Moth Spreadsheet Validator User Guide

Mark Cubitt – January 2024

Contents

Introduction

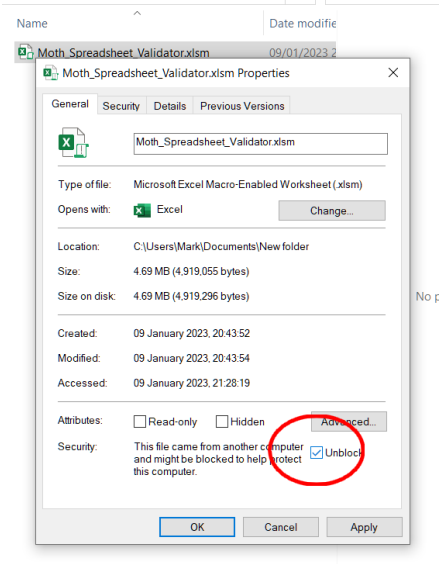
The macro and micro Moth Spreadsheet Validator provides four main types of function:

1. Parse record files of a variety of shapes and sizes, including Excel and comma-delimited files with either lists of records or matrix / pivot-table formats.
2. Create an output spreadsheet file with the records in one of three different formats including those that can be imported into MapMate and iRecord.
3. Resolves data values such as species names, life-stage and sex values and performs checks on the validity of the data in the records such as grid reference to vice county checking and setting to 100m level in GB when using a PC and 1km if using a Mac or for Ireland.
4. Provides a selectable variety of verification alerts to steer county recorders to the records needing most attention.

Download and installation

The Validator required Microsoft Excel to run and can be downloaded from https://butterfly-conservation.org/in-your-area/east-scotland-branch/moth-spreadsheet-validator.

The validator is an Excel Spreadsheet application that uses macros. It is recommended that you move it to a Documents folder that you use for your County Recorder work. Double-click the 'Moth_Spreadsheet_Validator' filename to open Excel. On a Mac then click on Enable Macros when prompted. On Windows, as it has been downloaded from the internet and contains macros Excel will not enable editing and will 'block' the macros from running. To resolve these you should click on the Enable Editing button and then view the macro block message. You now need to go back to Windows File Explorer and right-mouse-click the Moth_Spreadsheet_Validator file name and select 'Properties' from the shortcut menu. The Properties panel should have a Security section at the bottom with a checkbox that you can tick to Unblock the macros.



If you wish you can also create a shortcut to the validator by right-mouse-clicking the filename and selecting Create Shortcut. You can drag this to your Desktop from where you can start the validator in the future.

User interface

The Validator has two tabs or worksheets. The first is called 'Main' from where validations are run and the validation and verification options are chosen (see below). The second tab which will be displayed when you first open the validator contains some explanatory information and also some of the configuration that controls the validator that can be changed by the user, if required.

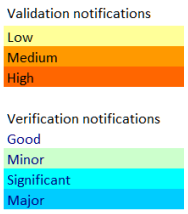
Running a validation

First we will just run the validator using the default settings and configuration; later in the guide we will discuss changing those settings.

From the 'Main' tab, click on the 'Validate Worksheet' button. This will open a file selection dialogue, allowing you to select an Excel spreadsheet or csv or tab delimited file to be validated. Select a records spreadsheet, txt or csv file and click Open. The Validator will now open the file in Excel and look for the column names for taxon, date, sitename etc. It does this based on the search words listed in the buff coloured panel in the README tab of the Validator spreadsheet. This process first finds-the first worksheet in the workbook that contains what look like records, so it doesn't need to be the first worksheet. If it can't find all of these in a single row of your records spreadsheet then it will try and work out if it is a matrix/pivot format file. If so, it will create a new worksheet and fill that with the records in a list format. Once it has found the columns it will put the names of the column headers into the table in the Main worksheet. A message in added to the Message Column (see below) to say which worksheet the Validator is processing. If there are multiple worksheets then you should check that it is the worksheet that your records are in.

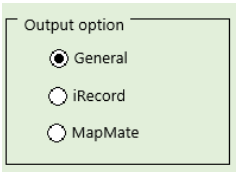
It will then start processing through the records. The validator's progress through the records can be observed at the bottom left in Excel's status bar where the row number being processed is shown.

On the right-hand side of the Main worksheet there is a wide column for messages. In this column there are messages about validator progress such as the name of the file and the worksheet containing the records. It will also contain validation and verification message alerts to guide you in checking the records. Each message is prefixed by the row number of the record and its background colour indicates the severity of the alert.



As the records are processed, as well as issuing the alert messages into the column of the Validator spreadsheet, the newly formatted records are inserted into a new worksheet called 'ValidatorOutput'. After processing all of the records you are prompted about whether to save the input file, with its new worksheet of formatted records, or whether to leave it to you to Save As the file with a different name or folder yourself.

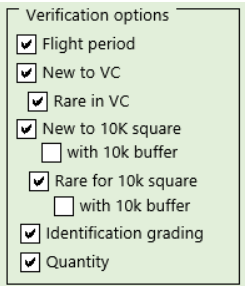
Output file formats



Three column formats and content are supported.

- iRecord – This format will import directly into iRecord allowing you to store your VC's data in the Portal. It uses the UKSI taxon names and the TVK species codes.
- MapMate – The MapMate import format is created. This uses the MapMate Taxon list and field terms. As this format only uses the scientific name, you might consider first doing your validation using the General format option, fixing any issues in the input dataset and then re-running the Validator to create the MapMate import format file.
- General – can be used if you keep your VC's records in your own database or in Excel files.

Verification options



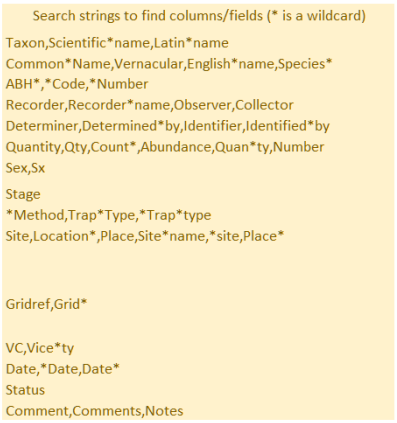
The Validator has a set of verification rules that can optionally be applied by ticking or un-ticking the checkboxes on the Main worksheet. There can be quite a large number of these alerts and so they are colour coded as mentioned above.

- Flight period – based on the first and last dates (and the reverse for over wintering flight periods) created from the national NMRS dataset for Record Cleaner. The warning uses the terms 'slight' if within 10 days of these limits, 'significant' within 50 days and 'major' for over this value outside the flight periods and has increasing severity background colours for each. The first and last dates are national and don't take account regional variation or multiple flight periods where these are distinct.
- New to VC – based on NMRS data which may be a year or two out of date (as are the verification rules below)
- Rare in VC – fewer than 5 records since 2000
- New to 10K square which also has a sub-option which does the check based on the 10K square itself plus the surrounding eight 10K squares. This latter option is for CMRs who have large, under-recorded VCs where the former option may produce too many alerts to be valuable. Not executed if 'New to VC'.
- Rare for 10K square - fewer than 5 records since 2000. Not executed if 'Rare in VC'.
- Identification grading – based on the published 1-3 grading numbers and guidance comments. The micro comments are life-stage specific and some of these have a Scottish bias. Warning are not issued for Grade 1 species and there is an option to only issue warnings for Grade 3 species which might be used for experienced recorders.

- Quantity – Broadly based on average number per trap plus two standard deviations by species, this check is to alert for typos and misidentifications.

Column finding configuration

The validator is focussed on being quick and easy to use and one feature of this is to try and avoid county recorders having to manually change record files before importing. Also in avoiding the need to manually tell the application which column is which. So, the Validator has a number of text strings that it searches for, some of which have wildcards. They have proven to be fairly robust in their ability to find the appropriate data, but if necessary you can tweak these column name finding search strings in the buff coloured panel on the README worksheet, if they don't suit your input files.



If you make changes to the configuration then, to retain these settings, just save the Validator spreadsheet file.

Reviewing validation and verification alerts

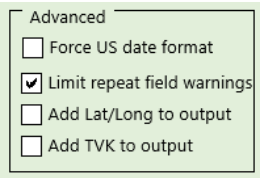
In addition to the alerts messages appearing on the Main worksheet, any records with alerts will have the background of the cells involved changed to the same colours as the alert messages in the input Spreadsheet. This makes them much easier to find. The cell backgrounds are changed I both the input worksheet and the ValidatorOutput worksheet. In the latter worksheet these messages will also be added as comments into those cells.

How you review the messages is up to you. Where there are only a small number of messages then it doesn't really matter, but for large numbers of messages then there are different approaches depending on your preference. One approach is to review the messages based on the Main worksheet and where needed then cross refer to the input or output worksheets. This is enabled by making the Excel windows sizeable and having the Validator spreadsheet on one side of your screen and the records spreadsheet on the other side. An alternative approach is just to go through the records spreadsheet and use the highlighted cells and comments as an aid.

In Excel, the individual cell comments can either be all shown or hidden from view using buttons the "Review" menu bar. If hidden then you can just hover your mouse pointer over the cell and the comment will be shown. By default only major alerts will have their comments shown, but even then the comments can cover up record information that you would like to read.

Advanced configuration

On the Main spreadsheet under the Verification Options panel there are some 'Advanced' options. These vary slightly depending on the Output format chosen.



It is highly recommended that all input dates are in UK date format dd/mm/yyyy. Other formats supported by Excel will work, but US mm/dd/yyyy format dates require the 'Force US Date Format' option to be selected in the Validator. If the wrong option is selected then dates that could be either format e.g 06/07/2019 will be processed as found, but Excel will switch round month and day if they don't match the format it is expecting and so cause unpredictable results. Date ranges with dates in US format are not currently supported.

The Limit repeat field warnings option will only repeat some types of alert up to 10 times. For example 'VC missing' will often apply to large numbers of records and you'll have got the idea that it is something to fix after 10 messages. The final message includes the fact if it is the last warning. Some warning such as a new or rare species at a 10k grid square will only appear once irrespective of this limit setting.

Including optional columns field values in the output files can be set as required. The Lat/Long can be useful for mapping tools that may require it, for example. Adding Taxon and Vernacular names to iRecord files or Code and CommonNames for MapMate is not necessary or allowed (respectively) to import them, but often useful when you want to do a sanity check before such an import.

The Stage option for the MapMate format is there because it can be desired for MapMate imports, but very often it clashes with Stage values in input files and so there is an option to turn on its use.

iRecord and MapMate have different interpretations of the meaning of zero for quantity. The former regards it as not present and the latter as present, but not counted. For the iRecord and General output formats there is a configuration option towards the bottom of the README worksheet where, by default, a 0 value will be converted to 'Present' unless there is a NonNumericQuantity column in the input file in which case the text value in that column will be used for the output Quantity value.

Remember to save your settings (and last used input folder) by saving the Validator spreadsheet.

Date ranges

Single dates and date ranges are supported. The format of dates accepted for input are any supported by Excel, but typically dd/mm/yyyy. Date ranges should be provided with the first and last dates separated by a hyphen or the word 'to', for example:

dd/mm/yyyy– dd/mm/yyyy
dd/mm/yyyy – dd/mm/yyyy
dd/mm/yyyy to dd/mm/yyyy

The output formats use either dd/mm/yyyy or, for date ranges, dd/mm/yyyy–dd/mm/yyyy.

Multi-column formats (such as Start Date, End Date) for date ranges are not supported by the Validator. These need to be combined into a single column called 'Date' using a formula such as =IF(F2 = G2, F2, TEXT(F2,"dd/mm/yyyy") & "-" & TEXT(G2,"dd/mm/yyyy")) assuming column F holds the start date and G the end date.

Matrix/Pivot input file considerations

Some recorders will use matrix style spreadsheets with species names down the left and dates across the top of a worksheet. The species counts are entered into the cells in the body of the worksheet. The Garden Moth Survey spreadsheet is one such, but there are many others. The Validator will handle these automatically by creating an extra worksheet called 'ValidatorImportList' which is subsequently processed.

The validator will also look for constant value 'column names', such as GridRef and Recorder in the rows above the Dates. These can either be constant for the whole workbook as in the first example. These rows/values can be added by the county recorder prior to validation.

	A	B	C	D
1	**Sitename**	**Nethy Bridge**		
2	**Recorder**	**Steve Brown**		
3	**VC**	**95**		
4	**Trap type**	**MV**		
5	**Gridref**	**NJ020207**		
6	**Taxon**	CommonName	15-Mar-21	16-Mar-21
7	*Ochropacha duplaris*	Common Lutestring		
8	*Achlya flavicornis*	Yellow Horned	2	2
9	*Archiearis parthenias*	Orange Underwing		
10	*Alsophila aescularia*	March Moth	1	
11	*Geometra papilionaria*	Large Emerald		

or even be constant for each dated column

	A	B	C	
1	Site	CarnDearg-700m	CarnDearg-700m	C
2	Gridref	NN8200090586	NN8200090586	N
3	VC	96	96	
4	Date	14/06/2021	08/10/2021	
5	Trap type	125W MV Robinson	125W MV Robinson	1
6	Recorder	Philippa G	Ellie D	P
7	Taxon			
8	Phlogophora meticulosa		2	
9	Enargia paleacea			

gVice counties

The vice county is set from the VC number or name in the input spreadsheet. There are no name synonyms, so numbers are preferred. If there are no VC numbers in the spreadsheet it will be set to the, optional, Default Vice County number which is set at the top of the ReadMe worksheet. These are then validated against the grid reference for each record. For GB this is done using the BSBI cucaera web service at 100m accuracy; for Irish grid references then against a 1km square lookup table. If a single VC is returned from the lookup then if it does not match then a warning will be issued and the value in the Output worksheet will be based on the grid reference. If the lookup spans VCs then if one of the looked-up VC numbers is in the input spreadsheet then a minor warning is issued and the input VC used; if it isn't then both VC numbers will be added and a major warning issued as this needs some action.

Combined Quantity and Sex fields

Some recorders put values such as '1m' or '7 f' into the quantity field/column. The Validator will recognise a number followed by one of the valid sex values (with or without a space between) and will place the quantity and the sex values into their own columns in the ValidationOutput worksheet.

Protected files

Excel allows spreadsheets and workbooks to be protected. A particular case in point is the GMS spreadsheet. The Validator handles these 'Protected' spreadsheet files by creating an unprotected copy of the file with the same name prefixed with 'Validation copy of'. This becomes the output file from the validator.

Importing Local Verification Rule Data

The data that drives the record verification is held in a hidden worksheet called VerificationData. This can be updated directly, if required although this will be lost next time the Validator is downloaded. On the ReadMe worksheet there is a button labelled 'Import Verification Rule Data' which allows a comma-delimited file (csv) to overwrite values in the VerificationData worksheet. This can be re-imported after a new version of the Validator is downloaded. The file format must be csv and the columns need to be in the following order:

Column header	Content
Taxon	UKSI Scientific name of species already in the VerificationData worksheet

FirstDate	The earliest emergence date in MMDD format (for overwintering species this will have an MM value greater than the LastDate)
LastDate	The latest adult flight period date in MMDD format
Grade	A value of 1, 2 or 3
AdultGradeComment	Text to be used for grade warnings for Grades 2 & 3 for adult stage
MineGradeComment	Text for mines
CaseGradeComment	Text for cases
Quantity	Value for which quantity values greater than will generate a warning

The file can be created using a spreadsheet, but must be saved as a csv file using comma as the field delimiter and double-quotes if there are commas in the cell/field value. A header row is not necessary, but it is advisable to use one and, if used, it must contain the word Taxon.

Only have rows for the species that you want to update the verification Data for. Only enter data in a cell that you want to update. Leaving a cell blank will not update the data.

The importer is not 'forgiving' and any invalid values will cause the import to terminate at that point. Any values imported to that point will be retained and not rolled back. The import can be re-run after the file is corrected. Closing the Validator spreadsheet without saving or redownloading it will avoid or resolve data corruption of the VerificationData worksheet.

Exporting data from MapMate

Although a standard export of records from a Browse window will work find in most circumstances, it is probably worth using a User (Custom) Query. Specifically, this format is required if data ranges are used.

To create the recommended User Query (courtesy of Martin Harvey)

1. Highlight and then copy the User Query Name below
2. Go to MapMate and use the normal method for creating customised queries: open the "Analysis" window, double-click on the "User Queries" sub-heading and choose <new user query>.
3. You will be prompted for a name; paste in the name you copied in step 1.
4. You will now see MapMate's SQL window with two lines of text - delete all the text that is there.
5. Highlight all three paragraphs of the SQL text below (the block of text in smaller font that has statements beginning SELECT, FROM, WHERE etc.), and copy this text. Make sure you highlight all the text from the beginning up to and including the semicolon at the end.
6. Return to MapMate and paste the SQL text into the blank window.
7. The WHERE clause (highlighted in blue below) restricts the records that will be selected. That whole line can be omitted if required. Defining that clause for your requirement is beyond this user guide, but note that when using a date field in the where clause the date must me enclosed between hask characters and the date is in US (mm/dd/yyyy) format.
8. Make sure you save the new query - click on the yellow folder icon at the top of the window.
9. Run the query, click on the Select All icon to select the result records and click on the Save Selection icon into a tab delimited file. (It is best not to copy and paste directly into a spreadsheet as the dates can be stored in US mm/dd/yyyy format which requires the 'Force US Date Format' option to be selected in the Validator.)

User Query Name
Browse records for Validator

User Query SQL:

```
SELECT Records.[_guk], ('MapMate export on ' & Format(Now(), "dd/mm/yyyy")) AS [Source of imported records], Records.[_guk] AS [Occurrence external key], Records.[_gen] AS [MapMate generation], [Taxa_1].Code AS MapMateSppCode, Taxa_1.Taxon AS Taxon, Taxa_1.Vernacular AS [MapMate vernacular], [Sites\Default].Name AS [Location name], [Vice-County].[VC] AS [Vice County (supplied)], [Sites\Default].OSGridRef AS [Grid ref], IIf(Records.Quantity=-1,'Dominant',IIf(Records.Quantity=-2,'Abundant',IIf(Records.Quantity=-3,'Frequent',IIf(Records.Quantity=-4,'Occasional',IIf(Records.Quantity=-5,'Rare',IIf(Records.Quantity=-6,'Very rare',IIf(Records.Quantity=-7,'Not present',IIf(Records.Quantity=-21,'Locally Dominant',IIf(Records.Quantity=-22,'Locally Abundant',IIf(Records.Quantity=-23,'Locally Frequent',IIf(Records.Quantity=0,'Present',Records.Quantity))))))))))) AS Quantity, IIf([Records].Date = [Records].DateTo,Format([Records].Date,"dd/mm/yyyy"),Format([Records].Date,"dd/mm/yyyy") & '-' & Format([Records].DateTo,"dd/mm/yyyy")) AS [Date], Recorders.Name AS [Recorder Name], Recorders_1.Name AS [Identified By], IIf(TaxonStage.Stage = 'Not recorded', '', IIf(TaxonStage.Stage = 'Larval', 'Larva', IIf(TaxonStage.Stage = 'Pupal', 'Pupa', TaxonStage.Stage))) AS [Stage (extensive)], IIf(TaxonSex.Sex = 'mixed sex group', 'mixed', TaxonSex.Sex) AS Sex, IIf(Methods.Method = 'Field record / observation', 'Field Observation', Methods.Method) AS [Method (free text)], Records.Comment AS [Occurrence comment], IIf(RecordStatus.Status = 'not recorded', '', RecordStatus.Status) AS [MapMate record status], ([References].[Author]&" "&[References].[Year]&". "&[References].[Title]) AS [Sample comment]

FROM ((((((((((Records INNER JOIN [Taxa\Default] ON Records.[*Taxon] = [Taxa\Default].[_guk]) INNER JOIN Taxa AS Taxa_1 ON [Taxa\Default].[**RefersTo] = Taxa_1.[_guk]) INNER JOIN [Sites\Default] ON Records.[*Site] = [Sites\Default].[_guk]) INNER JOIN [Vice-County] ON [Sites\Default].[ViceCounty] = [Vice-County].[_guk]) INNER JOIN Recorders ON Records.[*Recorder] = Recorders.[_guk]) INNER JOIN TaxonStage ON Records.[*Stage] = TaxonStage.[_guk]) INNER JOIN TaxonSex ON Records.[*Sex] = TaxonSex.[_guk]) INNER JOIN Recorders AS Recorders_1 ON Records.[*Identifier] = Recorders_1.[_guk]) INNER JOIN Methods ON Records.[*Method] = Methods.[_guk]) INNER JOIN [References] ON Records.[*Reference] = References.[_guk]) INNER JOIN RecordStatus ON Records.[*Status] = RecordStatus.[_guk]

WHERE Records.[Date]>=#01/01/2022#

ORDER BY [Taxa\Default].Code, [Taxa\Default].Taxon;
```

Exporting data from iRecord

From the https://irecord.org.uk/downloads form, create a Filter to select the desired records to be downloaded. Both the simple and Standard download csv formats are supported.

Importing the output into databases

You will need to refer to the documentation provided by your database for import instructions, but often these will require the 'ValidatorOutput' worksheet to be saved as a separate csv or xlsx file. To save as a separate Excel file, create a new blank Workbook, select all of the data in the ValidatorOutput worksheet and copy this and paste into the new workbook. To save as a csv or tab delimited file you just need to make the ValidatorOutput worksheet active by selecting its tab and then using File> Save As and changing the Type to, for example, comma-delimited. This will save only the data contents and ignore the comments in the worksheet cells.

iRecord imports can be done from https://irecord.org.uk/import. The 'Survey dataset' should be set to 'iRecord Import' and the Species List to 'UK Master List'.

Restrictions

Some species don't have all of the rule data with which to issue verification alerts.