Aculeate Hymenoptera Red List - iRecord data assessment

1. Introduction

This document details an assessment and comparison of the BWARS and iRecord Aculeate Hymenopteran datasets, with a view to incorporating iRecord data into a regional Red List assessment.

1.1. Why check data so carefully?

Due to the assessment processes of the Red List, combined with the manner in which data are collected, great emphasis is placed on maintaining a very high quality of data. There are a great many taxa within the area of interest which are under-recorded. The under-recording typically results from a lack of people with the requisite time and enthusiasm to devote years to honing identification skills to the level required to be proficient enough to reach an acceptably low level of error. As such, the small volume of high quality data being output by the skilled contingent is at high risk of being drowned out by large volumes of lower-skilled collectors. The consequences for data-drowning are severe and, in many cases, immediately impactful. Since the Red List protocol utilises a form of occupancy for the majority of Aculeate Hymenoptera assessments, small volumes of incorrect data can have vastly disproportionate impacts upon Red List criteria.

The BWARS (Bees Wasps and Ants) data has gone through a significant quality control process, including:

- Translation to safe nomenclature storage (Understandings)
- Human bulk-volume examination of data
- Personal assessment of contributors' ability to identify taxa
- Automated temporal range checking
- Automated spatial range checking
- Directed outlier verification
- Ongoing querying and resolution of potential errors in individual records

This project aims to bring as much of that quality control process to the iRecord data set. Doing so will advance our knowledge on:

- How complete the BWARS dataset is
- How complete the iRecord dataset is
- What iRecord data could be used for in future Red List Aculeate Hymenoptera assessments
- Areas for improvement in both BWARS and iRecord

1.2. What is contained

Contained within is the documentation and results for the various assessments and comparisons of the iRecord and BWARS datasets. At the end is a conclusion, where the project findings are summarised for presentation to the Red List core team. The core team, being subject experts, are then given the responsibility of deciding what data, if any, should be included for a parallel Red List assessment.

2. Data Breakdown

Base-level information relating to the iRecord dataset, as well as simple comparisons between the iRecord and BWARS dataset are contained in this section. In short, this section examines:

- Importing iRecord data to the BWARS system
- Taxonomic composition of each dataset
- De-duplication of data
- A breakdown of datasets based on various criteria

2.1. Importing iRecord data

This section examines the number of records in the incoming iRecord dataset, the number of records which failed to pass the Checker, as well as their reasons for failure.

2.1.1. Pre-Checker

How many records were pushed to the Checker

Included	Records	Percentage
Total	437,553	100%
Yes	399,105	91.2%
No	38,448	8.8%

2.1.2. Checker pass rate

How many records passed the Checker

Pass	Records	Percentage
All	399,105	100%
Yes	341,687	85.6%
No	57,418	14.4%

2.1.3. Checker failure reason

> Records may fail on multiple aspects. Totals may not match 100%

By far the most common cause of failure is the inability to assign a binomial to an Understanding (14.1%). This failure is directly equivalent to an inability to confidently assign the record to a Type. Inability to resolve a binomial to type is only resolvable through specific, highly skilled, human intervention - typically reserved for taxonomists with a specific goal in mind.

Reason	Records	Percentage
Multiple potential Understandings	56,302	14.1%
No Gridref/VC match found	131	0.0%
Species has no match	236	0.1%

Reason	Records	Percentage
Spatial resolution too low	154	0.0%
Subgrid present	785	0.2%

2.2. Simple Unique Record

A common problem in biological recording is the re-entry of data, leading to multiple copies of the same record. `Simple unique record` is a data processing stage designed to condense such records by removing duplication based on the attributes of What, Where, and When. Simple Unique Record has been used in multiple major research projects and is now considered the default export from the BWARS database.

The level of self-replication contained within a dataset is measured by 'compression' - the amount of records 'compressed' by the simple unique record procedure. A high compression indicates a high level of replication, which is undesirable. The low replication level in iRecord indicates that the scheme is functioning primarily as a primary source of data. Meanwhile BWARS is very firmly acting as a tertiary source.

Given the nature of the Red List investigation, converting the iRecord data to full simple unique record format would be an extraordinary amount of work. Collisions (records which match on What/Where/When) would need to be individually examined and differences in other attributes resolved to identify if these truly are the same record. Given the low (3.5%) rate of collision in the iRecord dataset, it is reasonable to compare BWARS's simple unique record data directly to iRecord data in terms of data volume.

Scheme	Raw records	Simple unique record	Compression
BWARS	1,135,135	834,568	26.5%
iRecord	341,687	329,796	3.5%

2.3. iRecord verification status

2.3.1. Verification status 1

Verification status 1 is of no use to the Red List project as it does not provide sufficient information to use a record in a scientific context.

2.3.2. Verification status 2

Verification status 2 contains the most precise categorisation of the two available statuses. Statuses in verification status 2 are not sorted by reliability. A `Correct` record is *not* less trustworthy than a `Considered correct` record. The iRecord verification statuses under verification status 2, as well as a short explanation, are:

Status	Explanation
Correct	The verifier is able to confirm that the species has been identified correctly, usually on the basis of photo/s within iRecord (or specimen/s outside iRecord)

Status	Explanation
Considered correct	The verifier has not seen photo/s or specimen/s but has a high degree of confidence that the record is likely to be correct, based on difficulty of ID, date, location and recorder skills/experience etc.
Unable to verify	The verifier has a high degree of confidence that the record is likely to be incorrect based on difficulty of ID, date, location and recorder skills/experience (and where no photo/s or specimen/s are available); or photos are available but do not show enough detail to confirm the identification; and/or the record is not sufficiently well documented to confirm (e.g. location is vague)
Incorrect	The verifier is able to confirm that the species has not been identified correctly, or the record is erroneous in other respects, on the basis of photo/s or specimen/s, or on information from the recorder
Plausible	The record is plausible based on species, date and location, but there is not enough supporting evidence for the possibility of misidentification to be ruled out, and the record remains within the "Unconfirmed" category
Not reviewed	The record is in the system but has either not been looked at, or a verification decision not yet been reached (all records start off in this category)

2.3.3. Verification status 2 breakdown

The low number of `Incorrect` records is due to the difficulty in making a definitive 'no' answer. Records are far more likely to fall into 'plausible' category as accurately identifying incorrect data points, especially in the absence of critical information, is very hard to do. Verifiers are also reluctant to assign data to `Incorrect` through fear of discouraging newcomers.

Records typically pass through one of two streams. The first stream is where the verifier trusts the originator of a dataset to be accurate. The dataset will then be assigned `Considered correct`. The second stream is where a verifier examines and assesses individual records one-by-one. This scenario is the only one whereby a record can reach `Correct`. There is no way to know whether a rejected record (`Plausible` or `Unable to verify`) was sourced from bulk-processing or individual examination.

From examination of the ratio of `Considered correct` to `Plausible` and `Unable to verify` data, it seems that verifiers do not attempt to verify non-photographic data unless they have prior connection to the originator which they can use as a basis for assessment. As such, records marked as `Plausible`, `Unable to verify`, and `Incorrect` stem from the assessment of individual records.

Status	Records	Percentage
Correct	102,835	25.8%
Considered correct	124,711	31.2%
Plausible	6,435	1.6%
Unable to verify	2,275	0.6%
Incorrect	598	0.1%

Status	Records	Percentage
Not reviewed	103,316	25.9%
Status not used	58,935	14.8%

2.4. Threshold for use

The criteria for acceptance into the BWARS database is that `a record must have been identified by a person known to be capable of identifying that taxa`. The current import stream for BWARS relies on a verifier approaching BWARS to donate their data, upon which experienced entomologists will examine the incoming data to look for potential problems or errors. The submitter may be asked to provide specimens or vouchers for particular records, especially those with significant potential for confusion. The experienced entomologist will then run through the key and specimen alongside the submitter, as well as provide assistance on other specimens as requested. In this way BWARS maintains high data quality whilst also providing a way to provide targeted help and tutoring to entomologists.

In the absence of checking individual determiner names, the only way to extract such information from the iRecord dataset is to use the verification status `Correct`. `Correct` is the only status which requires that the verifier has seen the individual record in question, and agrees that the record is valid. There will still be some level of error, but the level should, if all is proceeding correctly, be comparable to the level of accuracy found in BWARS.

2.4.1. Correct: simple unique record

The compression rate for simple unique records on the subset of iRecord data is still low enough to not be a concern in forming comparisons to the BWARS database.

Raw records	Simple unique records	Compression
91,207	88,990	2.4%

2.5. Scheme overlap

There are 52,441 (15.9% of iRecord) records which are directly found in both the BWARS database and iRecord. The 10 most common Understandings between the two are:

Understanding	Records:
Bombus pascuorum: iso. Cameron et al: 2007	4,824
Colletes hederae: iso. Schmidt & Westrich: 1993	3,780
Bombus lapidarius: iso. Cameron et al: 2007	3,133
Bombus hypnorum: iso. Cameron et al: 2007	2,255
Bombus pratorum: iso. Cameron et al: 2007	2,062
Bombus hortorum: iso. Cameron et al: 2007	1,284
Vespula vulgaris: iso. Archer: 1989	976

Understanding	Records:
Anthophora plumipes: iso. Amiet et al: 2007	973
Andrena haemorrhoa: iso. Amiet et al: 2010	927
Vespa crabro: iso. Archer: 1989	828

2.5.1. iRecord and BWARS userbase comparison

BWARS data is almost exclusively from people who have spent significant time and effort to develop their identification skills. Where the data is not from a highly capable source, it has invariably passed through such a source before it will be accepted. Of these sources, a large amount of the records from from users who perform contracted surveys. Samples of which include development work and nature reserve assessments.

Accuracy in aculeate hymenoptera identification is an *advanced* skill, as one of the requirements for accuracy is understanding the limitations of your knowledge. Such a level of skill development does not happen without significant time investment, which typically manifests as also producing large volumes of data. In turn, these large volumes of data get fed into a recording platform. Said platform can be anywhere from a handwritten field book to dedicated recording platforms such as Recorder and iRecord. Such people are *highly* unlikely to submit photographic vouchers for every single record, simply due to the time and inconvenience that would cause. Their data therefore defaults to `Considered correct`, leading to the majority overlap residing in the `Considered correct` verification category.

2.5.2. Verification status of overlap

Attempting to obtain exact figures for overlap is impossible without fully converting the iRecord dataset to simple unique record. None the less, it is possible to get an appropriate estimate (within compression + 3%).

This section examines the percentage overlap, of all processed iRecord records, to the BWARS simple unique record dataset. Note that some inflation of numbers is expected to happen here as including the verification status from the iRecord data will diminish the power of the simple unique record routine to condense data.

The results suggest that BWARS is managing to source iRecord data from far more non-photographic recorders than photographic recorders. This result is in line with expectations, as non-photographic recorders tend to be far higher in skill than photographic recorders.

Verification status	Records	Percentage
Correct	9,889	19.7%
Verified, other than correct	26,874	53.5%
Not verified	13,506	26.9%

2.5.3. Shortlist of available records

At this point in the selection process, the total pool of available records for immediate harvest in iRecord sits at ~79,101 records. This figure is calculated from the simple unique record (`Correct`) dataset figure of 88,990, minus the overlap in correct records of 9,889.

If verifiers which are known to have disobeyed the iRecord/BWARS verification guide are excluded, the pool decreases by approximately 2,600 records to ~76,600 `Correct`, unique, non-overlap records.

2.6. Taxon grouping

A breakdown of the broad-scale taxonomic composition of the iRecord and BWARS datasets. Only iRecord data which has passed the Checker is used in the breakdown due to the need for precision in nomenclature. The BWARS subset used is the de-duplicated `simple unique record` set.

2.6.1. Taxon groupings

Name	Description
Bombus	All non-aggregate species in the genus 'Bombus'
Other bee	All non-aggregate species in the family 'Apidae' other than in the genus 'Bombus'
Eusocial wasps	All non-aggregate species in the genera Vespa, Vespula, and Dolichovespula
Other wasps	All non-aggregate species other then eusocials, bees, or ants
Ants	All non-aggregate species of Formicidea

2.6.2. Absolute record count taxonomic composition

A breakdown of each scheme's broad scale taxonomic content, including a subdivision to `Correct` and `Considered correct`. Reported as absolute record counts

Scheme	Bombus	Other bee	Eusocial wasp	Other Wasp	Ant
BWARS	191,618	340,765	30,523	161,215	48,100
iRecord	139,019	141,190	30,512	21,630	9,027
iRecord Correct**	29,895	42,829	9,365	3,995	110
iRecord Considered correct**	42,389	38,312	8,085	7,636	189

** Verifiers known to have not followed the iRecord or BWARS verification ruleset removed. This will **very** strongly impact ant records.

2.6.3. Percentage-based taxonomic composition

A breakdown of each scheme's broad scale taxonomic content, including a subdivision to `Correct` and `Considered correct`. Reported as the percentage of each scheme's total

Scheme	Bombus	Other bee	Eusocial wasp	Other Wasp	Ant
BWARS	24.8%	44.1%	4.0%	20.9%	6.2%
iRecord	40.7%	41.4%	8.9%	6.3%	2.6%
iRecord Correct**	34.7%	49.7%	10.9%	4.6%	0.1%

Scheme	Bombus	Other bee	Eusocial wasp	Other Wasp	Ant
iRecord Considered correct**	43.9%	39.7%	8.4%	7.9%	0.2%

** Verifiers known to have not followed the iRecord or BWARS verification ruleset removed. This will **very** strongly impact ant records.

2.7. Identification difficulty

This section examines the breakdown of 'id difficulty', as created by Mike Edwards. The id difficulty rankings used for this were created for use in iRecord in 2013. The list has not been updated to Understandings, though the names have been converted using the Checker. The vast majority of nomenclature changes that cannot be automatically converted can be safely assigned to the highest difficulty. The assignment is based on the number of splits occurring that require highly detailed knowledge to identify being far greater than the number of simple 'new colonist' taxa.

In the case of sub-specialisations (e.g. '3 not males'), the additional information has been removed and the difficulty rating condensed to the major rank (in this case, 3). There are only two Understandings where re-classification has been applied and the impact is expected to be minimal.

Difficulty	Number of Understandings
1	18
2	47
3	53
4	188
5	323
6	1
NA	2
NULL	55

NA results are a product of the non-use of Understandings. NULL results are an indication of nomenclatural instability. Both classifications will be ignored from here on, as their results will be misleading.

2.7.1. Scheme difficulty comparison

The main take away from difficulty analysis is that iRecord data has a far higher weighting towards the 'easy to identify' taxa. Note that the iRecord dataset used here is 'all records which pass the Checker', so there is no subsetting to photo-identification records. The weighting reinforces the idea that iRecord and BWARS are capturing data from different sources. BWARS is capturing more 'advanced' users' data, whilst iRecord is capturing more 'beginner' data.

Note that *Melecta albifrons* was considered difficulty 5 but, due to the probable extinction of potential confusion taxa, should now be considered difficulty 1. Verifiers have treated this taxon as difficulty 1 and photo-identifiable. Of the 15,511 difficulty 5 records, 1,252 (8.1%) are of *M. albifrons*.

Difficulty	BWARS	iRecord
1	168,351	159,339
2	113,421	47,747
3	139,463	54,923
4	237,295	56,003
5	108,159	15,511
6	2	0

2.8. Scheme difficulty comparison as 'number of records per group'

The below table converts the scheme difficulty comparison table so that the values displayed are the number of records per member of each difficulty group according to the formula:

```
number of records in difficulty group / number of understandings in difficulty group
```

The result highlights the paucity of data on the very hard-to-identify taxa.

Difficulty	BWARS	iRecord
1	9,353	8,852
2	2,413	1,016
3	2,631	1,036
4	1,262	298
5	335	48
5**	335	44
6	2	0

** Without *M. albifrons* data

2.8.1. iRecord difficulty distribution by verification status 2

Due to space constraints, abbreviations have been used for iRecord's verification status 2 attribute. The abbreviations are as follows:

Abbreviation	Full
C	Correct
CC	Considered correct
P	Plausible
U	Unable to verify

There is a lot of information contained in the result of this section. On speaking with verifiers, there is considerable bias towards using `Plausible` over `Unable to verify` when a record is plausible but cannot be verified. Considering such information, it is clear that difficulty 3 & 4 taxa are considerably harder to verify.

It also seems that non-photographic recorders tend to record more of the common taxa than photographic, which is also expected. Surveys require that *all* taxa are recorded, not just interesting observations.

Difficulty	All	C	CC	P	U
1	159,339	40,702	51,147	911	469
2	47,747	16,755	14,205	505	274
3	54,923	13,579	15,572	1,109	309
4	56,003	14,188	17,295	1,483	611
5	15,511	3,099	5,423	629	173
6	0	0	0	0	0

Expressed as percentages of 'All'

Difficulty	All	C	CC	P	U
1	159,339	25.5%	32.1%	0.6%	0.3%
2	47,747	35.1%	29.8%	1.1%	0.6%
3	54,923	24.7%	28.4%	2.0%	0.6%
4	56,003	25.3%	30.9%	2.6%	1.1%
5	15,511	20.0%	35.0%	4.1%	1.1%

2.9. iRecord Verifiers & Rules

There are 252,671 verified records with a verifier's name against them in iRecord. Of these, 127,198 (50.3%) have been verified by one person. A further 80,430 records (31.8%) have been verified by a different person. The exact ruleset that these two people have been using has, according to conversation with them, varied over time. The variance has also included adherence to a 'BWARS accepted' verification score system. The BWARS rules differ slightly from the iRecord ruleset for `correct` and `considered correct` as follows:

2.9.1. iRecord

Status	Explanation
Correct	The verifier is able to confirm that the species has been identified correctly, usually on the basis of photo/s within iRecord (or specimen/s outside iRecord)
Considered correct	The verifier has not seen photo/s or specimen/s but has a high degree of confidence that the record is likely to be correct, based on difficulty of ID, date, location and recorder skills/experience etc.

2.9.2. BWARS

Status	Explanation
Correct	The verifier has examined the individual record and is confident that the information provided in the record provides enough to arrive at the determination
Considered correct	*Either* the verifier has not examined the individual record but trusts the origin of the record enough to accept the record, *or* the record has been examined but does not meet the criteria for `Correct`, and the verifier trusts the origin of the record enough to accept it

The BWARS verification system is designed to match the exact requirements of data accuracy over all, whilst maintaining a bypass for known and trusted submitters. Under the BWARS verification system, a `Correct` record's verifier takes on the same status as the record's determiner. Under `Considered correct`, the verifier is signifying their acceptance of the original determiner's capability.

2.9.3. Custom rules

There are also situations in which the verifier simply makes up their own rules. The third most prolific verifier in the system (8,583 records, 3.4%) is such a person.

2.10. Verification status breakdown

The range of approaches to verification in iRecord causes significant challenge to the ability to peer reviewed these verifications. It is at this point in the input and assessment process that 'invisible' errors - those which cannot be found without repeating the verification of the entire dataset - come to the front. Finding incorrect Correct records which have photos is only possible to do through human examination of individual records.

2.11. Misapplication of Correct

Under iRecord rules, records which are placed as `Correct` should have photos attached. Under BWARS rules, a photo *can*, under very specific circumstances, be marked as `Correct` without a photo. However, these records are very few and far between, applying to very few taxa. These taxa are ones which exhibit distinctive behavioural patterns in the absence of potential confusions and should be readily identifiable from the data.

The table below shows the number of records marked as `Correct`, followed by the subset of these without accompanying photos.

Verification status 2	Records
Correct	106,579
Correct with no images	11,425

2.11.1. Multi-record processing

Further examination of the 'Correct with no images' data paints a picture of misapplication of both the BWARS and iRecord scoring rulesets. It appears fairly common, on a verifier-by-verifier level, to apply the

Verification Status 2 ranks (`Correct`, `Considered Corrrect`, `Plausible`, `Unable to verify`, `Incorrect`) as pure confidence ranks based on any set of characteristics. Unfortunately, there is no way to identify multi-record processings reliably, due to the timestamp only reaching minute resolution. However, 14 situations where the verifier has verified > 60 records-per-minute have been identified. It is considered very unlikely that a verifier can reach a rate of > 1 record verification per second.

Further conversations and peer review have uncovered that these clusters of record verifications are a mixture of multi-record processing and 'spam clicking' through a pre-selected dataset using the individual processing tools. Both situations are most typically from either a period where the current verification ruleset was not in place, or from a verifier who was not aware of the current verification ruleset.

2.12. Verification Status 1 vs Verification Status 2

Two different sets of verification statuses exist within iRecord. These are referred to in the data as `verification status 1` and `verification status 2`. There have been issues in the past where records would be verified under `verification status 1` by one processing avenue and `verification status 2` by another avenue, all by the same verifier in the same session.

2.12.1. Number of records per year without `verification status 2`

`NULL` records seem to refer to records which are 'incomplete' (`Input still in progress`), as well as a number of what I believe were test records?

Despite the ability to use `verification status 1` being heavily reduced around 2020, someone *still* managed to find a way around that in 2023.

Year	# records
2010	58
2011	63
2012	1089
2013	16851
2014	9215
2015	10323
2016	5372
2017	8543
2018	7894
2019	858
2020	38
2022	51
2023	1
NULL	663

2.13. Number of known incorrect BWARS records present in iRecord

There are 95 records in the iRecord dataset which have been marked as suspended in the BWARS dataset.

2.14. Conclusion

An unexpectedly high rate of rejection was encountered when attempting to handle the conversion of binomials to binomial Understandings (14.1%). These rejections are almost entirely (98.0%) on the basis that the binomial cannot be resolved precisely to a type.

iRecord appears to be acting as a mainly primary data source i.e. most records are submitted by the originator of the record rather than a third party. BWARS is the opposite, functioning as a tertiary data source. 15.9% of iRecord data was found to already exist in the BWARS dataset, predominantly (53.5%) from the 'Considered correct' stream. Attempting to condense the iRecord data down to pure occurrence based on what/where/when only results in a 3.5% reduction in records, whilst BWARS has a 26.5% compression rate. The likely explanation is that BWARS acts as a tertiary data source and requests data from people repeatedly, whilst iRecord acts as a primary data source and relies on users to input their data directly. Experienced recorders do share records, particularly of rarities or notable observations, amongst themselves. Such sharing is potentially indirect, for example scraping the record from published material that mentions a particular occurrence.

The largest portion of records with a functional 'verification status 2' are those with `Considered correct` (31.2%), followed by `Unverified` (25.9%), then by `Correct` (25.8%). Whilst the majority of records *appear* to have been verified in accordance with either the iRecord or BWARS ruleset, there is no way to test the level of adherence. Conversation with verifiers uncovered two notable instances of people whose verifications are *not* in accordance. Instead, their verifications treat verification status 2 as a *measure of confidence* in the record being correct. When assembling the final count of 'harvestable records', these verifiers were specifically excluded from consideration.

The mistaken use of verification status 2 is driven by a mis-match between how the verifier reads the status and how iRecord/BWARS's rules mandate the status to be used.

In total, 52,441 records were found to match between the de-duplicated BWARS dataset and the raw iRecord dataset. This figure represents 15.9% of all of iRecord's data. The overlap is dominated by *Bombus*, with the notable exception of *Colletes hederae*, which was the subject of targeted effort. There are 95 records deemed incorrect by BWARS that are still live in the iRecord dataset.

Overlap between iRecord and BWARS was predominantly found in the `Considered correct` subset (53.5%), followed by `Unverified` (26.9%). The division is unsurprising and reflects the general difference in submitters between the two datasets.

Comparison of the 'difficulty of identification' between iRecord and BWARS reenforces the idea of an 'experienced identifier' cohort. As 'difficulty of identification' increases, the relative proportion of data between iRecord and BWARS favours BWARS more and more. *Melecta albifrons* data was specifically removed from 'difficulty 5' data due to changes in its circumstances.

Further investigation into difficulty took into account the verification status 2 of records in iRecord. The results here indicate that photographic recorders provide a very different species assemblage to non-photographic, though the difference is less stark than expected. The lack of starkness is most likely a result of outdated/incorrect classification of difficulty, as the table used has never been tested or fully reviewed

before. A re-assessment of identification difficulty, tailored to iRecord, would be required to investigate the topic of photographic vs non-photographic recording further.

Multi-record processing was very difficult to decisively identify, due to the minute-only resolution of the timestamps. Analysis did very heavily hint at the presence of multi-record verified data in places where it should not be, but no method could decisively identify such occurrences.

Verification status 1 has been almost entirely removed from use since 2020, though it was somehow still used by someone in 2023. The changes that deprecated verification status 1 likely occurred in early-mid 2019.

3. Spatial Range Checking

The spatial checking system works in a similar way to temporal range checking. The first step, in both spatial and temporal systems, is to identify a pool of 'similar as possible' records of the given taxon and year. The spatial distribution of the resultant pool of records is then plotted, followed by a 40 km buffer.

The data used in spatial range checking are only those which have passed the Checker. Automated bounds checking on a per-species level require exact-to-type precision in order to not fall foul of nomenclatural confusion. The requisite nomenclatural precision is supplied by Understandings assigned by the Checker.

Note that the spatial range checking routine does not evaluate any magnitude of rejection. For a Red List, magnitude of rejection is an incredibly important metric for the purpose of spatial threat categories. Two records which are extremely out of known range will likely have far more impact on spatially based threat categories than a hundred 'just outside' records. As the number of known records gets smaller and smaller, it becomes even more imperative to only use trusted data given the greater impact each individual records possesses.

> An attempt at examining the 'magnitude of rejection' was made during this project. The method looks at the distance from a given record to the edge of the 40km envelope. The distance is then combined with others in any applicable groupings to establish metrics, or individual records can be examined. Bear in mind that this *is* a first attempt to use this particular metric, even if the foundation is fundamental geometry.

3.1. Sources of major error

The 40 km buffer used for spatial range checking represents the generalised expected potential colonisation front of the taxon. Due to the lack of information on taxon-specific colonisation fronts it was not possible to create bespoke distances for taxa. The 40 km buffer region has been used in prior work successfully, including the BWARS import routine, and as such the distance has been used in this project as well.

However, as with all generalisations, there are likely taxa to which the 40 km radius is more applicable to than others. As before, the lack of strong comparable evidence makes the construction of an autecologically-based taxon-specific metric impractical.

BWARS does not seek or process data pertaining to *Apis* as such records are considered agricultural rather than ecological in nature. As such, all records of *Apis* are removed from analysis.

3.2. Breakdown of rejections per Understanding

This section breaks down the rejections from the routine to examine the distribution of Understandings in combination with different verification status 2.

3.2.1. Breakdown of rejections per Understanding, all records

A breakdown of where rejections occur, per Understanding.

Understanding	# rejections
Anthophora plumipes: iso. Amiet et al: 2007	388
Vespa crabro: iso. Archer: 1989	327
Vespula vulgaris: iso. Archer: 1989	297
Osmia bicornis: iso. Amiet et al: 2004	270
Bombus hypnorum: iso. Cameron et al: 2007	259
Andrena fulva: iso. Perkins: 1919	241
Lasius flavus: iso. Bolton & Collingwood: 1975	216
Anthidium manicatum: iso. Amiet et al: 2004	209
Colletes hederae: iso. Schmidt & Westrich: 1993	191
Vespula germanica: iso. Archer: 1989	187
Bombus muscorum: iso. Cameron et al: 2007	178
Andrena clarkella: iso. Perkins: 1919	173
Bombus pascuorum: iso. Cameron et al: 2007	163
Andrena haemorrhoa: iso. Amiet et al: 2010	155
Bombus humilis: iso. Cameron et al: 2007	152

3.2.2. Breakdown per Understanding, Correct only

Using only Correct verification status 2 records. Verifiers known to have not followed the iRecord or BWARS guidance have been removed from this step.

Understanding	# rejections
Anthophora plumipes: iso. Amiet et al: 2007	161
Vespa crabro: iso. Archer: 1989	122
Andrena fulva: iso. Perkins: 1919	92
Osmia bicornis: iso. Amiet et al: 2004	56
Bombus hortorum: iso. Cameron et al: 2007	54
Anthidium manicatum: iso. Amiet et al: 2004	49

Understanding	# rejections
Andrena bicolor: iso. Amiet et al: 2010	46
Nomada goodeniana: iso. Perkins: 1919	46
Andrena cineraria: iso. Perkins: 1919	44
Bombus vestalis: iso. Cameron et al: 2007	43
Andrena nigroaenea: iso. Perkins: 1919	43
Nomada marshamella: iso. Perkins: 1919	43
Bombus jonellus: iso. Cameron et al: 2007	40
Andrena clarkella: iso. Perkins: 1919	40
Bombus hypnorum: iso. Cameron et al: 2007	37

3.2.3. Breakdown per Understanding, `Considered correct` only

Using only `Considered correct` verification status 2 records. Verifiers known to have not followed the iRecord or BWARS guidance have been removed from this step.

Understanding	# rejections
Vespa crabro: iso. Archer: 1989	90
Anthophora plumipes: iso. Amiet et al: 2007	84
Bombus hypnorum: iso. Cameron et al: 2007	37
Bombus lapidarius: iso. Cameron et al: 2007	28
Bombus pascuorum: iso. Cameron et al: 2007	27
Bombus pratorum: iso. Cameron et al: 2007	25
Vespula vulgaris: iso. Archer: 1989	25
Andrena fulva: iso. Perkins: 1919	22
Bombus hortorum: iso. Cameron et al: 2007	22
Andrena cineraria: iso. Perkins: 1919	22
Bombus muscorum: iso. Cameron et al: 2007	22
Bombus monticola: iso. Cameron et al: 2007	22
Osmia bicornis: iso. Amiet et al: 2004	20
Bombus vestalis: iso. Cameron et al: 2007	19
Anthidium manicatum: iso. Amiet et al: 2004	17

3.3. Rates of rejection

A comparison of % error between various verification status 2 results. Verifiers known to have not followed the iRecord or BWARS guidance have been removed from this step.

Grouping	All	Errors	% error
Correct	91,207	1,999	2.2
Considered correct	105,568	1,157	1.1
Plausible	4,797	367	7.7
Incorrect	559	74	13.2
Unable to verify	1,916	277	14.5
Not reviewed	88,442	3,765	4.3

3.4. Magnitude of rejection

> This is an experimental metric which has been developed during this analysis. Take care interpreting the result as there has been minimal time to review and fully interpret the results. Conclusions drawn from this section are deliberately vague and restricted to only statements with perceived high confidence.

Magnitude of rejection is calculated as the direct distance between a record and the edge of the relevant year-matched 40 km buffer for that taxon. The sections below examine the mean magnitude for a given taxon, as well as the standard deviation for that taxon. Take care when interpreting these figures, as the number of records makes a significant impact on the reliability of these other metrics.

3.4.1. Magnitude, correct only

The 15 most common rejected Understandings for the `Correct` verification status 2 status. Sorted by mean, descending

Understanding	Mean (m)	Std Dev (m)	Num records
Anthophora plumipes: iso. Amiet et al: 2007	53,682	41,923	161
Bombus jonellus: iso. Cameron et al: 2007	41,942	35,035	40
Bombus vestalis: iso. Cameron et al: 2007	35,597	29,224	43
Andrena clarkella: iso. Perkins: 1919	27,428	20,256	40
Nomada goodeniana: iso. Perkins: 1919	25,827	26,793	46
Anthidium manicatum: iso. Amiet et al: 2004	23,801	21,174	49
Andrena fulva: iso. Perkins: 1919	23,520	22,257	92
Andrena nigroaenea: iso. Perkins: 1919	20,919	21,646	43
Vespa crabro: iso. Archer: 1989	18,302	17,994	122
Andrena bicolor: iso. Amiet et al: 2010	17,815	14,336	46

Understanding	Mean (m)	Std Dev (m)	Num records
Nomada marshamella: iso. Perkins: 1919	16,497	16,512	43
Bombus hypnorum: iso. Cameron et al: 2007	16,406	17,802	37
Andrena cineraria: iso. Perkins: 1919	15,663	10,894	44
Osmia bicornis: iso. Amiet et al: 2004	12,571	14,528	56
Bombus hortorum: iso. Cameron et al: 2007	12,028	24,870	54

3.4.2. Magnitude, considered correct only

The 15 most common rejected Understandings for the `Considered correct` verification status 2 status. Sorted by mean, descending

Understanding	Mean (m)	Std Dev (m)	Num records
Anthophora plumipes: iso. Amiet et al: 2007	38,099	32,458	84
Bombus muscorum: iso. Cameron et al: 2007	35,672	23,252	22
Anthidium manicatum: iso. Amiet et al: 2004	25,351	25,668	17
Bombus monticola: iso. Cameron et al: 2007	24,364	24,454	22
Bombus pratorum: iso. Cameron et al: 2007	21,069	24,612	25
Andrena fulva: iso. Perkins: 1919	20,193	22,970	22
Osmia bicornis: iso. Amiet et al: 2004	18,114	15,281	20
Vespula vulgaris: iso. Archer: 1989	15,798	22,918	25
Bombus hypnorum: iso. Cameron et al: 2007	15,400	26,099	37
Vespa crabro: iso. Archer: 1989	15,023	15,781	90
Bombus lapidarius: iso. Cameron et al: 2007	14,907	21,734	28
Bombus pascuorum: iso. Cameron et al: 2007	13,885	17,573	27
Andrena cineraria: iso. Perkins: 1919	13,881	11,593	22
Bombus hortorum: iso. Cameron et al: 2007	13,136	16,114	22
Bombus vestalis: iso. Cameron et al: 2007	11,919	15,824	19

3.4.3. Spatial magnitude per verification status 2

A breakdown of rejections, grouped by their verification status 2

Understanding	Mean (m)	Std Dev (m)	Num records
Correct	28,179	34,162	2,295

Understanding	Mean (m)	Std Dev (m)	Num records
Considered correct	24,409	31,712	1,511
Plausible	38,488	52,770	389
Unable to verify	95,041	116,600	341
Incorrect	95,704	110,405	103
Not reviewed	34,081	44,687	3765
NULL	34,852	50,341	1110

3.5. Discussion

3.5.1. Expansion potential

The spatial ranges of numerous taxa have expanded over the duration covered by iRecord and BWARS, with many species now entering geographic ranges where recorder coverage is poor. If gaps in recorder coverage are larger than, or even approaching, the 40 km buffer range, then it is possible for expansions of a taxon to go undetected until the taxon passes back into the range of a recorder. There appear to be signs of undetected expansions happening in the BWARS data that iRecord data could address, but entomologists have repeatedly expressed concerns and displeasure at the amount of spatial error also contained in iRecord data.

3.5.2. Major sources of rejection

Two major clusters of spatial rejections were identifed. One such cluster was, bar one record, a cluster of *Apis mellifera* records. Since records of *Apis* are considered records of livestock, BWARS does not incorporate or process many records of *Apis*. Those that are processed tend to be because it is more time consuming to remove them than to all the automated routines to process the data. In turn, the lack of data leads to many rejections of *Apis* records. To solve the issue of high false, and irrelevant, positives, *Apis* was removed from consideration. Removal of *Apis* records is akin to removing records of domestic poultry from ornithological recording.

The second main cluster of rejection was from the Isle of Mull (an island to the west of Scotland). BWARS records from the isle are very few and far between, which resulted in a large void in the 40 km buffer. An entomologist has been submitting data from within the void to iRecord, in excellent volume and with commendable caution, which resulted in a large cluster of rejected records. This cluster was considered a 'true' rejection - though not an error - and left in the analysis to avoid creating bias.

3.5.3. *Bombus hypnorum*

Bombus hypnorum underwent colonisation followed by significant, rapid expansion. The presence of the taxon near the top of the `Considered correct` rejections list is entirely expected, though records will still need examination to assess the degree of magnitude involved in the rejection.

3.5.4. *Anthophora plumipes*

An awkward species. Thought to have expanded rapidly over the past few years, there are still significant identification problems with two common bumblebees that led consulted entomologists to express concern and uncertainty over data quality.

3.5.5. Lack of general theme

In the temporal range checker there is a general theme of 'disturbed/interrupted behaviour' which explains a large number of the most common rejections. In the spatial set there is no such obvious, demonstrated, theme that covers any large group of taxa.

3.5.6. 'Correct' vs 'Considered correct'

On all examined metrics, `Considered correct` data has performed better than `Correct` data. Such a result mirrors all other checks performed on the iRecord dataset, though it is still impossible to say whether the subset of data *is* better, or whether the increase in quality metric is an artefact of increased recorder coverage.

3.5.7. Does verification work

Yes. Verifying data produces a higher quality result than the input. It is suspected that removing the records destined for `Considered correct` would only enhance the impact shown by photo verification. The exact impact of verification cannot be determined without a before & after comparison however, which would require a separate project.

3.5.8. Magnitude of rejection

As a first attempt, magnitude of rejection's strength in finding incorrect records cannot be established until a human re-examines the suspect records. As a technique for finding potentially impactful new data in a high-uncertainty area, magnitude seems very promising. Given the importance of maintaining clean data, especially in scarcer taxa, ways to separate range expansion from misidentification or errors in data entry are always worth exploring.

3.5.9. Further investigation

Magnitude of rejection provides the potential to investigate the appropriate size of the currently static 40 km buffer. Magnitude of rejection also provides a means to target human re-verification (or verification) to the size of rejection. Larger magnitudes should act as indicators to the verifier to be even more cautious of a given record.

Magnitude of rejection also provides a means to track the impact of different sized buffers, leading to a statistically backed per-taxon buffer size. In turn, accurate buffers will lead to more accurate data entry as well as more accurate buffer-union metrics for the Red List.

Many tests performed in this project require an A:B experiment to be performed to assess the 'true' impact of verification, as human intervention and re-verification is the only currently acceptable way to improve accuracy.

3.6. Conclusion

Many rejections were encountered, although without an immediately obvious overlying 'theme' to the rejections that would explain the majority of data. The 40 km figure used for buffers was untestable still, as tests require human opinion to re-verify data.

`Considered correct` data has a lower volume of rejection in both absolute and percentage terms than `Correct` data.

In both `Correct` and `Considered correct` statuses, there was a significant overlap in the Understandings with the majority of rejections. It is uncertain whether the overlap is the result of poor BWARS data volume, or reflects general difficulty in identification.

4. Temporal Range Checking

The temporal checking system works in a similar way to the spatial range checking. The first step is to identify a pool of 'similar as possible' records of the given taxon and year. From this pool, the minimum and maximum day of year (0-366) is identified. Incoming records of the given taxon and year are then checked to see whether they extend past these bounds, which would indicate a *potential* error in determination.

The term 'day of year' is frequently abbreviated to 'doy', especially in code snippets and explanations found online. Should there be a need to relate day of year back to a given Julian calendar date, NASA provide a useful calendar tool to help do that: https://www.esrl.noaa.gov/gmd/grad/neubrew/Calendar.jsp

The data used in temporal range checking are only those which have passed the Checker. Automated bounds checking on a per-species level require exact-to-type precision in order to not fall foul of nomenclatural confusion. The requisite nomenclatural precision is supplied by Understandings assigned by the Checker.

4.1. A word of caution

The temporal range checking performed here is very much in its infancy. A rejection of data does not in any way mean that the incoming record is at all wrong. Instead, the rejection is an indicator that the record is outside of expected timing and requiring specific attention from a human. Aculeate hymenoptera are known for their proclivity to adapt and interrupt their own phenology, with the result being records from species at very *odd* timings.

4.2. Which scheme is 'right'?

The short answer is 'neither of them'. iRecord data is being compared to BWARS data as the latter is 'stricter' than the former, due to the higher general threshold of entry. However, whilst the strictness improves data quality at an individual level, breadth of data does suffer as a result. A higher bar to entry creates a smaller pool of data, with the potential to miss many true outliers. The mechanism of comparison is also imperfect, being *incredibly* blunt.

4.3. Sources of major error

Ants are known to fail temporal range checks repeatedly, as their phenology leads to far greater year-round observations than most other aculeate hymenoptera.

Human intervention is the other source of major error, for example where phenology is intervened with. Stem nesters are at particular risk from intervention, where nests are reared out under altered conditions to identify the specimen.

For this section, the only data which can be used is that which has passed the Checker. Bounds identification requires accurate true-to-type nomenclature, which is provided by Understandings assigned in the Checker.

4.4. Impact of tolerance

Give that the temporal check is 'cutoff' check, there is minimal ability to adapt to records which are 'almost' within the time period. In the BWARS database, a 'tolerance' value can be applied to the outlier checking to ease the strain of checking large volumes of data. The following table examines the impact of various tolerances on the number of flagged records.

Tolerance	# records
0	5,047
1	4,814
2	4,576
3	4,366
4	4,173
5	4,019
6	3,825
7	3,668
10	3,235
14	2,807
21	2,239

4.5. Breakdown per taxon

A breakdown of temporal rejections, using various grouping metrics. This section uses a tolerance of 2 days, which is the same value used in BWARS import routines. In the absence of any statistical examination and experimentation of bounds checking, experience will have to suffice to provide a 'reasonable bound'. 2 days tolerance has proven to be a reasonable compromise between false positive/negative results during BWARS data entry.

Records must have their lower and upper dates be within the tolerance of one another in order for temporal range checking to work. Practically, this requirement removes records where the variance between the recorded lower and upper dates (e.g. month-only, year-only records) is greater than the permitted variance in checking. See table for example.

Tolerance	Lower doy	Upper doy	Acceptable
2	113	113	Y
2	187	189	Y
2	201	209	N

Tolerance	Lower doy	Upper doy	Acceptable
1	165	167	N
2	1	366	N

4.5.1. Breakdown of rejections per Understanding, all records

A breakdown of where temporal rejections occur, per Understanding

Understanding	# rejections
Vespa crabro: iso. Archer: 1989	362
Vespula vulgaris: iso. Archer: 1989	344
Bombus lapidarius: iso. Cameron et al: 2007	332
Bombus pratorum: iso. Cameron et al: 2007	239
Bombus hypnorum: iso. Cameron et al: 2007	148
Vespula germanica: iso. Archer: 1989	120
Bombus pascuorum: iso. Cameron et al: 2007	86
Colletes hederae: iso. Schmidt & Westrich: 1993	84
Andrena nitida: iso. Perkins: 1919	66
Bombus hortorum: iso. Cameron et al: 2007	62
Andrena cineraria: iso. Perkins: 1919	59
Anthophora plumipes: iso. Amiet et al: 2007	58
Dolichovespula media: iso. Archer: 1989	54
Bombus humilis: iso. Cameron et al: 2007	45
Andrena fulva: iso. Perkins: 1919	45

4.6. Breakdown per Understanding, `Correct` only

A breakdown of where temporal rejections occur, per Understanding, using only `Correct` verification status 2 records.

Understanding	# rejections
Vespa crabro: iso. Archer: 1989	168
Bombus lapidarius: iso. Cameron et al: 2007	117
Vespula vulgaris: iso. Archer: 1989	84
Bombus hypnorum: iso. Cameron et al: 2007	51

Understanding	# rejections
Bombus pascuorum: iso. Cameron et al: 2007	38
Bombus pratorum: iso. Cameron et al: 2007	35
Vespula germanica: iso. Archer: 1989	35
Anthophora plumipes: iso. Amiet et al: 2007	26
Andrena cineraria: iso. Perkins: 1919	24
Andrena nitida: iso. Perkins: 1919	23
Bombus sylvarum: iso. Cameron et al: 2007	22
Bombus hortorum: iso. Cameron et al: 2007	22
Osmia cornuta: iso. Amiet et al: 2004	19
Andrena fulva: iso. Perkins: 1919	18
Bombus vestalis: iso. Cameron et al: 2007	16

4.7. Breakdown per Understanding, `Considered correct` only

A breakdown of where temporal rejections occur, per Understanding, using only `Considered correct` verification status 2 records.

Understanding	# rejections
Bombus lapidarius: iso. Cameron et al: 2007	92
Vespula vulgaris: iso. Archer: 1989	83
Vespa crabro: iso. Archer: 1989	52
Bombus hypnorum: iso. Cameron et al: 2007	29
Bombus pratorum: iso. Cameron et al: 2007	29
Bombus pascuorum: iso. Cameron et al: 2007	15
Anthophora plumipes: iso. Amiet et al: 2007	15
Vespula germanica: iso. Archer: 1989	14
Bombus hortorum: iso. Cameron et al: 2007	14
Colletes hederae: iso. Schmidt & Westrich: 1993	11
Lasius flavus: iso. Bolton & Collingwood: 1975	11
Bombus sylvarum: iso. Cameron et al: 2007	10
Dolichovespula sylvestris: iso. Archer: 1989	10
Dolichovespula media: iso. Archer: 1989	8

Understanding	# rejections
Andrena fulva: iso. Perkins: 1919	7

4.8. Rates of rejection

A comparison of % rejection between various verification 2 statuses. We can see that photo-verifying incoming data drastically increases accuracy, but that data from experienced sources is yet drastically better still.

Grouping	All	Rejections	% rejection
Correct	85,845	1,022	1.2
Considered correct	98,154	572	0.6
Plausible	4,714	150	3.2
Considered correct**	98,154	507	0.5
Not reviewed	82,349	1,217	1.5
Verified, other than C/CC	6,595	257	3.9

** 'Considered correct' with the two main verifiers who used 'verification status 2' as confidence-based removed.

Comparison to BWARS is not possible at this point, since using BWARS data to check itself results in a perfect (meaningless) score.

4.9. Discussion

4.9.1. Human disturbance

Of the most common rejected Understandings, *Vespa crabro*, *Vespa vulgaris*, and *Vespula germanica* all hibernate in places which humans regularly disturb. Wood piles, chimneys etc. A quick sampling of records strongly suggests a pattern of disturbance, with many correspondents writing that they encountered the specimen through disturbing it whilst hibernating.

4.9.2. *Bombus lapidarius*

Many 'out of bounds' records in this sector are out by > 40 days compared to those in BWARS. Either BWARS data is highly insufficient, or the verifications of this taxon are incorrect. Subsequent peer review of these records has found that the majority of rejected *Bombus lapidarius* records in accurate, with the taxon able to emerge significantly earlier than anticipated.

4.9.3. *Bombus pratorum/pascuorum*

Conversations with Mike Edwards suggests that the majority of rejected records of these two taxa are the result of misidentification.

4.9.4. *Bombus sylvarum*

A lot of suspicious rejected records. Mike Edwards's comment was that it would be unusual to find *Bombus sylvarum* on Hawthorne flowers in mid February, as claimed by a subsample of records.

4.9.5. Does verification work?

Yes. The 'verified other than correct/considered correct' data runs a rejection rate of 3.9% against that of verified data at 1.2%. Incoming data sits at 1.5%, though the value is highly affected by the presence of large volumes of data destined for `Considered correct` i.e. acceptance based on the verifier's trust in the recorder. As such, photographic verification provides a greater than 0.3% decrease in rejections than suggested by pure figures.

4.9.6. Further investigation

The main avenue which entomologists would like to explore are the potential 'confusion' taxa which are leading to the out-of-bounds records. Knowing the potential confusions for photographic recording will aid future verifiers avoid such situations, leading to higher quality data output.

4.10. Conclusion

Far more temporal rejections than expected were encountered. The degree of tolerance for the day of year (doy) made a significant impact on the number of rejections. However, even unsuitable tolerances e.g. 21 days resulted in far more rejections than expected.

A value of '2 days' was selected for the tolerance in further exploration, as it is the value which has been used for BWARS. Note that a rejection cannot be considered an error until it has been checked by a human.

Examination of the Understandings present in rejections provides a good amount of insight into causes. Vespids make up a significant proportion of the most common rejections, with the likely reason that these are 'disturbances' of behaviour e.g. interruption of hibernation. Other taxa are far more concerning, especially given the lack of probability that the verifiers would get the identification incorrect given physical specimens.

`Considered correct` data has a lower volume of rejection in both absolute and percentage terms than `Correct` data. `Unverified` data has a higher volume of rejection than both `Correct` and `Considered correct`.

Further examination is required to delve further into temporal range checking. Enough information has been gathered, however, to provide a reasonable picture of the state of both iRecord and BWARS data to the Red List core team. Verification is resulting in fewer temporal rejections, but the exact size of the resulting data/rejection rate is unknowable without a before & after comparison. It is very plausible that the increased spread of recorders is detecting a higher number of 'true' outliers, but this cannot be automatically verified.

5. Introduction

A selection of records which were identified as being out of expected bounds were sent to an experienced verifier and data handler for a second opinion. The two sets of data identified were:

- All records which failed to pass the temporal boundary check process
- All verification status 2 'Correct' records which failed to pass the spatial boundary check process

Note that, in order to reach the boundary checking processes, the data must have passed through the BWARS Checker.

The re-verifier was asked to complete the iRecord verification status, using the BWARS version of the iRecord verification status 2 criteria:

- **Correct**: The verifier is able to confirm that the species has been identified correctly, usually on the basis of photo/s within iRecord (or specimen/s outside iRecord)
- **Considered correct**: The verifier has not seen photo/s or specimen/s (or the photograph is inadequate) but has a high degree of confidence that the record is likely to be correct, based on difficulty of ID, date, location and recorder skills/experience etc.
- **Unable to verify**: The verifier has a high degree of confidence that the record is likely to be incorrect based on difficulty of ID, date, location and recorder skills/experience (and where no photo/s or specimen/s are available); or photos are available but do not show enough detail to confirm the identification; and/or the record is not sufficiently well documented to confirm (e.g. location is vague or the date appears to be incorrect)
- **Incorrect**: The verifier is able to confirm that the species has not been identified correctly, or the record is erroneous in other respects, on the basis of photo/s or specimen/s, or on information from the recorder
- **Plausible**: The record is plausible based on species, date and location, but there is not enough supporting evidence for the possibility of misidentification to be ruled out, and the record remains within the "Unconfirmed" category

6. Temporal range checking

1,234 temporal outliers were supplied. Of these, 887 were re-verified and 347 verified for the first time. In the re-verification set, the highest degree of change was in the level of confidence that the verifier was willing to accept. 87 (21%) of the `Correct` records were downgraded to `Unable to verify`, as were 107 (39%) of the `Considered correct` records. Of all 188 non-correct data, only 4 records (2%) were moved into one of the 'correct' categories. `Correct` records were evaluated to be 299 (287 maintained and 12 incoming) against an initial 422, whilst `Considered correct` records were 153 (152 maintained and 1 incoming) against an initial 276.

In summary, of the 'potentially usable' `Correct`/`Considered correct` data, totalling 698 records, 452 records (65%) remained post re-verification.

Of the 347 newly verified records, 93 (27%) were accepted to either `Correct` or `Considered correct`. Such a low acceptance rate suggests that verifiers may be leaving records unverified rather than verifying them under a 'failing' grade.

7. Spatial range checking

Only `Correct` and `Considered Correct` records were sent to be re-verified due to the limited time available. Spatial extent checks, though usually done entirely by eye, are more familiar to the discipline than temporal checks. Accordingly, the error rate is expected to be significantly lower.

Of the 983 reviewed records, 981 were initially verified as `Correct`. The post review breakdown of these is as follows:

Review status	Count	Percentage
Correct	822	84%
Considered correct	86	9%
Incorrect	18	2%
Plausible	7	1%
Unable to verify	48	5%

All 2 of the `Considered correct` spatial outliers were accepted under review.

In total, 910 (93%) records were accepted as usable after review.

8. Post-review discussion

The large number of records being downgraded from `Correct` to `Considered correct` is the direct result of a change in application of verification standards from when the records were first evaluated to now. Historically, the 5 point scoring system was used as a confidence threshold. Records from known and capable recorders could be assigned as `Correct`, regardless of whether or not the record itself contained sufficient information to arrive at the determination. The entire iRecord dataset's `Correct` data must therefore be considered to contain a significant portion of `Considered correct` data.

Temporal outliers resulted in a far higher degree of true error, despite the extremely crude selection method used, than spatial errors. There is likely to be an element of tradition involved here, as all verifiers involved in this project and the iRecord data are used to the 'traditional' method of examining the spatial placement of a record on a map. In contrast, there is not yet a tradition of examining the date of a record in such context.

The high degree of acceptance of spatial outliers suggests that there exists a body of sporadic records which represent true occurrence not represented in the BWARS dataset.

9. Expert recommendation

The reviewer's recommendation on the question *'Should iRecord data be used for the current Red List attempt?'* is **no**. There were too many errors encountered for the verifier to be able to recommend the direct use of data. In particular, the history of iRecord, and the different ways in which verifiers have interacted with the system, have created a mixture of different results.

There is significant potential to the idea of using a similar review process, targeted at a specific species, in order to move accepted data into BWARS. However, this represents a significant amount of work that no volunteer is able to provide. Once the Red List is finished, priority taxa can be identified for specific attention in this regard, but the question of time still remains.

10. Further discussion

A few further specifically notable errors were observed. Firstly, the presence of data from 1st Jan 1970 for a taxon which was not present in GB at this time. It is strongly suspected that this date is from an epoch-date of 0, in turn likely from a record which did not have a date entered. This lack of date has, somewhere, been

interpreted as '0', leading to 1st Jan 1970. Such an situation is not unheard of, indeed being fairly typical. All incoming data should be searched for 1st Jan 1970 and, if found, discarded as being unusable.

Rearing records, as well as disturbance records, were present in low volume. Each of these records need to be converted to year-range records for BWARS purposes as the dates present in the record represent altered phenology. BWARS tracks date for the purpose of identifying flight period, not for tracking recorder activity.

11. Conclusion

Expert advice from peer review is to not incorporate iRecord data without further review. Analysis of records which are outside of 'known' spatial distribution according to the BWARS dataset suggests a reasonable degree of spatial expansion is possible for a select group of species. Temporal range analysis demonstrated the presence of a significant error within the iRecord dataset. A significant number of records which should be graded as `Considered correct` exist under the `Correct` verification status, which results from changes over time to the way in which `verification status 2` has been applied to records.## Conclusion Peer review and in depth analysis recommend against the use of iRecord data in a Red List without further verification to bring data up to the same standard as the BWARS dataset/current iRecord guidance. Whilst there is a degree of overlap between the iRecord and BWARS datasets, there is still high quality data which is held in iRecord and not present in BWARS. There are considerable difficulties in identifying the low-quality data, or mistakes in data entry, which preclude the immediate use of the entire iRecord dataset. The Red List process in particular is noted as having specific weakness to small volumes of incorrect data in already small datasets. The timescale of the iRecord dataset also invites skew as the scheme has not yet been recording for 30 years. However, this skew is fairly reliably identifiable and easy to therefore mitigate during interpretation of models.

The iRecord data generally adheres to a standardised format, making it very easy to manipulate, the need to translate binomials to binomial Understandings is a significant barrier to safe utilisation. There are also a number of general data errors e.g. a date of 1st Jan 1970 scatter into the data that require identification and correction (where possible).

Expert advice is that the overall iRecord dataset requires directed attention in order to reach a level where the data would be of sufficient quality for use. The majority of the attention stems from the age of the data and the changes over time that have been made to how verifiers have utilised the iRecord system.

Examination of flight period has identified a higher than expected number of records which are deemed to be highly likely to be inaccurate. Examination of spatial outliers found few records deemed highly likely to be inaccurate.

12. Glossary of terms for iRecord Data Analysis sub-project

This document contains definitions of terms which *only* apply to the iRecord Data Analysis sub-project.

12.1. Reviewer

The person reviewing the *verification* of a record

12.2. Verifier

The person verifying a record

12.3. Determiner

The person who makes the taxonomic determination of a record

12.4. Incorrect Correct error

A record has been given the status of 'Correct', without the record possessing the required voucher, in the assessment of the reviewer. Note that different verifiers may accept different levels of voucher completeness.

12.5. Voucher

A catch-all term for information contained with a record that may be used to arrive at a determination.