

DEVELOPMENT OF MATCHED (MIGRATORY ANALYTICAL TIME CHANGE EASY DETECTION) METHOD FOR SATELLITE-TRACKED MIGRATORY BIRDS

Tomoko Doko ^{a,b}, Wenbo Chen ^{a,b*}, and Hiroyoshi Higuchi ^b

^a Nature & Science Consulting Co., Ltd., Room 302, 3F, Chuo-dairoku-kannai Bld., 1-2-1 Furocho, Naka-ku, Yokohama 231-0032 Japan - doko@nature-science-consulting.co.jp, chen@nature-science-consulting.co.jp

^b Keio Research Institute at SFC, Keio University, 5322 Endoh, Fujisawa 252-0082 Japan - hhiguchi@sfc.keio.ac.jp

Commission II, WG II/3

KEY WORDS: change point analysis, data mining, MATCHED method, migration status, migratory birds, satellite telemetry, stopovers, threshold date

ABSTRACT:

Satellite tracking technology has been used to reveal the migration patterns and flyways of migratory birds. In general, bird migration can be classified according to migration status. These statuses include the wintering period, spring migration, breeding period, and autumn migration. To determine the migration status, periods of these statuses should be individually determined, but there is no objective method to define 'a threshold date' for when an individual bird changes its status. The research objective is to develop an effective and objective method to determine threshold dates of migration status based on satellite-tracked data. The developed method was named the "MATCHED (Migratory Analytical Time Change Easy Detection) method". In order to demonstrate the method, data acquired from satellite-tracked Tundra Swans were used. MATCHED method is composed by six steps: 1) dataset preparation, 2) time frame creation, 3) automatic identification, 4) visualization of change points, 5) interpretation, and 6) manual correction. Accuracy was tested. In general, MATCHED method was proved powerful to identify the change points between migration status as well as stopovers. Nevertheless, identifying "exact" threshold dates is still challenging. Limitation and application of this method was discussed.

1. INTRODUCTION

1.1 Satellite tracking

"Satellite tracking" or "satellite telemetry" is the process of measuring and processing information received from satellites to determine the location and/or characteristics of the signals. The development of satellite telemetry in the late 1980s allowed the first in-depth studies of animal movement patterns (Coyne and Godley, 2005). Satellite tracking technology has been used to reveal the migration patterns and flyways of migratory birds (Higuchi and Pierre, 2005).

In this field, the Service Argos emerged as the predominant satellite-based system for tracking wildlife (Argos, 1996). Satellite tracking involves attaching a tracking equipment, called a Platform Transmitter Terminal (or PTT) to target animals.

Despite the obvious power and advantages of the Argos system, the data generated are difficult for many biologists to fully exploit. Although steps have been taken to improve distribution mechanisms, the data stream still causes difficulty at various stages including downloading, collating, filtering and interpretation (Coyne and Godley, 2005).

1.2 Needs for data mining of satellite-tracked data

In general, bird migration can be classified according to migration status. These statuses include the wintering period, spring migration, breeding period, and autumn migration. Although variations in day length and weather conditions are important in determining the timing of their migration (Eileen, 2010), individual birds do not leave a wintering site on the same day due to various conditions, e.g., microhabitats and/or microclimates that each individual encounters in its habitat. Thus,

to determine the migration status, periods of these statuses should be individually determined.

To our knowledge, however, there is no objective method to define 'a threshold date' for when an individual bird changes its status. For instance, Limiñana et al. (2007) used distance to estimate ending date of migration but did not address detecting all migration status.

1.3 Research objective

The research objective of this study is to develop an effective and objective method to determine threshold dates of migration status based on satellite-tracked data. The developed method was named the "MATCHED (Migratory Analytical Time Change Easy Detection) method". In order to demonstrate the method, data acquired from satellite-tracked Tundra Swans were used.

2. METHODS

2.1 Satellite tracking

In total, 27 Tundra Swans were captured during the winters of 2009 through 2012 (October to April) at Kutcharo Lake (45°07'28.560" N, 142°20'45.743" E) near the town of Hamatonbetsu, Hokkaido, Japan. There were 12 females and 8 males. The sexes of 7 swans were unknown. All individuals were adults. The swans were captured using flat net traps or hands. These swans were equipped with green neck collars and alphanumeric codes to which satellite transmitters were attached. The seven PTTs (PTT ID: 96824 to 96830) were programmed to transmit for 8 h and turn off for 23 h. The other PTTs (PTT ID: 93911 to 93917, 96816 to 96823, 78021, and 89040) were programmed to transmit for 9 h and turn off for 35 h. Three PTTs (PTT ID: 96817, 96819, 96826) were collected in the field and

* Corresponding author

reattached to other individuals. The PTTs used were solar-powered “SOLAR BIRDBORNE PTT” that weighed 12 g and were made by North Star Science and Technology LLC (King George, VA, USA). A PTT plus neck collar weighs approximately 47 g and is approximately 0.9 % of the swan’s body weight.

The PTT locations were estimated using the Argos System (Argos, 1996; CLS (Collecte Localisation Satellites), 2007). Data collected using the Argos System included the date, time (Greenwich Meridian Time; GMT), and longitude and latitude (WGS84 datum) with location accuracy. The location accuracy (location class, LC), which is attributed based on the radius of error, was classified as 3 (estimated error < 250m), 2 (250m < estimated error < 500m), 1 (500m < estimated error < 1500m), 0 (estimated error > 1500m), A, B, and Z and was reported using the Argos System (Argos, 1996).

Prior to statistical analysis, the data were filtered using The Douglas Argos-Filter Algorithm (Douglas et al., 2012) which ingests Argos satellite tracking data and flags improbable locations based on user-defined distance and velocity thresholds.

Data of a total of 16 individuals (PTT IDs and sex: 78021U, 89040M, 93913M, 93915F, 93917F, 96816U, 96819F, 96821M, 96822F, 96823U, 96824M, 96825M, 96826M, 96827F, 96828F, 96830M) were used in this study because these data were considered to be durable for migration pattern analysis. Six of these swans were females (F), seven were males (M), and the rest were unknown (U). The maximum tracking period was from 2009 to 2012.

2.2 MATCHED method

“MATCHED (Migratory Analytical Time Change Easy Detection) method” was developed in this study.

The overall flow of MATCHED method is illustrated (Figure 1). There are six procedures in the MATCHED method: 1) dataset preparation, 2) time frame creation, 3) automatic identification, 4) visualization of change points, 5) interpretation of change points, and 6) manual correction of threshold dates.

The method contributes to data analysis on collating, filtering and interpretation of satellite-tracked data.

Following is a detailed description of the MATCHED method.

2.2.1 Dataset preparation: The dataset was created as follows. After filtering through the Douglas Argos-Filter Algorithm, the bad location classes of LC 0, A, B, and Z were eliminated, and only good location classes, such as LC 3, 2, and 1, were retained for analysis.

If there were several overlapping dates, a single date was selected by choosing better location classes. In situations when the same date had the same location class, the earlier record was selected.

Geographic coordinates (x_i, y_i) in meters were calculated based on latitude and longitude in degree. The data in this study are presented as a spatio-temporal surface pattern $\{u(x_i, y_i, t_i) | i = 1, 2, \dots, n\}$ with horizontal indexing space (x_i, y_i) (unit is m) in 2-D and t_i indicating the sampling time (the unit is the date).

The distance between two points is the length of the path connecting them. In the plane, the distance between points

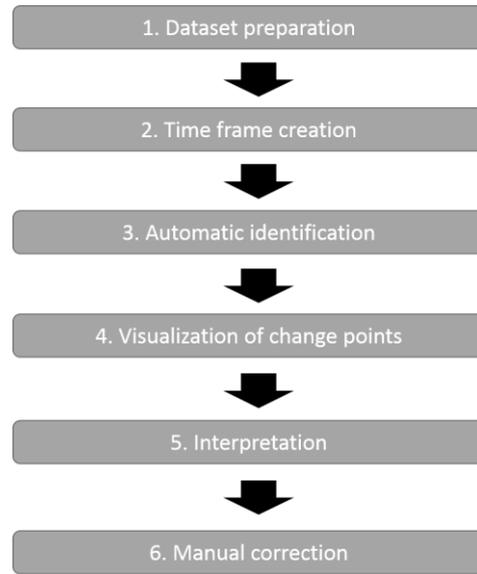


Figure 1. Overall flow of MATCHED method

(x_i, y_i) and (x_{i+1}, y_{i+1}) is given using the Pythagorean Theorem,

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} / 1000 \quad (1)$$

to acquire the distance in km. In this study, d_i ($i = 1, 2, \dots, n$) was calculated individually. The distance of the first record of an individual was treated as $d_1=0$.

Because data acquisition through the Argos System depended on variations in satellites, weather, etc., t_i could not be regularly sampled at an equal sampling spacing. As a result, there were numerous missing dates in the dataset.

Thus, the daily distance (v_i ; unit is km/day) was derived by the equation (2) to be comparable for each subsequent record:

$$v_i = d_i / (t_{i+1} - t_i) \quad (2)$$

The above data treatment was conducted using Microsoft Excel 2010. In the spatial analysis (ArcMap ver. 10[®], ESRI Inc.), simple location indicators, such as Japan, ocean, and Russia, were compiled to represent one column. The sample data of raw data acquired from ARGOS system and dataset prepared by this procedure is compared in Figure 2 (a) and (b), respectively.

	A	B	C	D	E
1	PTTID	Date(UTC)	Lat	Lon	LC
2	93911	2009/4/18 10:14	45.146	142.394	A
3	93911	2009/4/18 10:18	45.13	142.348	0
4	93911	2009/4/20 6:22	45.123	142.344	1
5	93911	2009/4/20 6:52	44.751	141.679	B
6	93911	2009/4/22 0:54	45.128	142.34	2
7	93911	2009/4/22 2:49	45.098	142.36	B
8	93911	2009/4/25 18:13	45.274	142.209	1
9	93911	2009/4/29 6:09	49.296	140.396	B
10	93911	2009/5/3 2:38	51.702	140.886	B

Figure 2 (a). Sample data acquired from ARGOS system

	A	B	C	D	E	F	G	H	I	J	K
1	Number	PflID	No	class	Date	Long	Lat	Dist_km	Duration_d	Speed_km_d	Location
56	18773	78021	55	2	2011/4/24	142.340000000000	45.119000000000	2.962	0.278	10.636	Japan
57	18774	78021	56	2	2011/4/26	142.330000000000	45.139000000000	3.336	1.602	2.083	Japan
58	18775	78021	57	1	2011/4/26	142.336000000000	45.119000000000	3.215	0.090	35.616	Japan
59	18785	78021	58	1	2011/5/5	148.917000000000	48.237000000000	889.502	9.197	96.714	Ocean
60	18787	78021	59	1	2011/5/5	148.910000000000	48.211000000000	4.401	0.058	76.354	Ocean
61	18795	78021	60	1	2011/5/7	153.726000000000	54.258000000000	1200.649	1.902	631.229	Ocean
62	18802	78021	61	1	2011/5/8	160.737000000000	58.821000000000	1207.017	1.632	739.619	Russia
63	18806	78021	62	1	2011/5/8	160.722000000000	58.809000000000	3.069	0.060	51.390	Russia
64	18810	78021	63	3	2011/5/9	160.736000000000	58.828000000000	4.365	0.059	73.955	Russia
65	18811	78021	64	2	2011/5/9	160.719000000000	58.823000000000	2.176	0.011	195.805	Russia
66	18812	78021	65	1	2011/5/9	160.747000000000	58.820000000000	3.183	0.009	352.552	Russia
67	18816	78021	66	2	2011/5/9	160.747000000000	58.841000000000	4.509	0.072	62.426	Russia

Figure 2 (b). Sample data after treatment of dataset preparation

2.2.2 Time frame creation: R software version 3.1.3 (R Development Core Team, 2006) was used to analyze migration patterns as time-series data and to determine the threshold dates of the migration status. The “zoo” (Zeileis and Grothendieck, 2005) and “changepoint” (Killick and Eckley, 2014) of R packages were used.

The change points in this study were defined as threshold dates between two consecutive migration statuses, e.g., wintering and spring migration. The migration status was defined as wintering, spring migration, breeding, or autumn migration. The migration status sequence is crucial for this analysis. In addition, threshold dates between stopovers during spring migration and autumn migration were defined as change points too. The concept of migration status is summarized (Figure 3).

The time frame was created for each individual using “zoo”. Each time frame was assigned to begin at the initiation of satellite tracking until the end date of satellite tracking for an individual. Because satellite tracking data usually have missing data, the time series analysis should be able to be conducted with irregular time series data. Hence, the package “zoo” was suitable to analyze such data because it has an advantage to unitize irregular time series data, compared to capabilities of other time series packages in R.

2.2.3 Automatic identification of change points of migration status: The “changepoint” package can automatically identify threshold values. It is possible to identify a single change point or multiple change points. An ordered sequence of data, $z_{1:n} = (z_1, z_2, \dots, z_n)$, is given. If a changepoint is said to occur within this set when a time is present, $\tau \in \{1, 2, \dots, n - 1\}$, two segments of $\{z_1, \dots, z_\tau\}$ and $\{z_{\tau+1}, \dots, z_n\}$ can be statistically differentiated. If a number of changepoints, m , together with their positions, $\tau_{1:m} = (\tau_1, \tau_2, \dots, \tau_m)$, is given, each changepoint position is an integer between 1 and $n - 1$ inclusive. It is defined that $\tau_0 = 0$ and $\tau_{m+1} = n$. The changepoints are assumed to be ordered as $\tau_i < \tau_j$ if $i < j$. Consequently, the m changepoints will split the data into $m + 1$ segments with the i th segment containing the data $z_{(\tau_{i-1}+1:\tau_i)}$. Each segment will be summarized using a set of parameters.

To identify the optimal combinations, the daily distance, latitude, and longitude were respectively tested via three scripts of changepoint package: “cpt.mean”, “cpt.var”, and “cpt.meanvar”. The scripts “cpt.mean”, “cpt.var”, and “cpt.meanvar” can calculate threshold values based on changes in the mean, variance, and mean and variance, respectively.

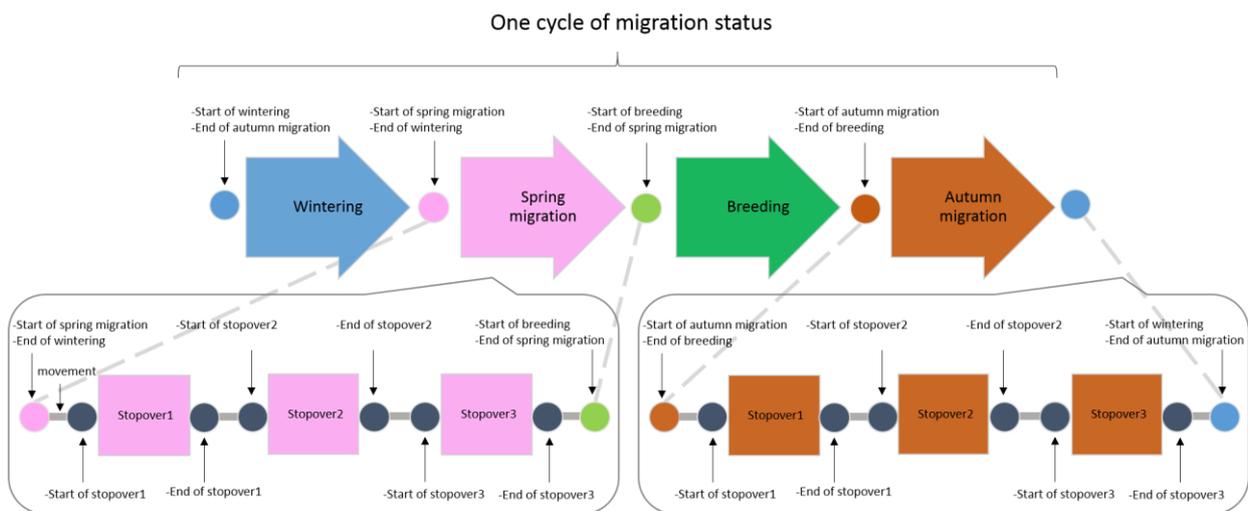


Figure 3. Conceptual diagram for explaining migration status and change points.

The migration status was defined as wintering, spring migration, breeding, or autumn migration. The migration status sequence is crucial for this analysis. This example illustrates the case when the number of stopover is three. A circle represents a change point.

The combination of the “cpt.meanvar” method for latitude best fitted the data based on a visual comparison with migration patterns (Figure 4). Hence, this combination was used for all individuals. The number of change points was adjusted based on the length of the tracking period.

2.2.4 Visualization of change points: After the change points were identified using the above method, these threshold dates were plotted against backgrounds showing the latitude, longitude, and daily distance with the simple location indicators of Japan, the ocean, and Russia. The scales of latitude and longitude were adjusted with the scale of the daily distance so that the change patterns in these variables were visible in one graph. Figure 5 (a) shows the example of PTT ID 96825. This graph was named “visualization plot”.

2.2.5 Interpretation of change points: Based on visualization plots, the meaning of this graph should be interpreted. One interpretation example of the MATCHED method was given using PTT ID 96825 (Figure 5 (b)). During wintering, the three lines of latitude, longitude, and daily distance were generally flat and stable (early December, 2009 to April 23, 2010 for this example). In this case, April 23 should be the threshold date between wintering and spring migration. The starting date of wintering was unknown for this example. Flat and stable patterns for daily distance, longitude, and latitude suggest migration ends. The latitude and longitude in wintering sites were lowest which suggested that the swans remained in Japan which is located southwest compared to Russia.

During spring migration, there was a clear step-like pattern indicating simultaneous increases in latitude and longitude (From 23 April 2010 to 1 June 2010). This pattern indicates that the swans moved northeast, i.e., from Japan to Russia. On the other hand, the daily distance had several peaks. The number of peaks in daily distance and the number of steps according to longitude and latitude synchronized. These changes indicate stopovers because the swans’ geographic location should be the same during each stopover, but the daily distance should increase prior to reaching the stopovers. This daily distance over a short period of time is represented as a peak.

After arriving at breeding or terminal sites, all daily distance, longitude, and latitude parameters became flat and stable again, and the latitude and longitude reached their highest values during migration (1 June 2010 to 23 September 2010). The highest latitude and longitude values indicate that the swans reached their terminal sites in Far East Russia in the Arctic.

The autumn migration showed a step-like shape with decreases in latitude and longitude (23 September 2010 to 2 November 2010). This result indicates that the swans moved to south-west during autumn migration from the Arctic in Russia to their wintering sites in Japan.

2.2.6 Manual corrections: All threshold dates were confirmed using the data after treatment of dataset preparation (Figure 2 (b)). If the threshold dates were considered to be incorrect or missed via automatic detection through “change point” analysis, these dates were manually identified based on visualization plot, and geographic location.

In general, the judgement of movement from one stopover to another stopover was most difficult. When the birds move from one stopover to another stopover, they move from dense area to another area suddenly. Hence, by judging geographic location from map, several samples which were inside one group, e.g.

circle of dense area, were considered to belong to one stopover. Moreover, if distance from one point to another point is much longer than the normal distance among points in one stopover at a scale where the whole study area can be seen in one map, this was considered movement from one stopover to another stopover.

Finally, identified threshold dates were manually added.

In this example given by Figure 5 (a) (b), the threshold dates, April 23, and September 23, were correctly identified by “change point” analysis. The detected threshold date, May 31, was different from correct threshold date, which is June 1. Another threshold date of November 2 was missed by “change point” analysis. In this case, November 2 should be added manually as a threshold date between autumn migration and wintering.

2.3 Mapping migration status

The migration status was mapped to demonstrate an example of application.

2.4 Accuracy assessment

The accuracy of MATCHED method was evaluated by a following method.

In this study, the data after manual corrections (chapter 2.2.6) was defined as the truth data, and the data from “change point” analysis (chapter 2.2.3) was defined as the test data. The threshold dates of test data was read from the visualization plots.

There were three types of classification results: 1) change point and threshold date were correctly detected (“matched date”), 2) change point was correctly detected but threshold date was slightly different (“incorrect date”), and 3) change point was missed by detection (“missing date”) (Figure 6). The number for matched date, incorrect date, and missing date were counted, and percentage of each type was calculated. For the case of incorrect date, the averaged difference of threshold dates was calculated. Accuracy of migration status and stopovers was calculated separately.

3. RESULTS

3.1 Visualization of MATCHED method

Figure 7 shows some examples of visualization plots. For example, threshold dates by PTT ID 96822 were successfully detected by “change point” analysis and visualized well by a visualization plot (Figure 7 (A)). Three cycles of migration status were detected in the case of PTT ID 96822.

On the other hand, PTT ID 96830 (Figure 7 (B)) should had three cycles of migration status too. However, because there were lacking tracking periods where migration status had been changed, threshold dates were not detected well, except the threshold dates between wintering and spring migration, and change points between stopovers during spring migration.

3.2 Mapping migration status

Figure 8 illustrates the map of classified migration status. The example was given by PTT ID 96819. There was a wintering area in Hokkaido of Japan, and a breeding area along Kolyma River at Russia. Between a wintering area and a breeding area, there were several stopover sites.

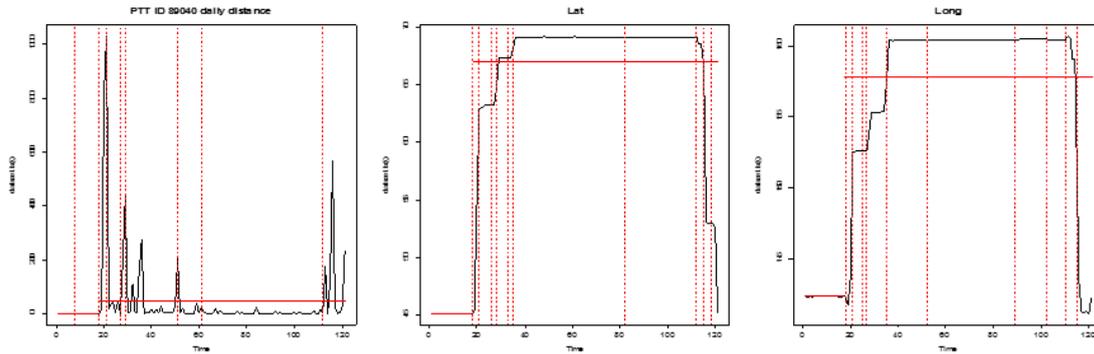


Figure 4. Change points detected by daily distance, latitude, and longitude

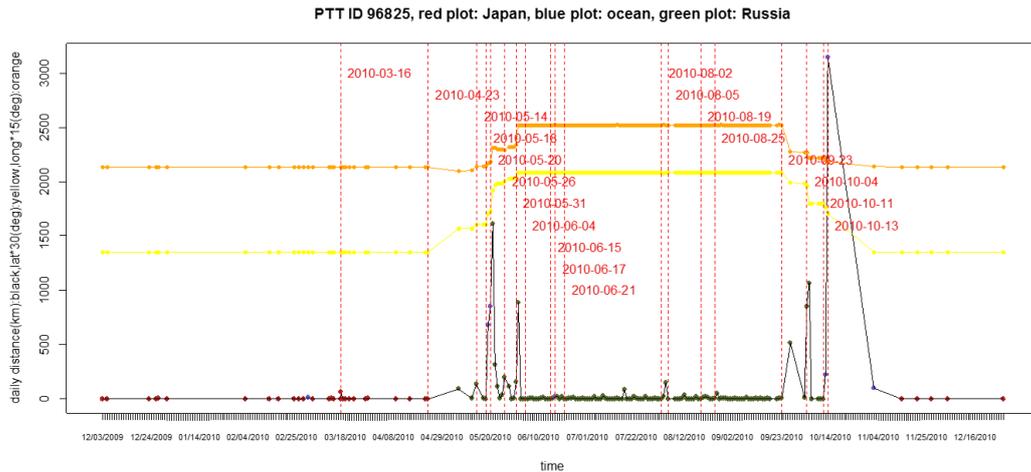


Figure 5 (a). Visualization of change points

The visualization plot can be created automatically by R scripts. Based on the visualization plot, interpretation should be done.

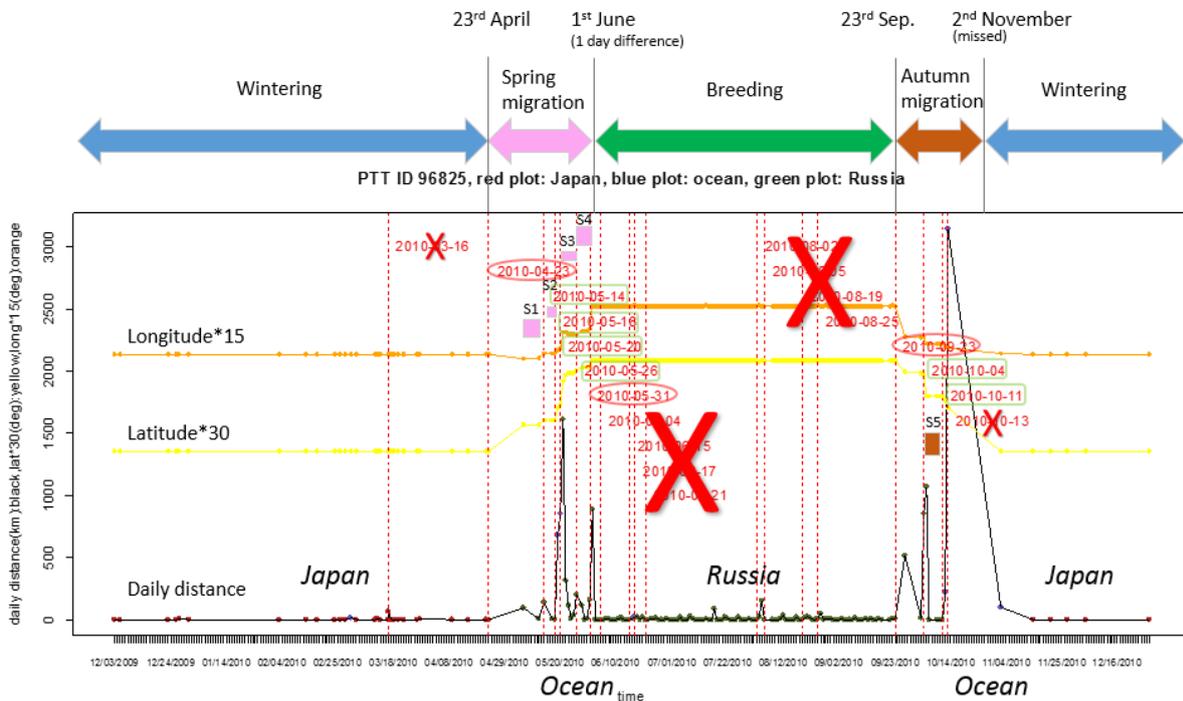


Figure 5 (b). Interpretation of change points

The migration status can be identified by interpreting the visualized graphic. The dates by red circles are threshold dates which were successfully detected to differentiate migration status. The dates by green rectangles are threshold dates which were threshold dates between stopovers. The dates by red crosses can be ignored, which were considered not be related to migration status.

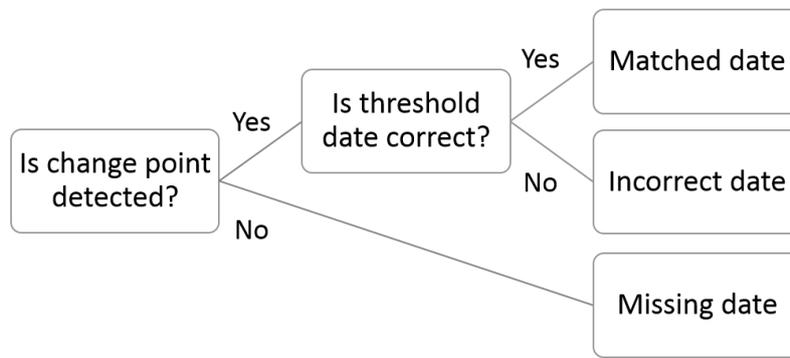
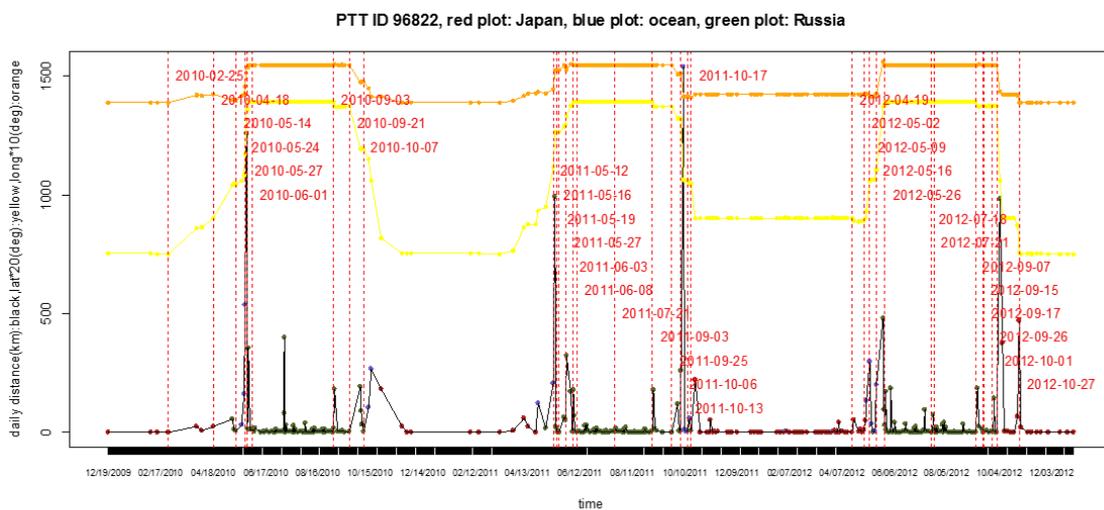


Figure 6. Classification of results

(A) Successful example by PTT ID 96822



(B) Not successful example by PTT ID 96830

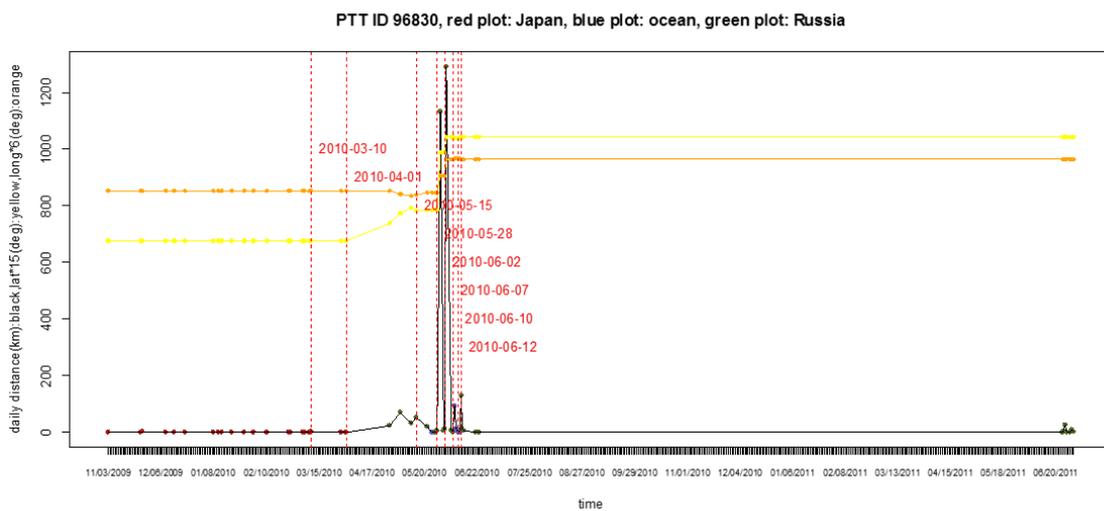


Figure 7. Examples of visualization plots



Figure 8. Classified migration status on the map

3.3 Accuracy assessment

Table 1 shows the results of accuracy assessment. The averaged difference of days between correct date and incorrect date was 2.68 ± 2.90 days, and 2.30 ± 1.34 days for migration status and stopovers, respectively.

	Migration status	Stopovers
Matched date	56 % (35)	31 % (66)
Incorrect date	31 % (19)	53 % (113)
Missing date	13 % (8)	17 % (36)

Table 1. Results of accuracy assessment
 The number in parenthesis indicates sample number.

4. DISCUSSION

4.1 General conclusion

In general, MATCHED method was proved powerful to identify the change points between migration statuses as well as stopovers. Nevertheless, identifying “exact” threshold dates is still challenging.

The threshold dates of migration status, which delineates wintering, spring migration, breeding, and autumn migration, are able to be identified using the MATCHED method. The accuracy was 87 % for the possible errors of 2.68 ± 2.90 days differences (Table 1).

The stopovers are also generally identifiable using the MATCHED method. However, compared to the migration status, the stopover identification is a bit more challenging. The accuracy of stopover identification was 84 % with possible errors of 2.30 ± 1.34 days (Table 1). The difficulties of automatic identification of stopovers are attributed to the complexity of its structure. As seen in Figure 3, conceptual model of migration status and stopovers is different, and model of stopovers are more complicated because it should take into account the duration of movement from one stopover and another stopover too.

The usage of automatic detection by “change point” analysis (step 3: automatic identification in Figure 1) depends on the purpose of application of research. If the researchers do not mind the error range of days which are mentioned above, automatic detection is considered fairly sufficient, which means researchers can omit step 6: manual correction in Figure 1. Still, step 4: visualization of change points, and step 5: interpretation are needed. On the other hand, if researchers are interested in detailed or accurate information on stopovers, e.g. duration of stopovers, exact geographic locations of one stopover, etc., manual corrections (step 6) after automatic detection would be recommended.

Even if manual corrections may be carried out, MATCHED method can contribute to save time and labour work drastically, because candidate threshold dates are automatically listed up so that researchers can check data only around those dates to confirm correct dates.

In addition, for applying MATCHED method, the length of tracked period does not matter. Rather, whether the threshold dates are included during tracked period is much important (see Figure 7 (A) (B)). Even if the tracked period is short, when the tracked period includes any period of threshold dates of migration status, MATCHED method should work.

4.2 Limitation

When manual correction is implemented, around the determined threshold dates of stopovers, the actual stopover periods are needed to be determined using geographical locations to confirm the proximity between consecutive points. Yet, density of points in one stopover, and distance between one stopover and another stopover could be different depending on target migratory birds. Hence, detecting and confirming stopovers' threshold dates could be more subjective than detecting threshold dates between migration statuses.

4.3 Application

Ornithologists are often interested in stage of migration of migratory birds, not only the geographic locations of these birds.

The MATCHED method can determine the migration status, e.g. wintering, spring migration, breeding, and autumn migration, as well as start and end of individual stopover site. Although interpretation of visualization plot requires some expertise, the MATCHED method can be expected to contribute to objective judgement of threshold dates of migration status for satellite-tracked migratory birds.

The demonstration of the MATCHED method was conducted by Argos service for Tundra Swans in this paper. The case of Tundra Swans is considered to have advantages which is suitable for MATCHED method as this species' geographic direction of migration was clear. This species started movement from Japan at south-west to north-east in Russia, which helped “change point” analysis to be implemented easier to enable detection of threshold dates just by changes of latitude. The MATCHED method is fundamentally applicable for any tracking data of migratory birds, but tracking data like satellite-tracked data is considered ideal because it can collect data by the same sampling frequency, e.g. every date. In order to discuss application for other migratory birds' application, more experiments will be needed.

ACKNOWLEDGEMENTS

This study was funded by the Ministry of the Environment of Japan.

REFERENCES

- Argos, 1996. User's manual. Service Argos, Inc., Landover, Md. CLS (Collecte Localisation Satellites), 2007. Argos User's Manual On Line.
- Coyne, M., Godley, B., 2005. Satellite Tracking and Analysis Tool (STAT): an integrated system for archiving, analyzing and mapping animal tracking data. *Marine Ecology Progress Series*, 301, pp. 1-7.
- Douglas, D.C., Weinzierl, R., C. Davidson, S., Kays, R., Wikelski, M., Bohrer, G., 2012. Moderating Argos location errors in animal tracking data. *Methods in Ecology and Evolution*, 3, pp. 999-1007.
- Eileen, R., 2010. Bewick's Swan (Poyser Monographs) Poyser, pp. 304.
- Higuchi, H., Pierre, J., 2005. Satellite tracking and avian conservation in Asia. *Landscape Ecol Eng*, 1, pp. 33-42.
- Killick, R., Eckley, I.A., 2014. changepoint: An R Package for Change point Analysis. *Journal of Statistical Software*, 58, pp. 1-19.
- R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Zeileis, A., Grothendieck, G., 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software* 14, 1-27.