS functions and tools have made it possible to implement cutting edge spatial autoregressive techniques to investigate many research study problems. Advances in the ArcGIS packages, a GIS application, in particular have greatly reduced time for estimating spatial parameters. For example, regression analysis allows users to model, examine, and explore spatial relationship, to better understand the factors behind observed spatial patterns. It further allows users to predict hypotheses based on the understanding of these factors. There are three types of regression; liner regression, local regression, and logistic regression. Linear regression has been used to predict values of *y* from values of x_i as computed by;

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_n x_n \qquad (9)$$

where *y* is the dependent variable, x_i is the independent variable *i*, and *bi*, …, *bn* are the regression confidents. However, linear regression requires several suppositions about error, or residual, between the predicted values and the actual values (Miles & Shevlin, 2001). Some errors are related to a normal distribution for a set of independent variables. Others have the expected mean value of zero. Linear regression has been used to model wildlife home ranges

(Anderson et al., 2005) and soil moisture (Lookingbill & Urban, 2004). According to Harris, Fotheringham, Crespo, and Charlton (2010), Local Regression or Geographically Weighted Regression (GWR) analysis has been used to predict information for each known point to derive a local model. However, parameters in this method may include variations in space, providing a basis to explore non-stationary spatial relationships. The logistic regression method can be applied to model spatial relationships between features, when the dependent variable is categorical (e.g., presence or absence) and the independent variables are categorical numeric, or both (Menard, 2002). The advantages of using the logistic regression method is that it is does not require the assumptions needed for the linear method.

Various studies have used autoregressive models to investigate the relationships between insect infestations and other factors based on environmental information. Munar-Vivas, Morales-Osorio, and Castañeda-Sánchez (2010) combined environmental information, spatial and attribute data in GIS-based maps to assess the impact of Moko disease *Ralstonia Solanacearum* on banana yield in Colombia. They used a regression model to investigate the relationship between infested areas and distances from *Moko foci* to cable-ways and drainage channels. Coops, Johnson, Wulder, and White (2006) researched the relationship between likelihood of attack, forest structure, and forest predisposition variables using regression tree models and found site index and slope to be the major factors for explaining variations in the probability of red attack. GWR model has been used to detect high-risk infestations caused by mountain pine beetle invasions of lodge-pole pine forests over large areas (Robertson, Wulder, Nelson, & White, 2008).

It is essential to begin by using single variables to develop correlations, before moving onto more complicated predictive models and regression analysis, in which all factors are incorporated to investigate which combination of factors is most conducive to the survival and spread of Dubas bug. For example, Geographically Weighted Regression (GWR) can be used to model correlation between Dubas bug infestation and meteorological variables such as humidity, rainfall, temperature, wind direction and speed. GWR can also be applied to model the correlations between Dubas bug infestations and the environmental variables such as soil type, slope, aspect, ecology, soil salinity and solar radiation. Additionally, autoregressive models may be used to investigate the relationships between Dubas bag infestations and human factors such as irrigation, plantation systems, insecticides and methods of spraying.

6. Use of Modelling Suitability Analysis for Detecting and Investigating Insect Infestations

A range of suitability models have been used to identify locations having a particular set of characteristics. These may also be used to predict areas that might be a susceptible to risks such as insect infestations, based on conditions of the species (Mitchell, 2011). For examples, Matern, Drees, Kleinwächter, and Assmann (2007) used the habitat suitability selection method to model the conservation of the rare ground bin beetle species. They used five different data sets to identify several key habitat factors for *Carabus variolosus* stress levels. C Bone, Dragicevic, and Roberts (2005), developed a model using fuzzy theory to identify susceptibility areas for *Dendroctonus ponderosae* Hopkins in Canada. However, GIS data has unique

characteristics which can impact on the results of the model (Allen, 2011; Crooks & Castle, 2012). Moreover, vector data uses coordinate pairs to define both location and shape of geographic features. Raster data is often used to represent geographic phenomena that are spatially continuous, such as elevation, temperature, soil moisture, rainfall, solar radiation, and slope at any location within a study area. Discrete features are typically represented with vector data, but raster data may be used if the analysis method requires. Thus, rasters are often used for finding and rating suitable locations, because in the raster overlay method results are formatted in a single layer of suitable and unsuitable cells known as a suitability layer, rather than a vector layer with many polygons and an attribute table containing the attribute values for each polygon.

There are two ways to create raster suitability layers. The first approach is to query the individual sources to create the suitability layer. The query can be used to create the suitability layer with two values, 1 for cells meeting all criteria of suitable habitat and 0 for all other cells. The second approach is to assign new values to each source layer to create a new layer with two values. Cells that meet the criteria are assigned one value and cell that not assigned another. However, the layer consists of only two values, indicating suitable and unsuitable cells, and these are termed binary suitability layers. A process that may be used to get an improved result is by combining the binary suitability layers into a single layer and identifying cells that were defined as suitable on all binary layers. Figure 1 shows the process that may be used for finding suitable location conditions for Dubas bug habitat using raster method overlay.
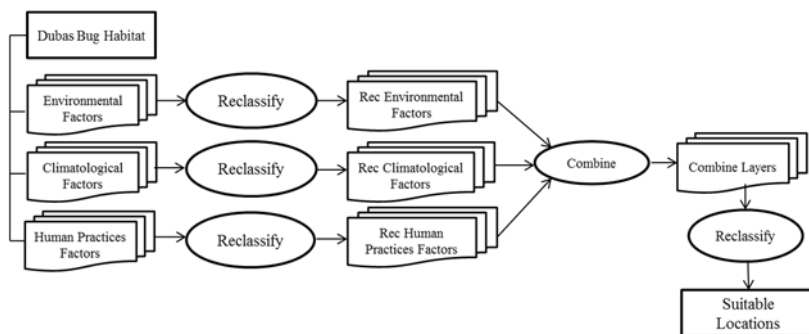

Figure 1. The process of finding the suitable location of Dubas bug habitat

The uncertainty resulting from geoprocessing operations demonstrates that sophisticated spatial analysis cannot be achieved using traditional, deterministic geoprocessing methods alone (M. F. Goodchild & Glennon, 2010; J. Zhang & Goodchild, 2002). Fuzzy logic method is a superset of Boolean logic lengthened to handle uncertainty in data that arises due to vagueness instead of to randomness alone (Li, Huang, Guo, & Nie, 2010). Fuzzy set theory was applied to information extracted from high resolution remote sensing data and combined in a raster-based GIS to create maps indicating the spatial variation of insect susceptibility in a landscape. Fuzzy sets allow for partial membership to one or more class, thus objects are represented by a value based on a membership continuum between 0 and 1. The membership function of an element *x* belonging to a fuzzy set A is computed by:

$$\mu_A : U \rightarrow [0,1] \tag{10}$$

14

where U is the universal set of x. The concept of fuzzy sets has also been employed for defining the spatial and attribute characteristics of geographic objects (Burrough & Frank, 1996; F. Wang & Hall, 1996). The results from the analysis can be fed directly into a decision via maps, tables and charts. The results can also be used in further analyses or to provide additional understanding of the problem. Furthermore, the result can be tested to see if layers are correlated using a statistical tool such as Pearson's correlation coefficient or Spearman's rank correlation coefficient, or by using regression analysis (Mitchell, 2005) . The Spearman's rank correlation coefficient can be computed by:

$$p = 1.0 - \frac{6 \times D^2}{(n^3 - n)} \tag{11}$$

A p value ranges from -1 to +1. The closer p is to +1 or -1, the stronger the positive or negative spatial correlation between two categories of features. For interpreting value p uses the following guides:

- $P = +1$ implies complete positive spatial decency

- $P = -1$ implies compete negative spatial dependency.

- $P = 0$ implies no spatial dependence.

The results of any model may be used to generate summary information that will be useful for decision makers or researchers. For example, we might calculate statistics such as the total amount of suitable habitat in the study area or the amount of suitable habitat in each landownership category. The challenge in any particular area of study is the geographic extent and resolution of analysis which is determined by the phenomenon being modelled. To achieve validity, researchers must ensure that they are using accurate, current data wherever possible. If the data is from one's own organization, one can rely on the data controls that are in place. Data quality should be checked against alternate sources, if possible, to ensure it meets the requirements of the analysis. Assessing the quality of data is a guideline to predict what level of confidence can be attributed to the result of the modelling.

7. Example of Using Solar Radiation Model to Investigate Insect infestation Level

The solar radiation model was developed by Kumar (1997) and has been applied to calculate the possible solar radiation at chosen locations for a 12 month period. The model may be applied to investigate infestations level differences in different locations during the infestation periods. For example, Kirkpatrick and Nunez 1980 achieved a positive result after investigating the relationships between solar radiations and the distribution of several species of eucalyptus a long a single transect in the Risdon Hills in Tasmania. The advantage of solar radiation is to calculate the potential on any slope and any latitude described (L. Kumar, Skidmore, & Knowles, 1997). The model has also been used for applied work in ecology, biology, forestry and agriculture where the spatial variation of solar radiation is more important than calibrated values. Incorporating solar radiation over long periods of time for huge areas can be used for modelling in a variety of areas such as plant growth, species locations, biodiversity and

identification of possible flora and fauna sites, and mapping topographic variations in direct shortwave radiation. The advantages and limitations of this model depend on the accuracy of the Digital Elevation Model (DEM) used. Mistakes in the DEM mean errors in the calculated values of slope and aspect which affect the accuracy of the computation of solar radiation. Another limitation is the sun position in the sky and the impact of shading by adjacent terrain.

8. Example of Using Humid-Thermal Index for Detecting Plant Disease

The Humid-Thermal Index, sometimes called the Humid-Thermal Ratio, has successfully been demonstrated in developing and testing relationships between several plant disease and varied climate conditions in regions including India, Australia, Europe and North America. An HTI model has been developed to simulate environmental conditions suitable for establishment and spread of plant diseases in India (Jhorar et al., 1992). The HTI method has been used to evaluate risk of *Karnal* bunt establishment and spread in wheat growing under a variety of climate conditions and in different areas (Jhorar et al., 1992; Stansbury & Pretorius, 2001; Workneh et al., 2008). The method has potential value in researching insect pests and associated diseases, allowing the possibility of accurately predicting occurrences and non-occurrences under specific climate and weather conditions. These predictions can be used to help reduce human efforts when conditions are expected to be unfavourable. Predictions can further be used for advance preparations to meet challenges in the event of threat of heavy crop losses. There is a need to study the biological systems of Dubas bug because these are dependent on two of the most important elements of climate, temperature (thermal) and humidity (moisture). Temperature is the derivative of solar radiation and thermal emission, while humidity is the derivative of dry or wet conditions associated with various levels and forms of precipitation.

9. Conclusions

A variety of applications of modern spatial information technologies such as GIS and RS have been shown to have application to areas of research on insect infestation worldwide. Environmental and climatological conditions are very important in determining the distribution and survival of any species including Dubas bug. This review highlighted that most of global research on Dubas bug has concentrated on its ecology, biology, or control mechanisms. There has been limited research on linking the presence, density and spatial temporal distribution of the Dubas bug with environmental, meteorological and human practices that promote the development and prevalence of this pest. Understanding the distribution and affinity of the Dubas bug in terms of variables, can play a key role in its mapping, control and management, including resource allocations. Remote sensed satellite data is familiar to GIS users. The spatial resolution of satellite images relates to ground pixels. Satellite images can be used processed digitally to produce a wide variety of thematic data for a GIS project such as land use, vegetation types, crop health, eroded soil, etc. Satellite images provide timely data, if collected at regular intervals; they can also provide temporal data valuable for recoding and monitoring changes in terrestrial and aquatic environments. A GIS can assist environmental modellers in data visualization, database management and exploration. A regression model relies on overlays to combine data in the statistical analysis of dependent and independent variables. A GIS and its functionalities can be used to build or help build, a spatially explicit model with

geospatial data. Climate conditions such as extreme temperature, high relative humidity, occurrences of cyclones and severe rain and hail storms, as well as soil and water salinity and traditional outdated methods of cultivation, need more investigations in the future research.