An investigation into the spatiotemporal patterns of the Nymphalid butterfly *Vagrans egista sinha* (Kollar, [1844])

Authors: Paul Pop¹, Kuldeep Singh Barwal¹, Randeep Singh², Puneet Pandey³, Harminder Pal Singh⁴, Sanjay Sondhi⁵

¹Zoology Department, Government College Bilaspur, Kosrian Sector, Bilaspur - 174001, Himachal Pradesh, India

²Amity Institute of Forestry and Wildlife (AIFW), Amity University, Sector 125, Gautam Buddha Nagar, Noida - 201313, Uttar Pradesh, India

³Conservation Genome Resource Bank for Korean Wildlife (CGRB), Research Institute for Veterinary Science and College of Veterinary Medicine, Seoul National University, Seoul, Republic of Korea.

⁴Department of Environment Studies, Panjab University, Sector 14, Chandigarh - 160014, India

⁵Titli Trust, 49 Rajpur Road Enclave, Dhoran Khas, near IT Park, P.O. Gujrada, Dehradun - 248001, Uttarakhand, India

Email: rsingh18@amity.edu, puneet.pandey09@snu.ac.kr, paulvpop@gmail.com

Abstract

Vagrans egista sinha (Kollar, [1844]), the Himalayan Vagrant is a subspecies of Nymphalid (Brush-footed) butterflies spread across Asia, whose western limit is in the north-west India. Observations of this subspecies have considerably increased over the past half-a-decade, with a spike in new sightings to the west of their previously known range. This has been considered as a range extension. The current study reports new records of this species from Bilaspur District, Himachal Pradesh, India (which are the first records for the district), through systematic and opportunistic sampling. This raises the question of whether the purported range extension towards the west could instead be a range shift or vagrancy, and whether there is any shift in elevational ranges in the populations across their known range. Questions pertaining to spatial differences in elevational ranges and seasonal variation, across their range, also piqued our curiosity. Using data from academic sources (such as published literature and museum collections), supplemented by data from public participation in scientific research and personal observations, these research questions are addressed. The accuracy of results when using citizen science data is also explored using the same dataset, focused on the impact of method of extraction of coordinates, and elevation derived from it under different scenarios. It was discovered that there has not been a range shift (either longitudinal or latitudinal) and observations don't suggest vagrancy but a case of range extension. Other results indicated that there was no climb of population to higher elevations, no spatial differences in elevational ranges in the populations, or seasonal variation in activities across their range. It was also discovered that the method of data collection by, and extraction from, citizen science databases, can influence the accuracy of the results. Some problems involved in collecting data are discussed, and remedial solutions are suggested.

Keywords: *Vagrans egista sinha*; Himalayan Vagrant; Citizen Science; Spatiotemporal patterns; Species distribution; Autecology

Introduction

The genus *Vagrans* is considered as monotypic or bitypic by various authors (Kirti, Mehra & Sidhu, 2016). *Vagrans egista sinha* (Kollar, [1844]), the Himalayan Vagrant is a subspecies of *Vagrans egista* (Cramer, [1780]) - the Vagrant butterfly, a species whose distribution is far and wide in the Indo-Australian region, with the eastern limit being French Polynesia in the Pacific Ocean; western and northern limits at India; and the southern limit being Australia (GBIF, 2021). They are a largely orange-and-black coloured butterfly belonging to the Nymphalidae (Brush-footed) family, which is the largest family of butterflies (see Fig. 1 for the photograph of an individual). Different authors have adopted different genus, species, and subspecies names as well as author citations to refer to the Vagrant butterfly found in India — *Vagrans sinha sinha* Kollar, 1848 (Larsen, 2004), *Issoria sinha sinha* (Kollar) (Singh et al, 2011) and *Vagrans sinha pallida* Evans (Smetacek, 2012). We will be using the name *Vagrans egista sinha* (Kollar, [1844]), which is the most widely adopted name in circulation (Gokhale, 2021).

Figure 1: Species under study — *Vagrans egista sinha* (Kollar, [1844]), the Himalayan Vagrant

Their local occurrence varies from very common (Sondhi & Kunte, 2016) to rare (Bhowmik & Chowdhury, 2021), depending on the location. In India, this species is reported in tropical deciduous forests within 400-1400 m above sea level and *Shorea robusta* forests below 1000 m asl, and habitat-wise, they are found within riverine habitat, village homes near forest edges,

and gardens (Smetacek, 2012; Sondhi & Kunte, 2018; Bhowmik & Chowdhury, 2021). Similarly, in Thailand and Vietnam, they are reported in very moist, lightly to partially shaded areas or gaps in the primary forest (Spitzer et al, 1997, Küppers & Janikorn, 2009). The immature stages of this subspecies is recorded to feed on plants of the following genera: *Dillenia* (Dilleniaceae), *Maytenus* (Celastraceae), *Flacourtia*, *Homalium*, and *Xylosma* (all Salicaceae members) (compiled in Gokhale & Yathumon, 2021). The same study reports the confirmed use of *Xylosma longifolia* Clos by this species as a larval host plant in Dehradun (Western Himalaya). The adults are observed to feed on flower nectar (*Polygonum hydropiper* L. - Water Pepper Plant and *Lantana camara* L. - Wild Sage, being two such plants) and animal excretions (bird droppings, cow urine and dung), and also involve in mud-puddling (Naro & Sondhi, 2014; Sondhi & Kunte, 2018; Sharma & Sharma, 2018; Kumar, Kushwaha & Namdev, 2020). They have been found to opportunistically feed on flower nectar in the forest canopy during some migrations in the Tam Dao mountains in Vietnam (Spitzer et al, 1993). Migrations are not recorded in India. Parsons & Cantlie (1948) report that they are known to sit on hot tarmac and gravel by the roadside in winter, most likely an adaptation to warm up during the cold season. They also hint at possible variation in size in Assam compared to adjacent areas.

This subspecies is distributed in 10 countries (earliest known explicitly mentioned record given in brackets) (Fig. 2): India (Bingham, 1905), containing the most northern and western distribution of the species as well as the subspecies; Nepal (Varshney, 1994); Bhutan (Singh, 2012); Bangladesh (Larsen, 2004); Myanmar (Bingham, 1905); Thailand (Varshney, 1994), containing the southernmost distribution of the subspecies; Cambodia (Savelle, nd); Laos (No record, but most likely); Vietnam (Spitzer et al, 1993); and China (D'Abrera, 1940), including the special autonomous region Hong Kong (Bascombe, Johnston & Bascombe,

1999) — containing the easternmost distribution of the subspecies. In India, they are found within 14 states and one union territory across the Himalayan range, its foothills, Eastern Ghats and near the Eastern coast — Jammu & Kashmir (Sharma & Sharma, 2018), which is the northernmost and westernmost distribution in their range; Himachal Pradesh (Kirti, Mehra & Sidhu, 2016); Chandigarh (Soman, 2015); Uttarakhand (Bingham, 1905); Uttar Pradesh (Kumar, Kushwaha & Namdev, 2020); Sikkim (Wynter-Blyth, 1957); West Bengal (Bingham, 1905); Arunachal Pradesh (Kunte, 2007), their easternmost distribution in India; Assam (Bingham, 1905); Nagaland (Tytler, 1911); Mizoram (Bhakare, 2009); Tripura (Zothansangi, 2015); Odisha (Wynter-Blyth, 1957); Chattisgarh (Naidu, 2018); and Andhra Pradesh (Saragada, 2020), which is their southernmost distribution within India.

Figure 2: (A) Global distribution of the subspecies on a WGS 1984 projection of Earth.

(B) Individual records of the subspecies (red pins) and species (black circles) on a globe.

Records of this subspecies have increased considerably over the past half-a-decade, with

many new sightings to the west of their previously known range in North-West India. Upon

cursory glance, there is also a lack of recent records in much of South-East. This apparent range extension towards the west could therefore be a range shift or a case of vagrancy. This may indicate disturbance in parts of their range or simply processes such as increasing population sizes, overshoot of dispersal range or getting blown off by wind. We adopt Jain et al's (2018) definition of vagrancy as sporadic sightings of a species with no breeding records. While Jain et al (2018) definition allowed for a maximum of three individuals in the span of 17 years in a country, we modify it and adopt the following criteria, failing any of which a species cannot be considered as a vagrant: i) maximum of 3 sightings per year in a geographical unit, which is a district in our case (which is far less than for a migratory species of butterfly) ii) no breeding records iii) no repeated sighting in the same unit (district) in at least two consecutive years. Range shift is defined as the changes in geographic dimensions of a species range (which are latitude, longitude, and elevation) over the fourth dimension i.e. time (Lenoir & Svenning, 2013). For analytical purposes, this indicates the presence of a leading edge with expansion into previously unoccupied area, as well as a trailing edge where the species is going locally extinct (Lenoir & Svenning, 2013). Range extension is equal to range shift minus trailing edge. This can be confirmed by default, if there is no evidence for range shift or vagrancy.

“Citizen science” refers to public participation in scientific research (‘c-science’ henceforth). C-science is extensively used in lepidopterology, especially for butterflies, with the number of species studied and sites monitored outweighing all vertebrate groups in Europe (Schmeller et al, 2009). It is used for several purposes, such as collecting photographs of rare butterflies (Mesaglio et al, 2021), monitoring populations at a national level (Sanderson, Braby & Bond, 2021), and even studying the range expansion of a Nymphalid butterfly species distributed through the Indomalayan and Australasian realms (similar to the current study) (Chowdhury

et al, 2021). Reliability of c-science data, especially when it comes to spatiotemporal patterns of butterflies depends on a wide-variety of components of a c-science model, including accuracy of identification of species by participants, particularly for cryptic species (Vantieghem et al, 2016); irregular nature of survey efforts over time; incompleteness and selectivity in observation; geographical bias (van Strien, van Swaay, & Termaat, 2013). Studies of butterflies have been done to compare the reliability of c-science data (opportunistic data) in comparison to standardized monitoring data. In assessment of distribution trends using occupancy models (van Strien, van Swaay, & Termaat, 2013), and species population trends (Dennis et al, 2017), it was shown that the former performs almost as well as the latter. Differences have been found in the comparison of quantity and quality of outputs of two regions (Matteson, Taron & Minor, 2012) when the monitoring protocols are different: c-science resulted in more species due to unrestricted sampling but standardized monitoring were more likely to detect trends due to consistent sampling. While much work has been done on the reliability of c-science data, as far as we know, there has been no study to look at the influence of reliability of coordinate data derived through various means, and multiple scenarios of secondary data derived from these coordinates (such as elevation) to be used as variables/covariates.

We addressed five questions, namely: 1) are increasing records of the subspecies westwards of their known range indicative of a range extension, range shift, or vagrancy?; 2) are there seasonal changes in distribution across its range?; 3) are populations shifting to higher elevations across their known range, possibly indicative of climate change?; 4) is there a gradient of elevation across distributions of populations; and 5) how reliable is c-science data for addressing such ecological questions? Questions 1 and 3 are prespecified (hypothesis-

based) whereas the rest are exploratory in nature (some subsections within the prespecified questions were also exploratory).

Materials and Methods

Data characteristics

As part of the project “High Resolution Spatial Mapping of Bird Phenology as an Indicator of Ecosystem Health in Relation to Climate Change in Himalaya”, surveys (which includes invertebrate monitoring) were undertaken in Bilaspur District of Himachal Pradesh from March 2020 to April 2021, through which some of the data were collected. These surveys included both reconnaissance surveys (done to design a stratified random sampling protocol) and stratified random sampling points (designed for systematic surveys of birds, invertebrates and plants). Identification was carried out using Sondhi & Kunte (2018), an easy task as there are no species that are confusingly similar in the region. Bulk of data were collected for the current study from a c-science website (Butterflies of India); a data aggregator (GBIF, which contained data from c-science portals such as iNaturalist and India Biodiversity Portal, as well as records of specimen collections in museums); peer-reviewed and grey literature; and communication with authors of the literature. A total of 216 usable data points were obtained from this (Fig. S1) for a total of seven countries (Table S1; a record from one country — Cambodia was missed at the time of analysis, and hence excluded), the contribution from literature and communication with authors constituting the bulk at ~55.1%, followed by c-science at ~36.1% . Of the 24 authors contacted (including multiple authors of the same paper) through email or websites such as Research Gate and Academia Edu regarding 20 papers for more information regarding their records, only nine people responded (37.5% response rate) regarding 10 papers (50%) covering various countries such as India (Singh, Gupta & Varatharajan, 2011), Bangladesh (Khandokar et al, 2013), China (Kendrick (Ed), 2005), and

Thailand (Küppers & Janikorn, 2020). All except one was through email (one being Research Gate messaging). Coordinates for location names were deduced using the given landmarks, elevation information, given coordinates of adjacent locations, comparison between historical names and current names etc. For the current study, no records were available to a desired resolution needed for use in analysis from Myanmar, Laos and Cambodia.

Data analysis

Data sorting, analysis and visualisation were carried out using WPS Excel, R (v 3.6.1), R Studio (v 4.1.0), and QGIS (v 3.8.2) for studying the autecology of the *V. e. sinha*. Selection: Some records were omitted due to locations covering a large area and ambiguity in their name (due to recent changes in the name). In one case where there were two records within half an hour interval on the same site, only one record was retained. When there were entries duplicated in multiple sources, only the one with the most reliable information (such as closeness of name of location and coordinates provided) was retained. For sight-based records with no coordinate uncertainty information, different coordinate uncertainty values were assigned for analysis, ranging from 0 km to 10 km. For museum specimens with no coordinate uncertainty information, at least 5 km was assumed unless more details were available. The assignment of coordinate uncertainty was carried out under manual supervision, based on all the available information for each record (such as coordinates within an area such as a National Park having a coordinate uncertainty equal to the radius from the point that which could be realistically covered by walking or in a vehicle depending on the availability of trails or roads, whose limit is set by the boundaries of the Park).

To retain a greater number of records, a novel approach has been taken wherein median dates were calculated, as well as adjusted-median dates. Median calculation was done as

follows: When date range was only mm-yyyy, dd was assumed to be 01 (first day of that month) for begin dates and 28 to 31 (last day of that month) for end dates. Similarly, for yyyy, dd-mm were assumed to be 01-01 (first day of that year) for begin dates and 31-12 (last day of that year) for end dates. For dates with only one mm-yyyy, dd was assumed to be 14-16 (mid-month) and similarly for only one yyyy, dd-mm were assumed to be 01-07 (mid-year). Median date was also adjusted for the information about the months reported. For example, if the study was done from March 2012 to February 2013, the median date would be the 30th of August 2012. If it was given that the species was recorded from September-October, then the date was adjusted to the closest month i.e. the 30th September, 2012. This is the adjusted-median date. Dates were converted to date values to use as a continuous measurement. Since individual zones can have different patterns, three clusters were compared using histograms: the Western Himalayan cluster, including the foothills; the North-East Indian cluster, including Bhutan and Bangladesh; and the South-East Asian cluster, including China.

Data were tested for normality, using the Shapiro-Wilk test and q-q plots (values from Shapiro-Wilk test and q-q plots are available in Table S2 and Fig. S2-S5, respectively). Shapiro-Wilk test is inadvisable when sample sizes are extremely large, which as a rule of thumb is >250 (Ruxton, Wilkinson & Neuhäuser, 2015). As our data were a maximum of 216 datapoints, it was used. However q-q plots were used to make the final decision. When the p-value was <0.05, the null hypothesis that the data are normally distributed was rejected (alpha level = 0.05). If the p-value was > 0.05, then the null hypothesis was not rejected. For hypothesis testing using correlation, it was decided that Pearson's Product-Moment Correlation (PPMC) would be used if the data were normal, and if non-normal, Spearman's (rho) and Kendall's rank correlation (tau) were to be used (Crichton, 2001). Spearman's rho was ignored and Kendall's tau was given importance when the used data had a considerable

amount of ranking ties because tau is more accurate in that condition (Nie, Bent & Hull, 1970). The statistical interpretation of PPMC and rho were based on Dancy & Reidy (2007) (+/-1 = perfect correlation, +/-0.9 to +/-0.7 = strong, +/-0.6 to +/-0.4 = moderate, +/-0.3 to +/-0.1 = weak, and 0 = zero), which is a rather conservative scale compared to other approaches (see Akoglu, 2018). Tau was interpreted as follows: +/-1 = perfect correlation, +/- 0.99 to +/- 0.40 = strong, +/-0.39 to +/-0.35 = moderate, and +/-0.34 to +/-0.1 = weak, and 0 = zero. Final decision was made based on the context of questions addressed, as is generally advised (Schober, Boer & Schwarte, 2018). To avoid p-hacking (Head et al, 2015), the tests were carried out under different scenarios, including outliers removed, and p-values for all these tests were reported. To avoid overemphasis on p-values due to its rather unreliable nature (Nuzzo, 2014), visualisation of data was carried out when necessary and results were interpreted with biological plausibility and relevance in consideration. Histograms were created for all the variables considered.

For addressing the first question, check for negative correlations between longitude and dates or years (which indicates westward shift) were done, and observations were checked for fulfillment of the criteria for vagrancy. For understanding seasonal variation, correlations between months and coordinates were done. Third, the question was addressed by checking for a positive correlation between elevation and dates. The exploratory analysis to answer the fourth question involved checking for positive or negative correlations between elevation and coordinates.

The final question was tackled using the following steps. To find out the differences between coordinate deduced from name and that directly acquired from the source, 39 unique datapoints from the Butterflies of India website (Gokhale, 2021) were used to find the shortest

distance between them and 39 corresponding acquired ones. The distance was found using the Vincenty Ellipsoid method which is accurate up to a few millimetres, using the R package geosphere. A subsample of these distances was manually verified for accuracy using Google Earth. For points with no reported coordinate uncertainty (a total of 63 points, a majority of them from c-science sources), four scenarios were considered. In the first scenario, the coordinates reported were assumed to be completely accurate. In the second, third and fourth scenario, buffers of 0.5, 5, and 10 km were drawn with the coordinates as the centre, and the mean elevation for each of such coordinates was derived from the ASTER GDEM raster layer (ASTER GDEM is a product of METI & NASA) using zonal statistics plugin in QGIS (Fig. S6). Note that due to the 2D nature of the buffer and 3D nature of the Earth, the fixed-width buffer actually had variable length from the centre, which is offset by the fact that this variation was present across the range. The biological relevance of this step is that it calculates an elevation where the subspecies probabilistically has a higher chance of being for the largest amount of time. Even in cases where the elevation of a record is known to 1 m accuracy, buffers like these give a more biologically realistic idea of the elevation due to the motile nature of butterflies. All such scenarios were compared by checking for correlation with other variables, by comparing the correlations between scenario 1 and scenarios 2-4 and, as well as doing a one-way ANOVA. For the latter two steps, 63 points alone as well as these points included in the 216 total points were analysed separately to understand the variations of the values of variable buffer points alone, as well as their influence on data when mixed with unchanging elevation values. Tests to satisfy the assumptions of ANOVA were also carried (independence was assumed, qq plots for normality, and Bartlett's test for homogeneity of variance).

Results

During field surveys, five sightings of *V.e.sinha* in their adult stage were made in the district (see Fig. 3 for sighting details). Two of these sightings were made during reconnaissance surveys and three of these sightings were made within stratified random sampling points. In Rani Kota, two individuals were sighted, whereas in Kiarian, three individuals were sighted, two of them likely mating in the air near and on a *Mangifera indica* (Mango) tree. At rest of the sighting localities, only one individual was found. In a majority of the sightings, they were feeding on the invasive plant *L. camara* (Fig. 1), a behaviour which has also been seen in Uttar Pradesh (Kumar, Kushwaha & Namdev, 2020).

Figure 3: Sightings of the species during the current study from Bilaspur District, Himachal Pradesh.

Date, latitude and longitude were found to be non-normal data in all scenarios, except for elevation which was found to follow a normal distribution. Hence, non-parametric tests were used for most cases. Rho and tau were used for most of the scenarios, except for addressing the question on the reliability of c-science data, where PPMC was used. It can be observed from the histogram of years (Fig. 4E), that most of the records collected were after 2010, a lesser proportion between 2000 and 2010, and even lesser from before the start of the 21st century, indicating a skew towards the last two decades. When histograms were generated cluster-wise (Fig. 5A), it has been found that most records from earlier years were from the SE Asian cluster, and over there, records in recent history are comparatively few. In the NE Indian and the W Himalayan clusters, records started growing from 2008, with the former having tall peaks during the last decade, whereas the latter showed a relatively consistent and moderate number of records. Month-wise distribution of records show a bimodal distribution nature with a small peak at spring/post-winter and a taller and wider peak during the final months of monsoon and most of the winter/post-monsoon months (Fig. 4C). Individual histograms for the three clusters (Fig. 5B) indicated bimodal distribution in all the clusters. However due to data deficiency and spread in the SE Asian cluster, it cannot be asserted with confidence that this is the pattern for this area.

The elevation histogram (Fig. 4D) shows two peaks, indicating most sightings at the elevation between 0 and 200 m asl and around half that many between 600-800 m asl. Individual histograms (Fig. 5C) show that this pattern is mostly contributed by records from the NE Indian cluster and the W Himalayan cluster which show the largest peak from 1000-1200 m asl (but could be an artefact of sampling bias). In the longitude histogram (Fig. 4A), there are two major uphills — between 75-80° E being Uttarakhand, Himachal Pradesh, and Jammu & Kashmir (most records from Uttarakhand), and between 90-95° E being North-East India, and

a small adjacent peak from 95-100° E showing records from Thailand. Latitudinal distribution (Fig. 4B) shows a marked high frequency between 25-30° N, which is where North-East India and parts of Uttarakhand (Nainital District) are located.

Figure 4: Histogram of variables across the species range. (A) Histogram of longitude (n = 216). (B) Histogram of latitude (n = 216). (C) Months under different scenarios — S4 being

334 actual months (n = 168) and S1 being actual+median+adjusted median dates (n = 208). (D)

335 Elevation under different scenarios (n = 216, four figures in the same column) — Elev 0, Elev

336 0.5, Elev 5, Elev 10 (with 0 km coordinate uncertainty kept as zero or changed to 0.5, 5, and

337 10 km respectively, for 63 points). (E) Years under different scenarios (three figures in the

338 same column) — Year S1, Year S4, and Year S6 (YearS1 = actual years + median years +

339 adjusted median years, n = 198; YearS4 = actual years, n = 165; and Year S6 = YearS1 - pre-

340 2000 records, n = 190).

341

Figure 5: Histograms of variables in different clusters — the North-West Himalayan

cluster, the North-East Indian cluster, and the South-East Asian cluster. (A) Years in

various clusters, combined. Please note that in areas of overlap, colours are additive. In areas

where the three clusters overlap, colour is bright dark green; Indian clusters together =

medium green; and NE Indian+SE Asian = purplish-green. (B) Months in various clusters. (C)

Elevation in various clusters.

Main findings are as follows:

1) No significant positive or negative linear relationships between dates (as absolute dates and years) and coordinates were found, indicating no range shift, either longitudinal or latitudinal. The results of this are summarized in Table 1 and 2. Note that in scenario 5 for correlations between coordinates and years, the p-value was > 0.05 . This indicates that the null hypothesis that these two variables are uncorrelated is not rejected. However, this may be due to low sample sizes. This scenario is the least realistic as the actual known dates are removed. In scenario 6 and 7 which are realistic ones, where the sparse (outlier) data from pre-2000 and pre-2005 have been removed, the p-value was almost always > 0.05 , indicating non-correlation.

Although vagrancy seems like a plausible explanation for the sightings of the subspecies in Himachal Pradesh and Jammu & Kashmir, it is unlikely due to repeated sightings in Solan District across years and the relatively high number of observations in Bilaspur District in a multi-season multi-year study (March 2020 to April 2021). Surveys were often single-season or single-year, and not multi-season multi-year. This sampling bias has resulted in knowledge gaps. The recent observations in their north-western-most part is, therefore, most likely a case of range extension.

Table 1:
Correlations between dates and coordinates.

Spearman's rank correlation			Kendall's rank correlation		
Test statistic (S)	p-value	Sample estimate (rho)	Test statistic (z)	p-value	Sample estimate (tau)

	Date (n = 161), Scenario 1 = only actual dates					
Latitude	598069	0.07627	0.1401123	2.0147	0.04393	0.107455
Longitude	892801	0.0002662	-0.2836457	3.6815	0.0002319	-0.196351
	Date (n = 198), Scenario 2 = actual dates + median dates + adjusted median dates					
Latitude	1079744	0.01989	0.1653826	2.502	0.01235	0.1200731
Longitude	1631627	0.0002016	-0.2612106	-3.826	0.0001302	-0.1836157
	Date (n = 190), Scenario 3 = scenario 2 - outliers					
Latitude	991019	0.06721	0.1330689	1.9469	0.05154	0.09543789
Longitude	1397614	0.002021	-0.2226151	-3.1475	0.001647	-0.1542937

Table 2:
Correlations between years and coordinates.

	Spearman's rank correlation			Kendall's rank correlation		
	Test statistic (S)	p-value	Sample estimate (rho)	Test statistic (z)	p-value	Sample estimate (tau)
	Year (n = 198), Scenario 1 = actual years + median years + adjusted median years					
Latitude	1094096	0.02998	0.1542888	2.3691	0.01783	0.1170505
Longitude	1606976	0.0005879	-0.2421557	-3.5747	0.0003507	-0.1766173
	Year (n = 195), Scenario 2 = actual years + median years					
Latitude	1058328	0.04521	0.1435955	2.1907	0.02848	0.1091555
Longitude	1516382	0.001412	-0.2270651	-3.3009	0.0009638	-0.1644785
	Year (n = 168), Scenario 3 = actual years + adjusted median years					
Latitude	664048	0.03867	0.159693	2.3041	0.02122	0.1240884
Longitude	1016653	0.0001665	-0.2865052	-3.8052	0.0001417	-0.204935
	Year (n = 165), Scenario 4 = actual years					
Latitude	639561	0.06181	0.1457253	2.0872	0.03687	0.1135471
Longitude	950502	0.000462	-0.2696045	-3.5193	0.0004327	-0.1914519
	Year (n = 33), Scenario 5 = median years + adjusted median years					
Latitude	7597.9	0.129	-0.2697014	-1.7569	0.07893	-0.2331204
Longitude	5614.3	0.7327	0.06178231	0.11495	0.9085	0.01526533

	Year (n = 190), Scenario 6 = Scenario 1 - pre-2000 records					
Latitude	1005034	0.09685	0.1208094	1.804	0.07123	0.09126776
Longitude	1373053	0.005393	-0.2011293	-2.8784	0.003997	-0.1456269
	Year (n = 177), Scenario 7 = Scenario 1 - pre-2005 records					
Latitude	874965	0.4815	0.05324805	0.69082	0.4897	0.03637827
Longitude	1038081	0.1022	-0.1232504	-1.6799	0.09298	-0.08846514

2) No significant correlations between months and coordinates were found, indicating no linear pattern of seasonal variation (summarised in Table 3). However, Spearman's correlation between Longitude and Months showed a moderate negative relationship. In this case, Tau was given priority due to a considerable amounts of tied ranks. Like earlier, probably as a result of a lower sample size, scenario 5 showed p-values of > 0.05 .

Table 3:
Correlations between months and coordinates.

	Spearman's rank correlation			Kendall's rank correlation		
	Test statistic (S)	p-value	Sample estimate (rho)	Test statistic (z)	p-value	Sample estimate (tau)
	Month (n = 208), Scenario 1 = actual months + median months + adjusted median months					
Latitude	1259502	0.0208	0.160211	2.3129	0.02073	0.113716
Longitude	2047400	5.873e-08	-0.3651301	-5.1845	2.166e-07	-0.2549036
	Month (n = 195), Scenario 2 = actual months + median months					
Latitude	948767	0.001086	0.2322526	3.1877	0.001434	0.1620078
Longitude	1744817	2.193e-09	-0.4119154	-5.6958	1.228e-08	-0.2894842
	Month (n = 181), Scenario 3 = actual months + adjusted median months					
Latitude	756835	0.001509	0.2341742	3.0401	0.002365	0.16099
Longitude	1348105	4.683e-07	-0.3641196	-4.8537	0.000001212	-0.2570329
	Month (n = 168), Scenario 4 = actual months					

Latitude	527106	0.00001032	0.3329839	4.1498	0.00003327	0.2285832
Longitude	1130038	5.999e-09	-0.4299857	-5.5386	3.048e-08	-0.305083
Month (n = 41), Scenario 5 = median months + adjusted median months						
Latitude	14776	0.06872	-0.2871365	-1.4322	0.1521	-0.1666446
Longitude	14602	0.08545	-0.2719221	-1.7583	0.07869	-0.2047052

3) No significant linear correlations between elevation and date were found, indicating that no climbing up to higher elevations occurred over the years (Table 4). P-values never dropped below 0.05 but were consistently high. Hence they are likely to be highly uncorrelated.

Table 4:

Correlations between elevations and dates.

	Spearman's rank correlation			Kendall's rank correlation		
	Test statistic (S)	p-value	Sample estimate (rho)	Test statistic (z)	p-value	Sample estimate (tau)
Elevation, Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for 63 points						
Actual dates (n=161)	717927	0.685	-0.03221634	-0.20615	0.8367	-0.01099546
Actual dates + median dates + adjusted median dates (n=198)	1279680	0.8796	0.01083613	0.4002	0.689	0.01920503
Elevation, Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for 63 points						
Actual dates (n=161)	727313	0.5648	-0.04571181	-0.43716	0.662	-0.02331297
Actual dates + median dates + adjusted median dates (n=198)	1293101	0.9948	0.000462296	0.20063	0.841	0.00962677
Elevation, Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points						
Actual dates (n=161)	713556	0.744	-0.02593102	-0.12428	0.9011	-0.006627434
Actual dates + median dates + adjusted median dates	1259871	0.7146	0.02614833	0.58474	0.5587	0.02805663

(n=198)						
	Elevation, Scenario 4 = with 0 Km coordinate uncertainty changed to 10 Km for 63 points					
Actual dates (n=161)	697627	0.9696	-0.003029815	0.16521	0.8688	0.008810589
Actual dates + median dates + adjusted median dates (n=198)	1233450	0.5147	0.0465713	0.84867	0.3961	0.04072072

4) No significant linear correlations between latitude and elevation, or longitude and elevation, were found, indicating no linear pattern of elevation across populations (Table 5).

Table 5:
Correlations between elevations and coordinates.

	Spearman's rank correlation			Kendall's rank correlation		
	Test statistic (S)	p-value	Sample estimate (rho)	Test statistic (z)	p-value	Sample estimate (tau)
	Elevation (n = 216), Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for 63 points					
Latitude	1216277	0.00003947	0.2758446	4.1632	0.00003138	0.1911641
Longitude	2031022	0.00199	-0.2092438	-2.7739	0.005539	-0.1273734
	Elevation (n = 216), Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for 63 points					
Latitude	1208391	0.00002871	0.2805397	4.377	0.00001203	0.2009557
Longitude	1967936	0.01149	-0.1716836	-2.1965	0.02806	-0.1008476
	Elevation (n = 216), Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points					
Latitude	1175178	0.000007048	0.3003145	4.4241	0.000009684	0.2031179
Longitude	1987323	0.006931	-0.1832264	-2.4621	0.01381	-0.1130427
	Elevation (n = 216), Scenario 4 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points					
Latitude	1155782	0.000002955	0.3118626	4.3864	0.00001152	0.2013881
Longitude	1974993	0.009592	-0.1758851	-2.3566	0.01844	-0.1081993

5) Using the Vincenty Ellipsoid method, the mean shortest (straight line) difference in distances between corresponding pairs of points was found to be 7855.95 m (Median = 4989.53 m, St dev (pop) = 9475.03 m, Range = 0.298 - 38809.5 m). The correlation between elevations under different scenarios have shown that there is a linear decline in correlation between coordinates with fixed or no buffers versus coordinates that had changing buffers, both individually and when mixed with coordinates of fixed buffer values (Table 6).

Table 6:

Correlations of elevations with and without buffer corrections.

	Test statistic (t)	Df	p-value	95% CI	Sample estimate (PPMC)
Elevation, Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for 63 points					
Elevation without buffer correction (n=216)	37.24	214	< 2.2e-16	0.9103866 - 0.9466366	0.9307636
Elevation without buffer correction (n=63)	Inf	61	< 2.2e-16	1 - 1	1
Elevation, Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for 63 points					
Elevation without buffer correction (n=216)	28.828	214	< 2.2e-16	0.8607228 - 0.9161812	0.8917529
Elevation without buffer correction (n=63)	12.39	61	< 2.2e-16	0.7568081 - 0.9041972	0.8459488
Elevation, Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points					
Elevation without buffer correction (n=216)	27.126	214	< 2.2e-16	0.8460870 - 0.9070821	0.8801679
Elevation without buffer correction (n=63)	10.344	61	4.82e-15	0.68592 - 0.87317	0.7980576
Elevation, Scenario 4 = with 0 Km coordinate uncertainty changed to 10 Km for 63 points					
Elevation without buffer correction (n=216)	24.521	214	< 2.2e-16	0.8192044 - 0.8902195	0.8587826
Elevation without buffer correction (n=63)	7.9218	61	6.00e-11	0.5638130 - 0.8159125	0.7121022

Using one-way ANOVA, it was found that under two of the aforementioned conditions ($n = 216$ & $n = 63$), there was no significant difference in means among the four scenarios of elevation ($p = 0.969$ & $p = 0.797$ respectively). The results are visualised in Fig. 6 and enumerated in Table S3. In pairwise comparison of four scenarios of elevations, the value when $n = 216$, was 0.96 for all, and value when $n = 63$, was 0.92 for all. Bartlett's test indicated that the variances were homogeneous in both cases ($n = 216$, Bartlett's K-squared = 0.37976, $df = 3$, $p\text{-value} = 0.9444$ & $n = 63$, Bartlett's K-squared = 1.0746, $df = 3$, $p\text{-value} = 0.7832$). There was also no discernible difference of means between latitudes and longitudes from the two groups (see Fig. S7).

Figure 6: Comparison of spread of elevations under different scenarios. The four scenarios here are Elev 0, Elev 0.5, Elev 5, Elev 10 (with 0 km coordinate uncertainty kept as zero or changed to 0.5, 5, and 10 km respectively, for 63 points). (A) $n = 216$. (B) $n = 63$ (all points with varying buffer).

Discussion

The sightings at Bilaspur district are the first records of the species/subspecies in this district, thereby filling a void in distribution along the western end of their range. They constitute the bulk of records from Himachal Pradesh (with only three pre-existing records from the state). If the westward expansion of the subspecies continues, sightings from the Hindukush mountain range of northern Pakistan and north-eastern Afghanistan can be expected, after which they will not be able to survive due to the presence of dry deserts all around. It is unlikely that they will climb up the Himalayan range to higher elevations, as they prefer lower altitudes (this study provides partial evidence of this species' resilience to climate change). Nor are they expected to cross the Himalayan peaks, which acts as tall barriers for dispersal, a prediction affirmed by the absence of records from the northern side of the Himalayan range.

Studies using c-science data (primarily or along with other data) have found climate-induced expansions to (as well as retreat from) the northern latitudes in areas such as North America (Breed, Stichter & Crone, 2012) and northern Europe (Pöyry et al, 2009, Hällfors et al, 2021). The current study doesn't provide any evidence for such a range shift but an in-depth study into the population biology could reveal more about this phenomenon. It could be that the numbers of this subspecies are growing in the northern portion of their range compared to stable populations in the southern portion.

In terms of seasonal variation, no significant linear patterns were observed, although Spearman's correlation between longitude and months showed a moderate negative relationship. If this were to be a true correlation, then it would imply that as we move east, the activity of species is expected to peak earlier in the year. Histogram of months (Fig. 4C) shows peak activity in March and August-December, indicating annual bimodal distribution, with more activity in spring/post-winter, monsoon/pre-winter and winter/post-monsoon. This

pattern may vary across their range. Due to the data limitations in some parts of their range, and the absence of long-term data, it is difficult to understand if the species has advanced their phenology as a response to climatic changes, as has been observed for a large number of butterflies in Finland (Hällfors et al, 2021), particularly relevant to less mobile or movement-restricted species (Amano et al, 2014, Socolar et al, 2017) like *V. e.sinha*.

When looking at seasonal patterns, one must look at the detection and non-detection data. Boruah et al (2018) reports non-detection during summer and monsoon months (March-October) and detection only during winter (November-February) in the Manchabandha Reserve Forest, Odisha. Similarly, non-detection was noted in the Barail Wildlife Sanctuary (Assam), for the months April-November and detection from December-March (Gogoi, Singha & Deb, 2016). In Tipong (eastern Assam), they were detected in September and not detected during the other months surveyed, representative of most seasons (Singh, 2017). In Pakke and Sessa in Arunachal Pradesh, they have been detected during May and September-November, and not detected during the other months surveyed (Sondhi & Kunte, 2016). Non-detection in summer and detection in winter, at the Baghmara Reserve Forest (Meghalaya), is reported by Kunte et al (2012). In the Santosh river catchment (Bhutan), non-detection was in March, May and July, whereas detection was in November and January. In Uttarakhand too, a similar pattern is seen — detection in March, and September-December, and non-detection during the rest of the year (Singh & Sondhi, 2016; Sondhi, 2021, pers. obs.). Smetacek (2012) reports activity in May and July-October (and hence bins them under the multivoltine category which is used to refer to species that usually have multiple overlapping generations). This is also a little different from the current study (August-November). From all this, it can be inferred that they are more active during the winter months, at least in and near the North-East Indian cluster and the Western Himalayan cluster.

Butterfly species, especially mountain butterflies have been reported to climb up to the summits over the years as a result of climate change (Molina-Martínez et al, 2016; Rödter et al, 2021). No patterns associated with elevation were observed for *V. e. sinha*. This means that the species has not climbed up over the years as a result of climate change. Molina-Martínez et al (2016) reported that the species occurring below the elevation of 1000 m, as well as all the species combined, had significantly less range shifts towards higher elevations than what was expected based only on global warming. Our results are consistent with this finding because *V. e. sinha* occurs mostly below 1000 m. This could either be as a result of their resilience as a species or their annual life cycle where they are more active during the colder part of the year. The populations seems to be randomly spread across the elevational gradient across their range of distribution. This could alternatively be a complex pattern which is yet to be understood.

Mean distance between the sets of points (~7.85 km), is higher than the known daily dispersal of the species — around 1-5 km (Smetacek, 2012). This can negatively affect studies that require a high resolution of information such as studies on microhabitats of the species. This difference is unlikely to affect research such as Species Distribution Modelling (Wisz et al, 2008; Mizsei et al, 2016) because of high performance even with low spatial accuracy and lower number of points for models such as MaxEnt. As is seen in Fig. 6b, when using just the coordinates with unknown uncertainty as opposed to using them when mixed with coordinates with fixed buffers (Fig. 6a), there is a notable variation in means, which can become significant when the dataset is largely or wholly the former type of coordinates. So, metadata associated with datapoints in c-science datasets can affect the results. For this reason, biological plausibility and relevance should be considered while deciding which datum to

keep and which to exclude. Some butterflies are known to be high-flyers and others low-flyers (DeVries, Penz & Hill, 2010). The former may use canopies to move around which are parallel to the slopes in hilly areas which translates into wide variation in elevation, whereas the latter may use roads which are generally along the contours or slopes with a low gradient of elevation. So, averaging elevation using a buffer may have different effects as covariates on different species. For proper use and application of data (usually from c-science sources), one must have some idea about the behaviour and movement ecology of the species under study.

The problems associated with the collection and use of data, and potential remedies, are elucidated in Table 7.

Table 7:

Problems with, and remedies/best practices for data collection and use.

Problems	Remedies/Best Practices
Lack of reporting of coordinate uncertainty. Only iNaturalist and some museums reported this.	Always submit coordinates along with coordinate uncertainty. C-science repositories should include this data.
Difficulty extracting coordinates because: 1) Old names are given (e.g. Gogoi, Singha & Deb, 2016) 2) Names are too broad to pinpoint (e.g. Coene & Vis, 2008) 3) Coordinates are incorrectly given. It was found that even in coordinates within main tables of papers, there have been significant errors in the accuracy (e.g. Singh, 2017).	1) Always report coordinates 2) Report coordinates along with coordinate uncertainty instead of just names 3) Report coordinates in decimal degree (but carefully converted) and thoroughly review data. Authors need to verify their coordinates using visualization of their coordinates in some online map or a GIS software before submission, especially when reporting them in supplemental materials which are not reviewed by most journals. Reviewers need to do the same, especially when the coordinates are reported in main tables.
Coordinate mismatch for same record in different sources. Eg: Sharma & Sharma (2018) reports accurate coordinates of	Consistent reporting of coordinates should be done by authors/c-science contributors.

the Manwal record in J&K in their paper whereas it is ~ 3,568 m away in Butterflies of India website (because the Google map coordinates for Manwal was used as default).	
Time wasted in extracting data from papers. The data extraction alone, from published literature for this study, took around 3 weeks. This could have been reduced to a few days if they were stored in bulk somewhere. Data extraction for studies involving a large number of species would be highly impeded if this is not done.	Submit the same data in c-science repositories (ideally somewhere which is connected to data aggregators like GBIF, due to the ability of analytical and visualization packages in R to directly download and handle these datasets with ease) or give the same information in supplemental materials.
Lack of reporting of exact dates and time (e.g. Kirti, Mehra & Sidhu, 2016). Butterflies of India website currently do not collect information about time of record, which is a loss of valuable data.	Exact dates should be reported along with time, if possible.
Problems with correspondence: 1) Low response rate. The response rate for papers was merely 50% 2) Lack of corresponding email address. This is often a problem with articles published in lower quality journals, but has valuable data which can't be discarded.	1) Corresponding authors must be responsive 2) Corresponding email address must always be given and a shift towards atleast 3 authors sharing their email addresses should be encouraged, for better communication for science. Orcid or some unique author ID from where any changes in corresponding address can be tracked should ideally be given in the journal articles.

Conclusion

In the current study, it has been found that *V. e. sinha* has likely undergone a range expansion. None of the other potential spatiotemporal phenomena that were investigated, exhibited any discernible patterns. The current study has helped in adding to the knowledge on this subspecies, such as their distribution patterns and seasonal activity. It has brought more private data into the public domain (Supplemental Table S4 contain the raw and derived data

used for the research). This data can be used for some more computationally expensive analyses such as maximum entropy modelling which can address the some of the questions asked in the current study with more accuracy. It has identified knowledge gaps which other researchers could look into. It has been found that we need more c-science data, especially from before 2000. This can be achieved by writing to older researchers and helping them to digitise the data which are gathering dust in their field notes, thus adding professionally-collected data to c-science repositories. As has been noticed in this study, we need more precise data. This would require teaching the c-science contributors about the best practices in data entry. To that extent, we would need better curation and peer-review of data, whether in journals or in c-science repositories. The results from the current study come with the caveat of the obvious sampling bias, with a high amount of sampling in much of North-East India and Uttarakhand compared to other regions. It was found that the way c-science data are extracted and handled can potentially affect the results of a study. As is evident from the varying taxonomy of the genus, species, and subspecies (briefly described in the introduction), this genus needs a taxonomic revision. Studies like this help in such an endeavour by providing ecological details that may be useful in taxonomic revisions based on multiple lines of evidence, including genetics (as has been carried out for the genus *Cethosia* Fabricius belonging to the same subfamily as *Vagrans* - Heliconiinae (Müller & Beheregaray, 2010)).

Acknowledgements

We thank the Himachal Pradesh Forest Department for the permission granted to carry out our research in Bilaspur district. We acknowledge Dr. Krushnamegh Kunte for sharing the dataset of contributions to the Butterflies of India website. We thank Tshetsholo Naro, Bitupan Boruah and Dr. Monwar Hossain for providing details of records from their

respective papers. Our gratitude goes towards Dr. Peter V. Küppers, Dr. Mutum Ingobi Singh (prompted by Dr. Ramaiyer Varatharajan), and Dr. Roger Clive Kendrick, for not only giving details of records from their respective publication, but other unpublished records of the species. We would also like to thank all the remaining contributors of the records through science, published literature and other sources. We would like to thank Faezal Yunus, Dr. Shankha Banerjee, Abhishek Hariharan and Benny Malone for proofreading and bettering this manuscript.

References

- Akoglu H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18:91–93. DOI: [10.1016/j.tjem.2018.08.001](https://doi.org/10.1016/j.tjem.2018.08.001).
- Amano T, Freckleton RP, Queenborough SA, Doxford SW, Smithers RJ, Sparks TH, Sutherland WJ. 2014. Links between plant species' spatial and temporal responses to a warming climate. *Proceedings of the Royal Society B: Biological Sciences* 281:20133017. DOI: [10.1098/rspb.2013.3017](https://doi.org/10.1098/rspb.2013.3017).
- Bascombe, MJ, Johnston G, Bascombe FS.1999. *The Butterflies of Hong Kong*. London: Academic Press.
- Bhakare, M. 2009. Vagrants egista (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P. Roy (Chief Editors). *Butterflies of India*, v. 3.17. Indian Foundation for Butterflies and National Centre for Biological Sciences. Available at <http://www.ifoundbutterflies.org/sp/908/Vagrants-egista> (accessed 06 August 2021)
- Bhowmik S, Chowdhury S. 2021. Butterflies (Lepidoptera: Rhopalocera) in and around the Unakoti Archaeological Site with 59 new additions to the Butterfly Fauna of Tripura,

578 North-eastern India. *Asian Journal of Conservation Biology* 10:68–95. DOI:
579 [10.53562/ajcb.NDZF2600](https://doi.org/10.53562/ajcb.NDZF2600).

580 Bingham CT. 1905. *The fauna of British India including Ceylon and Burma — butterflies Vol*
581 *I*. London: Taylor and Francis.

582 Boruah B, Das GN, Payra A, Gogoi MJ, Dash SK, Tamuly T, Sethy J, Mishra RK, Rout SD.
583 2018. Assessment of Butterfly (Lepidoptera, Rhopalocera) Diversity in Manchabandha
584 and Budhikhamari Reserve Forest, Mayurbhanj, Odisha, India. *Asian Journal of*
585 *Conservation Biology* 7:51-65.

586 Breed GA, Stichter S, Crone EE. 2013. Climate-driven changes in northeastern US butterfly
587 communities. *Nature Climate Change* 3:142–145. DOI: [10.1038/nclimate1663](https://doi.org/10.1038/nclimate1663).

588 Chowdhury S, Braby MF, Fuller RA, Zalucki MP. 2021. Coasting along to a wider range:
589 niche conservatism in the recent range expansion of the Tawny Coster, *Acraea terpsicore*
590 (Lepidoptera: Nymphalidae). *Diversity and Distributions* 27:402–415. DOI:
591 [10.1111/ddi.13200](https://doi.org/10.1111/ddi.13200).

592 Coene HA, Vis R. 2008. Contribution to the butterfly fauna of Yunnan, China (Hesperioidea,
593 Papilionoidea). *Nota lepidopterologica* 2:231-261.

594 Crichton N. 2001. The dialysis diet and fluid non-adherence questionnaire: validity testing of
595 a self-report instrument for clinical practice INFORMATION POINT: Kendall's Tau.
596 *Journal of Clinical Nursing* 10:715. DOI: [10.1046/j.1365-2702.2001.00537.x](https://doi.org/10.1046/j.1365-2702.2001.00537.x).

597 D'Abrera, B. 1940. *Butterflies of Oriental Region. Part II: Nymphalidae, Styridae,*
598 *Amathusidae*. Singapore: Kyodo-Shing Loong.

599 Dancy CP, Reidy J. 2007. *Statistics without maths for psychology*. Harlow: Pearson
600 Education Limited.

601 Dennis EB, Morgan BJT, Brereton TM, Roy DB, Fox R. 2017. Using citizen science butterfly
602 counts to predict species population trends. *Conservation Biology* 31:1350–1361. DOI:
603 [10.1111/cobi.12956](https://doi.org/10.1111/cobi.12956).

604 DeVries PJ, Penz CM, Hill RI. 2010. Vertical distribution, flight behaviour and evolution of
605 wing morphology in *Morpho* butterflies. *Journal of Animal Ecology* 79:1077–1085. DOI:
606 [10.1111/j.1365-2656.2010.01710.x](https://doi.org/10.1111/j.1365-2656.2010.01710.x).

607 GBIF.org. 2021. GBIF Occurrence Download. Available at
608 <https://doi.org/10.15468/dl.35yvnz> (accessed 29 July 2021)

609 Gogoi MJ, Singha HJ, Deb P. 2016. Butterfly (Lepidoptera) Diversity in Barail Wildlife
610 Sanctuary, Assam, India. *Journal of Entomology and Zoology Studies* 4:547–560.

611 Gokhale P. 2021. Vagrans egista (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P.
612 Roy (Chief Editors). *Butterflies of India*, v. 3.17. Indian Foundation for Butterflies and
613 National Centre for Biological Sciences. Available at
614 <http://www.ifoundbutterflies.org/sp/908/Vagrans-egista> (accessed 06 August 2021)

615 Gokhale P, Yathumon MA. 2021. Descriptions of the early stages of Vagrans egista sinha
616 (Lepidoptera: Nymphalidae) with notes on its host plant *Xylosma longifolia* Clos from
617 the western Himalaya of India. *Journal of Threatened Taxa* 13:19144–19148. DOI:
618 [10.11609/jott.6441.13.8.19144-19148](https://doi.org/10.11609/jott.6441.13.8.19144-19148).

619 Hällfors MH, Pöyry J, Heliölä J, Kohonen I, Kuussaari M, Leinonen R, Schmucki R,
620 Sihvonen P, Saastamoinen M. 2021. Combining range and phenology shifts offers a
621 winning strategy for boreal Lepidoptera. *Ecology Letters* 24:1619–1632. DOI:
622 [10.1111/ele.13774](https://doi.org/10.1111/ele.13774).

623 Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. 2015. The Extent and
624 Consequences of P-Hacking in Science. *PLOS Biology* 13:e1002106. DOI:
625 [10.1371/journal.pbio.1002106](https://doi.org/10.1371/journal.pbio.1002106).

- 626 Jain A, Khoon KS, Gan CW, Webb EL. 2018. Butterfly extirpations, discoveries and
627 rediscoveries in Singapore over 28 years. *Raffles Bulletin of Zoology* 66:217–257.
628 [https://lkenhm.nus.edu.sg/wp-content/uploads/sites/10/app/uploads/2018/01/66-rbz217-](https://lkenhm.nus.edu.sg/wp-content/uploads/sites/10/app/uploads/2018/01/66-rbz217-257.pdf)
629 [257.pdf](https://lkenhm.nus.edu.sg/wp-content/uploads/sites/10/app/uploads/2018/01/66-rbz217-257.pdf)
- 630 Kendrick R, ed. 2005. Kadoorie Farm & Botanic Garden 2005 Fauna Checklist. Available at
631 <https://www.kfbg.org/images/download/KFBG-Fauna-Checklist-2005.pdf> (accessed 12
632 August 2021).
- 633 Khandoker F, Rashid M, Das D, Hossain M. 2013. Species diversity and abundance of
634 Butterflies in the Lawachara National Park, Bangladesh. *Jahangirnagar University*
635 *Journal of Biological Science* 2:121–127.
- 636 Kirti JS, Mehra D, Sidhu AK. 2016. Vagrans egista sinha (Kollar, 1844) (Lepidoptera:
637 Papilionoidea: Nymphalidae): range extension and an addition to the butterfly fauna of
638 Himachal Pradesh, India. *Check List* 12:1923–1923. DOI: [10.15560/12.4.1923](https://doi.org/10.15560/12.4.1923).
- 639 Kumar A, Kushwaha S, Namdev A. 2020. First record of Vagrant Vagrans egista sinha from
640 Uttar Pradesh, India. *Zoo's Print* 35:12–14.
641 <https://www.zoosprint.zooreach.org/index.php/zp/article/view/5673>.
- 642 Kunte K. 2007. Vagrans egista (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P. Roy
643 (Chief Editors). *Butterflies of India*, v. 3.17. Indian Foundation for Butterflies and
644 National Centre for Biological Sciences. Available at
645 <http://www.ifoundbutterflies.org/sp/908/Vagrans-egista> (accessed 06 August 2021).
- 646 Kunte K, Sondhi S, Sangma BM, Lovalekar R, Tokekar K, Agavekar G. 2012. Butterflies of
647 the Garo Hills of Meghalaya, northeastern India: their diversity and conservation. *Journal*
648 *of Threatened Taxa* 4:2933–2992. DOI: [10.11609/JoTT.o2945.2933-92](https://doi.org/10.11609/JoTT.o2945.2933-92).

- 649 Küppers PV, Janikorn L. 2009. Ergänzungen zu: Vorläufige Liste der von uns in der Provinz
650 Phang Nga (S-Thailand) beobachteten Tagschmetterlinge (Lepidoptera, Diurna). *Atalanta*
651 40:240 - 250. https://www.zobodat.at/pdf/Atalanta_40_0240-0250.pdf.
- 652 Küppers PV, Janikorn L. 2020. Schmetterlingsfauna der Provinz Phang Nga in Süd-
653 Thailand. Teil 2 Spezieller Teil: Nymphalidoidea II and Papilionoidea II. *Neue*
654 *Entomologische Nachrichten* 78:1-524. [https://www.zobodat.at/pdf/Neue-](https://www.zobodat.at/pdf/Neue-Entomologische-Nachrichten_78_0001-0524.pdf)
655 [Entomologische-Nachrichten_78_0001-0524.pdf](https://www.zobodat.at/pdf/Neue-Entomologische-Nachrichten_78_0001-0524.pdf).
- 656 Larsen TB. 2004. *Butterflies of Bangladesh – an annotated checklist*. Dhaka: IUCN The
657 World Conservation Union.
- 658 Lenoir J, Svenning J-C. 2013. Latitudinal and Elevational Range Shifts under Contemporary
659 Climate Change. In: Levin SA ed. *Encyclopedia of Biodiversity (Second Edition)*.
660 Waltham: Academic Press, 599–611. DOI: [10.1016/B978-0-12-384719-5.00375-0](https://doi.org/10.1016/B978-0-12-384719-5.00375-0).
- 661 Matteson KC, Taron DJ, Minor ES. 2012. Assessing Citizen Contributions to Butterfly
662 Monitoring in Two Large Cities. *Conservation Biology* 26:557–564. DOI:
663 [10.1111/j.1523-1739.2012.01825.x](https://doi.org/10.1111/j.1523-1739.2012.01825.x).
- 664 Mesaglio T, Soh A, Kurniawidjaja S, Sexton C. 2021. ‘First Known Photographs of Living
665 Specimens’: the power of iNaturalist for recording rare tropical butterflies. *Journal of*
666 *Insect Conservation* 25:905–911. DOI: [10.1007/s10841-021-00350-7](https://doi.org/10.1007/s10841-021-00350-7).
- 667 Mizsei E, Üveges B, Vági B, Szabolcs M, Lengyel S, Pfliegler WP, Nagy ZT, Tóth JP. 2016.
668 Species distribution modelling leads to the discovery of new populations of one of the
669 least known European snakes, *Vipera ursinii graeca*, in Albania. *Amphibia-Reptilia*
670 37:55-68. DOI: <https://doi.org/10.1163/15685381-00003031>.
- 671 Molina-Martínez A, León-Cortés JL, Regan HM, Lewis OT, Navarrete D, Caballero U, Luis-
672 Martínez A. 2016. Changes in butterfly distributions and species assemblages on a

Neotropical mountain range in response to global warming and anthropogenic land use.

Diversity and Distributions 22:1085–1098. DOI: [10.1111/ddi.12473](https://doi.org/10.1111/ddi.12473).

Müller CJ, Beheregaray LB. 2010. Palaeo island-affinities revisited – Biogeography and

systematics of the Indo-Pacific genus *Cethosia* Fabricius (Lepidoptera: Nymphalidae).

Molecular Phylogenetics and Evolution 57:314–326. DOI: [10.1016/j.ympev.2010.07.002](https://doi.org/10.1016/j.ympev.2010.07.002).

Naidu R. 2018. *Vagrans egista* (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P. Roy

(Chief Editors). *Butterflies of India*, v. 3.17. Indian Foundation for Butterflies and

National Centre for Biological Sciences. Available at

<http://www.ifoundbutterflies.org/sp/908/Vagrans-egista> (accessed 06 August 2021)

Naro T, Sondhi S. 2014. Butterflies (Lepidoptera) of Chizami, Phek District, Nagaland, India.

Journal of Threatened Taxa 6:6593–6634. DOI: [10.11609/JoTT.o3995.6593-634](https://doi.org/10.11609/JoTT.o3995.6593-634).

Nie NH, Bent DH, Hull CH. 1975. *SPSS: Statistical package for the social sciences*. New

York: McGraw-Hill.

Nuzzo R. 2014. P values, the ‘gold standard’ of statistical validity, are not as reliable as many

scientists assume. *Nature* 506:150–152.

Parsons RE, Cantlie K. 1948. The Butterflies of the Khasia and Jaintia Hills, Assam. *The*

journal of the Bombay Natural History Society 47:498–522.

Pöyry J, Luoto M, Heikkinen RK, Kuussaari M, Saarinen K. 2009. Species traits explain

recent range shifts of Finnish butterflies. *Global Change Biology* 15:732–743. DOI:

[10.1111/j.1365-2486.2008.01789.x](https://doi.org/10.1111/j.1365-2486.2008.01789.x).

Rödger D, Schmitt T, Gros P, Ulrich W, Habel JC. 2021. Climate change drives mountain

butterflies towards the summits. *Scientific Reports* 11:14382. DOI: [10.1038/s41598-021-](https://doi.org/10.1038/s41598-021-93826-0)

[93826-0](https://doi.org/10.1038/s41598-021-93826-0).

696 Ruxton GD, Wilkinson DM, Neuhauser M. 2015. Advice on testing the null hypothesis that a
697 sample is drawn from a normal distribution. *Animal Behaviour* 107:249–252. DOI:
698 [10.1016/j.anbehav.2015.07.006](https://doi.org/10.1016/j.anbehav.2015.07.006).

699 Sanderson C, Braby MF, Bond S. 2021. Butterflies Australia: a national citizen science
700 database for monitoring changes in the distribution and abundance of Australian
701 butterflies. *Austral Entomology* 60:111–127. DOI: [10.1111/aen.12513](https://doi.org/10.1111/aen.12513).

702 Saragada A. Vagrans egista (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P. Roy
703 (Chief Editors). Butterflies of India, v. 3.17. Indian Foundation for Butterflies and
704 National Centre for Biological Sciences. Available at
705 <http://www.ifoundbutterflies.org/sp/908/Vagrans-egista> (accessed 06 August 2021).

706 Savela M. nd. Vagrans Hemming, 1934. Available at
707 http://ftp.funet.fi/index/Tree_of_life/insecta/lepidoptera/ditrysia/papilionoidea/nymphalid
708 [ae/heliconiinae/vagrans/](http://ftp.funet.fi/index/Tree_of_life/insecta/lepidoptera/ditrysia/papilionoidea/nymphalid/ae/heliconiinae/vagrans/) (accessed 06 August 2021).

709 Schmeller DS, Henry P-Y, Julliard R, Gruber B, Clobert J, Dziock F, Lengyel S, Nowicki P,
710 Déri E, Budrys E, Kull T, Tali K, Bauch B, Settele J, Van Swaay C, Kobler A, Babij V,
711 Papastergiadou E, Henle K. 2009. Advantages of volunteer-based biodiversity monitoring
712 in Europe. *Conservation Biology* 23:307–316. DOI: [10.1111/j.1523-1739.2008.01125.x](https://doi.org/10.1111/j.1523-1739.2008.01125.x).

713 Schober P, Boer C, Schwarte LA. 2018. Correlation Coefficients: Appropriate Use and
714 Interpretation. *Anesthesia & Analgesia* 126:1763–1768. DOI:
715 [10.1213/ANE.0000000000002864](https://doi.org/10.1213/ANE.0000000000002864).

716 Sharma S, Sharma N. 2018. New nymphalid butterfly records from Jammu & Kashmir, India.
717 *Journal of Threatened Taxa* 10:12602–12606. DOI: [10.11609/jott.3874.10.11.12602-](https://doi.org/10.11609/jott.3874.10.11.12602-12606)
718 [12606](https://doi.org/10.11609/jott.3874.10.11.12602-12606).

719 Singh AP. 2012. Lowland forest butterflies of the Sankosh River catchment, Bhutan. *Journal*
720 *of Threatened Taxa* 4:3085–3102. DOI: [10.11609/JoTT.o2625.3085-102](https://doi.org/10.11609/JoTT.o2625.3085-102).

- 721 Singh AP. 2017. Butterflies of eastern Assam, India. *Journal of Threatened Taxa* 9:10396-
722 10420. DOI: [10.11609/jott.3177.9.7.10396-10420](https://doi.org/10.11609/jott.3177.9.7.10396-10420).
- 723 Singh AP, Sondhi S. 2016. Butterflies of Garhwal, Uttarakhand, western Himalaya, India.
724 *Journal of Threatened Taxa* 8:8666–8697. DOI: [10.11609/jott.2254.8.4.8666-8697](https://doi.org/10.11609/jott.2254.8.4.8666-8697).
- 725 Singh MI, Gupta A, Varatharajan R. 2011. Butterfly fauna of the Keibul Lamjao National
726 Park, Manipur, North East India. *Current Science* 101:719-720
- 727 Smetacek P. 2012. Butterflies (Lepidoptera: Papilionoidea and Hesperoidea) and other
728 protected fauna of Jones Estate, a dying watershed in the Kumaon Himalaya, Uttarakhand,
729 India. *Journal of Threatened Taxa* 4:2857–2874. DOI: [10.11609/JoTT.o3020.2857-74](https://doi.org/10.11609/JoTT.o3020.2857-74).
- 730 van Strien AJ, van Swaay CAM, Termaat T. 2013. Opportunistic citizen science data of
731 animal species produce reliable estimates of distribution trends if analysed with
732 occupancy models. *Journal of Applied Ecology* 50:1450–1458. DOI: [10.1111/1365-
733 2664.12158](https://doi.org/10.1111/1365-2664.12158).
- 734 Socolar JB, Epanchin PN, Beissinger SR, Tingley MW. 2017. Phenological shifts conserve
735 thermal niches in North American birds and reshape expectations for climate-driven
736 range shifts. *Proceedings of the National Academy of Sciences* 114:12976–12981. DOI:
737 [10.1073/pnas.1705897114](https://doi.org/10.1073/pnas.1705897114).
- 738 Soman, A. 2015. Vagrans egista (Cramer, [1780]) – Vagrant. Kunte, K., S. Sondhi, and P.
739 Roy (Chief Editors). Butterflies of India, v. 3.17. Indian Foundation for Butterflies and
740 National Centre for Biological Sciences. Available at
741 <http://www.ifoundbutterflies.org/sp/908/Vagrans-egista> (accessed 06 August 2021)
- 742 Sondhi S, Kunte K. 2016. Butterflies (Lepidoptera) of the Kameng Protected Area Complex,
743 western Arunachal Pradesh, India. *Journal of Threatened Taxa* 8:9053–9124. DOI:
744 [10.11609/jott.2984.8.8.9053-9124](https://doi.org/10.11609/jott.2984.8.8.9053-9124).

745 Sondhi S, Kunte K. 2018. *Butterflies of Uttarakhand: A Field Guide*. M/s Bishen Singh
746 Mahendra Pal Singh (Dehradun), Titli Trust (Dehradun), National Centre for Biological
747 Sciences (Bengaluru) & Indian Foundation of Butterflies (Bengaluru).

748 Spitzer K, Jaroš J, Havelka J, Lepš J. 1997. Effect of small-scale disturbance on butterfly
749 communities of an Indochinese montane rainforest. *Biological Conservation* 80:9–15.
750 DOI: [10.1016/S0006-3207\(96\)00079-1](https://doi.org/10.1016/S0006-3207(96)00079-1).

751 Spitzer K, Novotný V, Tonner M, Lepš J. 1993. Habitat Preferences, Distribution and
752 Seasonality of the Butterflies (Lepidoptera, Papilionoidea) in a Montane Tropical Rain
753 Forest, Vietnam. *Journal of Biogeography* 20:109–121. DOI: [10.2307/2845744](https://doi.org/10.2307/2845744).

754 Tytler HC. 1911. Notes on butterflies from the Naga Hills - Part II. *Journal of Bombay*
755 *Natural History Society* 21:588–606.

756 Vantieghem P, Maes D, Kaiser A, Merckx T. 2017. Quality of citizen science data and its
757 consequences for the conservation of skipper butterflies (Hesperiidae) in Flanders
758 (northern Belgium). *Journal of Insect Conservation* 21:451–463. DOI: [10.1007/s10841-](https://doi.org/10.1007/s10841-016-9924-4)
759 [016-9924-4](https://doi.org/10.1007/s10841-016-9924-4).

760 Varshney RK. 1994. Index Rhopalocera Indica Part III. Genera of butterflies from India and
761 neighbouring countries [Lepidoptera: (B) Satyridae, Nymphalidae, Libytheidae and
762 Riodinidae]. *Oriental Insects* 28:151–198. DOI: [10.1080/00305316.1994.10432303](https://doi.org/10.1080/00305316.1994.10432303).

763 Wisz MS., Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting
764 Species Distributions Working Group. 2008. Effects of sample size on the performance
765 of species distribution models. *Diversity and distributions* 14:763–773. DOI:
766 <https://doi.org/10.1111/j.1472-4642.2008.00482.x>.

767 Wynter-Blyth, MA. 1957. *Butterflies of the Indian region*. Mumbai: Bombay Natural History
768 Society.

769 Zothansangi. 2015. Molecular Characterization and Phylogeny of Heliconiinae Butterflies
770 (Nymphalidae: Lepidoptera) in Northeast India Butterflies. D. Phil. Thesis, Mizoram
771 University.

772

773 **Supplemental Information**

Supplemental Figure S1 - Contribution of records by sources.

Supplemental Figure S2: q-q plots for dates, latitudes, and longitudes. (a) Dates ($n = 161$, Scenario 1 = only actual dates), (b) Dates ($n = 198$, Scenario 2 = actual dates + median dates + adjusted median dates), (c) Latitudes ($n = 161$), (d) Latitudes ($n = 198$), (e) Latitudes ($n = 216$), (f) Longitudes ($n = 161$), (g) Longitudes ($n = 198$), and (h) Longitudes ($n = 216$). Please note that only two or three examples are given for each here, but q-q plots were actually generated prior to each correlations. This is justified for the following reasons - i) Once it's established that a set with highest number of observations is non-normal (e.g. longitudes with 216 observations), it can be inferred that normality tends to be non-normal for a smaller subset of it (e.g. longitudes with 33 observations), as there is more variability ii) Non-parametric tests were carried out even if one in a pair of variables is found to be non-normal.

Supplemental Figure S4: q-q plots for months and elevations. (a) Months ($n = 208$, Scenario 1 = actual months + median months + adjusted median months), (b) Months ($n = 195$, Scenario 2 = actual months + median months), (c) Months ($n = 181$, Scenario 3 = actual months + adjusted median months), (d) Months ($n = 168$, Scenario 4 = actual months), (e) Months ($n = 41$, Scenario 5 = median months + adjusted median months), (f) Elevations ($n = 216$, Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for 63 points), (g) Elevations ($n = 216$, Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for 63 points), and (h) Elevations ($n = 216$, Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points).

778

779

Supplemental Figure S6 - Buffers around coordinate points in North-East India cluster over hillshade map.

Supplemental Figure S7 - Comparison of spread of coordinates amongst two groups - derived from names and acquired through request. (A) Spread of longitudes. (B) Spread of latitudes.

Supplemental Table S1:

Sources of records for each country.

Country	Sources
India	Parsons & Cantlie, 1948; Ramakrishna et al, 2006; Singh, Gupta & Varatharajan, 2011+Singh, 2021, pers comm.; Smetacek, 2012; Kunte et al, 2012; Gogoi, 2012; Zothansangi et al, 2013; T. Sengupta, 2013, wikimedia; Bora & Meitei, 2014; Naro & Sondhi, 2014+Naro, 2021, pers. comm.; Zothansangi, 2015; Singh & Sondhi, 2016+S Sondhi, 2021, pers. obs.; Payra et al, 2016; Gogoi, Singha & Deb, 2016; Kirti, Mehra & Sidhu, 2016; Sondhi & Kunte, 2016+S Sondhi, 2021, pers. obs.; Singh, 2017; Boruah et al, 2018+B Boruah, 2021, pers. comm.; Sharma & Sharma, 2018; Irungbam et al, 2020; Kumar, Kushwaha & Namdev, 2020; P Pop, 2020, pers. obs.; Bhowmik & Chowdhury, 2021; GBIF, 2021; Gokhale, 2021; S Sondhi, 2021, pers. obs.
Nepal	GBIF, 2021; Gokhale, 2021
Bhutan	Singh, 2012
Bangladesh	Larsen, 2004; Khandokar, et al, 2013+M Hossain, 2021, pers. comm.
China	Bascombe, Johnston & Bascombe, 1999; Huang & Xue, 2004; Kendrick (Ed), 2005+Kendrick, 2021, pers. comm; Coene & Vis, 2008; GBIF, 2021
Vietnam	Spitzer et al, 1993, 1997; GBIF, 2021
Thailand	Smithsonian Tropical Research Institute, 2014; Küppers & Janikorn, 2020+P Küppers, 2021, pers. comm.; GBIF, 2021; P Küppers, 2021, pers. comm.

References (other than the ones given in the main manuscript)

- Bora A, Meitei LR. 2014. Butterfly fauna (order: Lepidoptera) in five major tea gardens of Sivasagar District, Assam, India. *Biological Forum – An International Journal* 6:7–15.
- Gogoi MJ. 2012. Butterflies (Lepidoptera) of Dibang Valley, Mishmi Hills, Arunachal Pradesh, India. *Journal of Threatened Taxa* 4:3137–3160. DOI: [10.11609/JoTT.o2975.3137-60](https://doi.org/10.11609/JoTT.o2975.3137-60).
- Huang H, Xue YP. 2004. A Contribution to the Butterfly Fauna of Southern Yunnan (Lepidoptera, Rhopalocera). *Neue Entomologische Nachrichten* 57: 135-154.
- Irungbam JS, Meitei LR, Huidrom H, Soibam BS, Ngangom A, Ngangom B, Meitei R, Fric ZF. 2020. An inventory of the butterflies of Manipur, India (insecta: lepidoptera). *Zootaxa* 4882:1–91. DOI: [10.11646/zootaxa.4882.1.1](https://doi.org/10.11646/zootaxa.4882.1.1).
- Payra A, Das GN, Boruah B, Dash SK, Das UP, Sethy J. 2016. Butterfly Diversity in Two Selected Fringe Area of Similipal Biosphere Reserve, Odisha, India, With Notes on Some Important Sightings. *Journal of Wildlife Research* 4:17-25.
- Ramakrishna, Siddiqui SZ, Sethy P, Dash S. 2006. *Faunal Resources of Similipal Biosphere Reserve, Mayurbhanj, Orissa*. Kolkata: Zoological Survey of India.
- Sengupta T. 2013. Open wing position of Vagrans egista Cramer, 1780 – Vagrant. https://commons.wikimedia.org/wiki/File:Open_wing_position_of_Vagrans_egista_Cramer,1780_%E2%80%93_Vagrant.jpg (accessed 06 August 2021).
- Smithsonian Tropical Research Institute. 2014. Public Data Portal - Record List – Vagrans[tax]. [http://www.boldsystems.org/index.php/Public_SearchTerms?query=Vagrans\[tax\]](http://www.boldsystems.org/index.php/Public_SearchTerms?query=Vagrans[tax]) (accessed 06 August 2021).
- Zothansangi, Ruati CVL, Kumar SN, Lalremsanga HT, Gurusubramanian G. 2013. Diversity and vertical distributions of Heliconiinae butterflies in Mizoram. In: Singh KK, Das, KC, Lalruatsanga H, *Bioresources and traditional knowledge of Northeast India*. Aizawl: Mizoram University, 229–235.

Supplemental Table S2 - Shapiro-Wilk values of variables		
Variable	W	p-value
Date (n = 161, Scenario 1 = only actual dates)	0.79526	9.701E-14
Date (n = 198, Scenario 2 = actual dates + median dates + adjusted median dates)	0.76713	< 2.2E-16
Latitude (n = 161, corresponding to Dates - Scenario 1)	0.80722	2.679E-13
Latitude (n = 198, corresponding to Dates - Scenario 2)	0.84975	4.969E-13
Latitude (n = 216)	0.86059	3.891E-13

Longitude (n = 161, corresponding to Dates - Scenario 1)	0.92683	0.00000027
Longitude (n = 198, corresponding to Dates - Scenario 2)	0.92254	1.025E-08
Longitude (n = 216)	0.92263	3.222E-09
Year (n = 198, Scenario 1 = actual years + median years + adjusted median years)	0.76388	< 2.2E-16
Year (n = 195, Scenario 2 = actual years + median years)	0.75486	< 2.2E-16
Year (n = 168, Scenario 3 = actual years + adjusted median years)	0.76362	3.616E-15
Year (n = 165, Scenario 4 = actual years)	0.75303	2.297E-15
Year (n = 33, Scenario 5 = median years + adjusted median years)	0.63378	7.453E-08
Year (n = 190, Scenario 6 = Scenario 1 - pre-2000 records)	0.94735	0.000001851
Year (n = 177, Scenario 7 = Scenario 1 - pre-2005 records)	0.95856	0.00004373
Month (n = 208, Scenario 1 = actual months + median months + adjusted median months)	0.90442	2.732E-10
Month (n = 195, Scenario 2 = actual months + median months)	0.90965	1.547E-09
Month (n = 181, Scenario 3 = actual months + adjusted median months)	0.89125	3.196E-10
Month (n = 168, Scenario 4 = actual months)	0.89414	1.354E-09
Month (n = 41, Scenario 5 = median months + adjusted median months)	0.84783	0.0000663
Elevation (n = 216, Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for 63 points)	0.92148	2.625E-09
Elevation (n = 216, Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for 63 points)	0.91845	1.549E-09
Elevation (n = 216, Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for 63 points)	0.92817	8.846E-09
Elevation (n = 216, Scenario 4 = with 0 Km coordinate uncertainty changed to 10 Km for 63 points)	0.9345	2.975E-08
Elevation (n = 63, Scenario 1 = with 0 Km coordinate uncertainty values kept as zero for all points)	0.81991	2.641E-07
Elevation (n = 63, Scenario 2 = with 0 Km coordinate uncertainty changed to 0.5 Km for all points)	0.79862	7.311E-08
Elevation (n = 63, Scenario 3 = with 0 Km coordinate uncertainty changed to 5 Km for all points)	0.86418	0.000005268
Elevation (n = 63, Scenario 4 = with 0 Km coordinate uncertainty changed to 10 Km for all points)	0.8964	0.00006658

810

811 Supplemental Table S3:

812 One-way ANOVA for testing difference in means between elevations under different

813 scenarios.

814

	Df		Sum Sq		Mean Sq		F value		p(>F)	
	n=216	n=63	n=216	n=63	n=216	n=63	n=216	n=63	n=216	n=63
as.factor(Group)	3	3	54226	185917	18075	61972	0.084	0.339	0.969	0.797
Residuals	860	248	185795797	45381125	216042	182988				

815

816 Supplemental Table S4:

817 Raw and derived data

818 Mendeley Data, V1, doi: 10.17632/mpn65fcnwb.1 (All other supplementary materials
819 included in the text can also be found here)