

## 蝶モニタリングのためのデータ解析ツール

安川 雅紀<sup>†</sup> 須田 真一<sup>‡</sup> 鷺谷 いづみ<sup>‡</sup> 喜連川 優<sup>§</sup>

<sup>†</sup> 東京大学地球観測データ統融合連携研究機構 〒153-8505 東京都目黒区駒場 4-6-1

<sup>‡</sup> 東京大学大学院農学生命科学研究科 〒113-8657 東京都文京区弥生 1-1-1

<sup>§</sup> 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: <sup>†</sup> yasukawa@iis.u-tokyo.ac.jp,

<sup>‡</sup> s-suda@jcom.home.ne.jp, awashi@mail.ecc.u-tokyo.ac.jp, <sup>§</sup> kitsure@tkl.iis.u-tokyo.ac.jp

**あらまし** 我々は、生物多様性における市民参加型によるモニタリングデータのための Web サービスシステムを構築中である。モニタリングによって得られる東京の蝶のデータに関するデータアップロードツール、データ編集・品質管理ツール、データ公開ツール等を既に開発したが、本研究では上位層であるデータ解析ツールを開発した。本ツールは、データから表や図、地図を作成して、蝶の生息分布やその変化を知ることができるような基本解析が可能である。保全生態学の研究者による実利用のフィードバックから本ツールの改良および拡張を行っている。

**キーワード** 生物多様性, 市民参加型モニタリング, 科学データ, データ解析

## A Data Analysis Tool for Butterfly Monitoring

Masaki YASUKAWA<sup>†</sup> Shin-ichi SUDA<sup>‡</sup> Izumi WASHITANI<sup>‡</sup> and Masaru KITSUREGAWA<sup>§</sup>

<sup>†</sup> Earth Observation Data Integration and Fusion Research Initiative, the University of Tokyo

4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

<sup>‡</sup> Graduate School of Agricultural and Life Sciences, the University of Tokyo

1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-8657 Japan

<sup>§</sup> Institute of Industrial Science, the University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

E-mail: <sup>†</sup> yasukawa@iis.u-tokyo.ac.jp,

<sup>‡</sup> s-suda@jcom.home.ne.jp, awashi@mail.ecc.u-tokyo.ac.jp, <sup>§</sup> kitsure@tkl.iis.u-tokyo.ac.jp

**Abstract** We are constructing a Web service system for citizen-based monitoring data on biodiversity. In this process, various tools such as data upload, data edit, data quality control and data publication are had developed for butterfly monitoring in Tokyo metropolitan, already. In this paper, we describe a data basic analysis tool newly developed. In this tool, generating the table, the chart and the map from the monitoring data, users can understand habitat distribution of butterfly and its time series change. Moreover, this tool is being improved by the feedback from the researchers of conservation ecology.

**Keyword** Biodiversity, Citizen-based monitoring, Science data, Data analysis

### 1. はじめに

1992年にブラジルのリオ・デ・ジャネイロで開かれた地球サミットで「生物多様性条約」が採択されて以降、健全な生態系を維持し、持続可能な社会を築くためのキーワードとして「生物多様性」という言葉が多く使われてきた。2010年に名古屋で開催された生物多様性条約第10回締約国会議(COP10)では「愛知ターゲット」が採択され、生物多様性のデータや情報の収集、解析やそれによって得られる新しい知識の共有の強化についても記載された[1]。

上記のターゲットや政策に貢献するには、まず、生物多様性の現状や傾向を知る必要があり、モニタリングによる大量のデータ取得が不可欠である。モニタリ

ングする手法としては様々な方法があるが、市民参加型モニタリングは、大量のデータが得られ、時間的・空間的に高密度の広域モニタリングが可能である[2]。一般の調査員が調査マニュアルに従って定期的に生き物を調査して調査用紙に記入し、事務局に提出する。しかし、一般の調査員は生き物調査に関する知識が少ないことが多く、調査に参加した初期の頃は種の同定での誤判別や調査項目の記入ミスが多い。データが紙媒体であったり、データに誤判別・記入ミスが含まれていたりするため、このデータを生態学の解析に用いるには、データのデジタル化やクレンジングが必要であり、データ利用者にかかる負担は大きかった[3]。

また、デジタルデータが大量に収集された際には、

そこから新しい知見が容易に得られる様な高価値の解析ツールが不可欠である。従来、生態学の研究者は、表計算ソフトやGISソフトなどを用い手間暇かけて解析結果の図を作っていた。しかし、取得されるデータが多くなればなるほどデータの整理やハンドリングに要する時間が増え、研究者が本来行うべき解析の時間が減る傾向があった。そのため、データ管理ツールとデータ解析ツールが同一プラットフォーム上にあり、データ解析ツールが容易なユーザビリティかつ高機能であることが必要である。

現在、我々は保全生態学の研究者および市民生協の一つであるパルスシステム東京との協働で、東京の蝶のモニタリング活動を行っている。蝶は幅広い環境に生息し一般市民に馴染み深く種を識別し易いことから、モニタリングターゲットとして採用している。IT側のミッションとして、データアップロードツール、データ編集・品質管理ツール、データ公開ツール、データ解析ツールを有するシステムを構築中である。このなかで、データアップロードツール、データ編集・品質管理ツール、データ公開ツール等を既に開発した[4]。

上記の開発をさらに進め、本論文では、東京の蝶のモニタリングデータのためのデータ解析ツールを開発したので報告する。本ツールは、約2万件の品質管理された信頼性の高い調査データを用い、データベースからのデータ抽出・データセット作成・データダウンロード機能、データのテーブル表示機能、データのグラフ表示機能、データのマップ表示機能等を実装し、Web上で生態学における基礎解析が行えるようにした。本ツールがユーザによる新しい知見の発見をサポートし、研究の作業効率が高まることを目指す。

## 2. 関連研究

生物多様性に関するデータベースは、分類学的ライブラリや知識情報のライブラリや、モニタリングのデータを蓄積したデータベースが存在する。

ESABII Databaseは様々な種について学名やDNA情報、標本情報と標本画像、分布情報を収録したデータベースで、地球規模生物多様性情報機構(GBIF)の日本ノードを担当している。百科事典的要素が強く、データ解析を有する本研究のようなツールとは異なる[5]。

NaGISA (Natural Geography In Shore Areas: なぎさ) Databaseは海洋生物センサスの野外研究プロジェクト(Census of Marine Life: CoML)で、海辺を調査してCoMLのデータベースOBISに登録するもので、データの品質管理はされておらず、品質はデータ提供者次第である。データ解析ツールは考慮されていない[6]。

英国は市民参加型の生物モニタリング活動が非常に活発な国として知られている。1995年から英国蝶類

保全協会では、国土全体を10km区画に分けてボランティアが蝶の種類や数を数える調査を行っており、データ量は2009年までの15年間で500万件以上をアーカイブサイトで公開している。画像アップロードやデータ品質管理の機能は有しておらず、データ品質は調査員毎の習熟度に依存し、データ全体としては均一な品質ではなく、データの可用性に欠点があり、データ解析ツールはない[7]。

## 3. ツールの機能要件および設計指針

本章では蝶のモニタリングで取得されるデータについて述べ、そのデータを用いた提案ツールの実現のための機能要件および設計指針について述べる。

### 3.1. 調査データ

調査員は、調査マニュアルに従って蝶の調査を行う。その際、調査データは、図1のように、調査項目と画像の組で記録する。

調査項目について、天候、蝶の性別、蝶の行動等は選択肢の中から該当するものを選ぶ。各調査項目の記録方法は、調査員自身の調査スタイルを考慮して規定せず、紙、携帯電話のメモ帳やボイスレコーダー等、自由である。蝶の画像はデジタルカメラで撮影する。画像は、蝶の羽の表、蝶の羽の裏、周辺環境の3枚を記録することが望ましいが、いずれか1枚でも、撮影が困難な場合は0枚でも可である。デジタルカメラでの撮影方法は、特に指定せず、羽の模様を分かりやすく撮影することを目標とし、解像度は自由で、携帯電話での撮影も可としている。複数の蝶の存在を確認した場合は、その個体数分の調査データを作成する。



図1 調査データ

### 3.2. 機能要件

蝶モニタリングのためのデータ解析ツールの開発において、下記のニーズがある。

品質の高い調査データを用いて、蝶の生息に関する解析をデータがアーカイブされているシステム上で行

いたいという、保全生態学の研究者からのニーズがある。従来は表計算ソフトや GIS 系ソフトがインポートできるように調査データのフォーマットを成型し、それらソフトにインポートして試行錯誤しながら表やグラフ、マップ等に加工していたが、これは手間がかかるので、データをアーカイブしたシステム内で簡単な操作で効率的に生態学の基礎解析を行いたい、という理由があるためである。基礎解析では、「年」対「種名」の存在確認件数の表、存在確認件数の時系列グラフ、存在確認種数の時系列グラフ、「区市町村」対「種名」の存在確認件数表、種名毎の存在確認マップ等を手軽に表示できる必要がある。

### 3.3. 設計指針

前節の機能要件をもとに、解析ツールを設計する。調査員や保全生態学の研究者側でデータをアーカイブし一括管理することは、データ量が膨大であることからみて非現実的である。近年のインターネットの普及を考慮し、データは大規模なストレージ基盤のデータベース上に保存することを基本とし、データ解析機能は Web 上で行うことができるようにする。なお、データベースのスキーマは、下記の設計指針を満たすよう決定し、次章でスキーマの詳細を述べる。ユーザはそれぞれの機能ページにて所望のパラメータを設定してデータのリクエストを行う。データベースから検索されたデータに対して可視化処理を施し、Web ブラウザに結果を表示する。前節の機能要件に対して、下記のように設計する。

データ解析ツールは、品質管理された信頼性の高い調査データを用いて Web 上で生態学における基礎解析が行えるようにする。本ツールには、データベースからのデータ抽出・データセット作成機能、データのテーブル表示機能、データのグラフ表示機能、データのマップ表示機能が含まれる。機能の詳細は以下の通りである。

データベースからのデータ抽出・データセット作成機能では、年月の範囲、空間範囲、種名を条件にクエリを発行してデータベースから該当する調査データを抽出し、データセットのナイスなリスト表示のための処理をオンラインのサーバ上で行う。リスト表示では、調査データ内の概要(日時、種名、調査地情報等)を表示し、リンクによってそのデータの詳細が表示できるようにする。また、データセットを CSV 形式で一時的作業領域に保存する。作成したデータセットファイルは、ユーザがローカル PC にダウンロードすることも可能とする。

存在確認テーブル表示機能では、年月、空間範囲を検索条件として、データ抽出・データセット作成機能のデータ抽出モジュールを用いてデータを抽出し、デ

ータのテーブル形式表示のための処理をオンラインのサーバ上で行う。テーブル形式は、「年」対「種名」の存在確認件数表示である。「年」対「種名」の存在確認件数のテーブル表示では、表の対象カラムについて調査データの有無を表示し、調査データの件数を表示する。調査データが存在するカラムはリンクによって、該当する調査データのリストを表示し、データの詳細を閲覧できるようにする。

グラフ表示機能では、空間範囲を条件として、データ抽出・データセット作成機能のデータ抽出モジュールを用いてデータを抽出し、時系列グラフ表示のためのグラフ化処理をオンラインのサーバ上で行い、単位およびタイトルを埋め込んで、Web ページに結果を表示する。時系列グラフは、単位時間毎の存在確認件数のグラフ、あるいは存在確認種数のグラフが選択できるようにする。このグラフ表示では、FusionCharts を使う。これはインドの FusionCharts Technologies LLP が開発した Flash のグラフ描画コンポーネントであり、2次元や3次元グラフに関する様々な表示機能を持っている[8]。グラフ上で、マウスカーソルをプロット点に置くと、その点の値を表示し、その点をクリックすると、リンクによって対象データのリストを表示し、データの詳細を閲覧できるようにする。

種名毎の存在確認マップ機能では、種名と年を条件として、データ抽出・データセット作成機能のデータ抽出モジュールを用いてデータを抽出し、区市町村毎に仕分けし、Google Maps 上に区市町村の境界線をポリゴンで描き、存在確認件数が1件以上の区市町村は白以外の色で境界線内部を色塗りし、存在確認件数が0件の区市町村は白色で境界線内部を色塗りする。マップ上では、区市町村の領域をマウスでクリックすると、その区市町村での存在確認件数を吹き出しで表示し、吹き出し内のリンクをクリックすると、対象の調査データのリストを表示し、データの詳細を閲覧できるようにする。

## 4. システムの構築

### 4.1. データベースのスキーマ

本プロジェクトでは、調査員が調査した蝶のデータを、Web の入力ページに調査員が入力してサブミットし、スクリプト処理によって図2のリレーショナルデータベースの調査データテーブルに格納する。なお、調査データテーブルにおいて、調査員 ID、天気、風、種名、調査員の種名同定正否、蝶性別、蝶行動、蝶画像有無、データ公開フラグ、編集レベルの項目については入力値の種類が限られており正規化する。これらの項目は外部キーとし、それぞれの項目に対してテーブルを用意し、各テーブルの主キーである ID を指

定する。例えば、図2の調査員データテーブルの天気  
の項目は外部キーであり、天気テーブルの天気 ID を  
指定する。天気テーブルは主キーとして天気 ID を持  
つ。

天気、風、蝶性別、蝶行動の各テーブルの表記の種  
類については、蝶のモニタリングの専門家の意見によ  
って決めた。調査場所の住所は、個人情報保護する  
ことを考慮して町名や丁目までの公開可能な住所と、  
オリジナルデータである詳細な住所の、2つの住所を  
保存することとした。調査場所の緯度経度については、  
国土交通省国土政策局の街区レベル位置参照情報のフ  
ァイルをダウンロードし、データベースに格納し、調  
査員が入力した住所を自動的に緯度経度に変換して、  
調査データテーブルに入力している[9]。

調査員の種名同定正否は、調査員による種名同定が  
専門家による種名同定と一致したかどうかを保存する。  
これは、調査データ公開後に調査員に対して種名同定  
の学習に役立てるために項目を設定した。画像につい  
ては、画像のパス、画像のファイル名をそれぞれの項  
目に保存する。調査データに画像が存在する場合は、  
「画像あり」として、調査データテーブルの蝶画像有  
無に「1」を入れる。

データ公開フラグは、蝶の専門家によるデータ品質  
検査によって与えられる。データ解析に使用できるデ  
ータは「公開(信頼高)」とし、解析には使用できない  
が調査員へのフィードバックとして公開するものは  
「公開(信頼低)」とし、蝶以外の生き物については「非  
公開」、データ不備やデータ重複については「データ削  
除」に区別する。

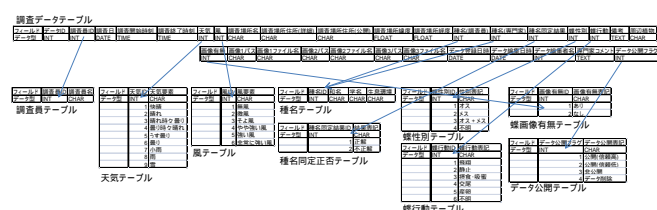


図2 データベースのスキーマ

## 4.2. システム構成

図3は、開発中の蝶データのWebサービスシステム  
の概念図であり、本ツールを「データ解析ツール」と  
して、このシステムに組み込む。このシステムは、蝶  
のデータと気候データとの統合など様々な機能の将来  
的な拡張を考慮し、データ統合・情報融合コアシステ  
ムのプラットフォーム上で稼働中である。データ統  
合・情報融合コアシステムは、ペタバイト級ストレ  
ージに地球環境に関する様々なデータをアーカイブし上  
層に多様なアプリケーションを構築して統合・解析す

ることで、データを科学的・社会的に有用な情報に変  
換できることを目指している[10]。

本データ解析ツールの利用は、図3の「データ解析  
ツール」のページにアクセスする。このページでは、  
データの解析の方法を選ぶ。データ解析方法は、デー  
タ抽出、存在確認テーブル表示、存在確認グラフ表示、  
存在確認マップ表示がある。

データ抽出では、検索条件として、種名、年月、蝶  
画像有無、空間範囲が設定でき、これらを用いてクエ  
リを生成して、データベースの調査データテーブルか  
ら抽出する。抽出されたデータは表形式のhtmlにペ  
ージを自動生成してクライアントに結果を返す。デー  
タをクリックすると、データIDを用いてデータ詳細表  
示のページに移ることができ、データの詳細を表示す  
ることができる。データ抽出結果のページにはデータ  
ダウンロードのリンクをはり、クリックしたときは、  
上記の抽出結果をCSV形式に成型してファイルを作  
成し、ファイルをクライアントに送信する。

存在確認テーブル表示では、検索条件として、年月、  
空間範囲が設定でき、これらを用いてクエリを生成し  
て、データベースの調査データテーブルから抽出する。  
抽出されたデータは種名と年によってグルーピングを  
行い、1件以上であれば「○」とその件数を表示し、0  
件であれば「x」を表示する。表形式のhtmlにペ  
ージを自動生成してクライアントに結果を返す。「○」の部  
分はリンクをはり、データ抽出機能によって、該当す  
る調査データを閲覧することが可能である。

存在確認グラフ表示では、検索条件として、空間範  
囲が設定でき、これらを用いてクエリを生成して、デ  
ータベースの調査データテーブルから抽出する。抽出  
されたデータは種名と年によってグルーピングを行い、  
蝶全体あるいは蝶の科毎の「年」対「存在確認件数」  
あるいは「年」対「存在確認種数」のテーブルにまと  
め、そのテーブルからグラフ作成の скрипт を実行  
し、クライアントに結果を返す。

存在確認マップ表示では、検索条件として、蝶の種  
名と年が設定でき、データ抽出機能を用いてデータを  
抽出し、区市町村毎にデータをグルーピングし、対象  
の種名の存在が確認された区市町村の区域を赤く塗り、  
存在確認件数が0件の場合は区域を白く塗る。ただし、  
白地図とその色付けはGoogle Maps上にポリゴンで描  
く。マウスで区市町村をクリックすると吹き出しが現  
れ、区市町村名とその区域での存在確認件数を表示す  
る。また、その部分にリンクをはり、該当する調査デ  
ータを表示する。

Webの機能を利用した解析結果表示において、表の  
各項目、グラフのプロット点、マップ上の対象地域等  
からリンクによってデータの詳細を閲覧できることが、

既存の表計算ソフトや GIS 系ソフトにはない機能であり、興味のある部分について詳細を素早くチェックできるため、データ解析におけるユーザの作業スピードを速める効果が期待できることが本機能の特徴である。

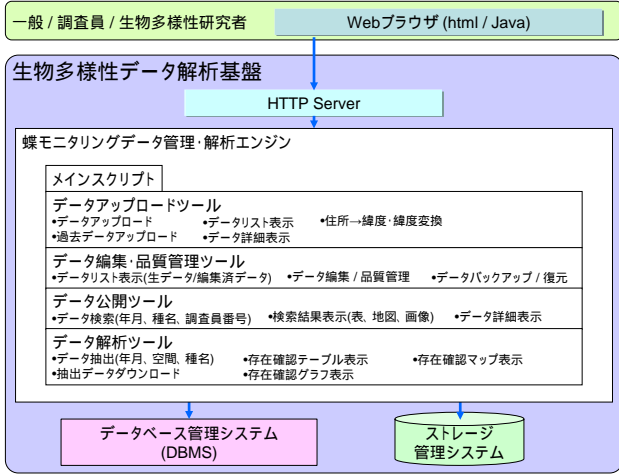


図 3 蝶データ Web サービスシステムの概念図



図 4 データ抽出とデータダウンロード

## 5. ツールの利用

本章では登録された調査データを用いた蝶の解析例について述べる。

図 4 は、図 3 のデータ解析ツールのデータ抽出機能を用いた、データの抽出例である。検索条件として、種名は「全て」、年月は「2011 年全て」、蝶画像有無は「全て」、空間範囲はマップ上で新宿御苑をクリックした場所を中心とした 1km 四方である。地図の下方にデータベースから抽出した調査データが表形式で表示された。対象区域での調査状況を知るのに便利である。虫眼鏡のアイコンをクリックすると調査データの詳細を閲覧することが可能である。また、地図と表の間にはデータダウンロードのリンクがあり、図 3 の抽出データダウンロード機能により、表で示されたデータをダウンロードすることが可能である。

図 5 は、図 3 のデータ解析ツールの存在確認テーブル表示を用いた種名毎の存在確認件数を示す。行が種名、列に年毎の存在確認件数であり、表のタブは種名を科で分類した。調査でその種が確認されれば「」とその件数を表示し、確認されなければ「x」を表示する。種によって生息できる環境が異なるため、蝶と温暖化・都市化との関係を調査することが可能である。「」の部分はリンクをはり、データ抽出機能によって、該当するデータのリストを閲覧することができる。デフォルトは全空間範囲が対象となるが、範囲を緯度・経度で指定すると、対象範囲での蝶の生息件数とその時系列変化を知ることができる。



図 5 種名毎の存在確認件数表

図 6 は、図 3 のデータ解析ツールの存在確認グラフ表示を用いた、存在確認種数の時系列グラフを示す。縦軸は、対象の空間範囲の蝶の存在確認種数であり、横軸は年であり、横軸の右端は対象期間の累積である。この累積は、ユーザのニーズにより、表示するよう表示ツールを改良した。このグラフにより、対象の空間範囲における、蝶の多様性の状態と時系列変化を知ることができる。

象の種名が生息できると言われている環境と、実際の生息分布との比較や時系列変化等の解析が可能である。

## 6. まとめ

本研究では、東京の蝶のモニタリングデータのためのデータ解析ツールを開発した。本ツールは、約2万件の品質管理された信頼性の高い調査データを用い、データベースからのデータ抽出・データセット作成・データダウンロード機能、データのテーブル表示機能、データのグラフ表示機能、データのマップ表示機能等を実装し、Web上で生態学における基礎解析が行えるようにした。前章のツールの利用では、保全生態学の研究者から概ね好評を得たが、保全生態学の分野における新しい知見が得られるよう改良を重ねている。

今後は、解析ツールの拡張、蝶以外の生物モニタリングデータへの適用等を行っていく。また、システムの改良を重ねていくことで、市民参加型の生物多様性モニタリングデータに対するクラウド・ソーシングのフレームワークを具体化させる。

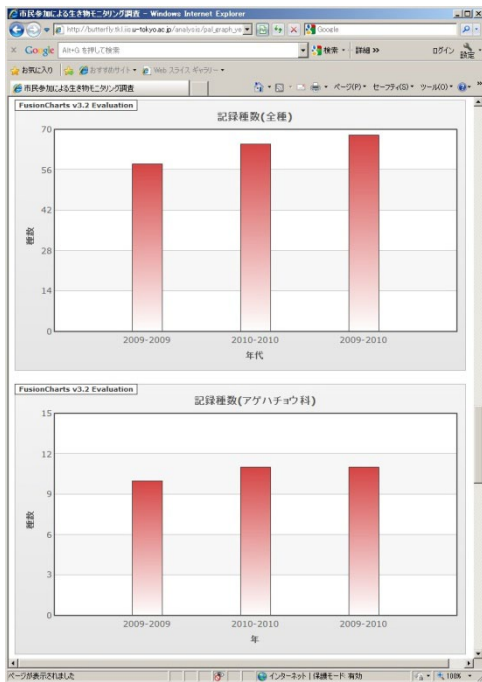


図6 存在確認種数の時系列グラフ

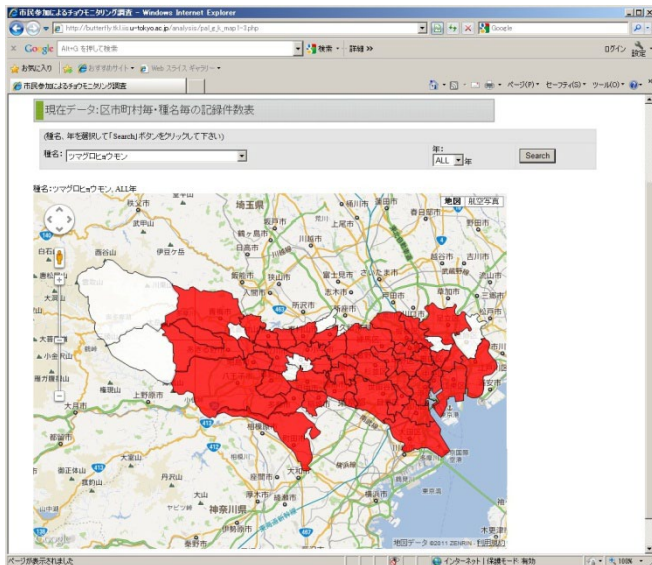


図7 生息分布(マップベース)

図7は、図3のデータ解析ツールの存在確認マップ表示を用いた、区市町村毎の存在確認マップを示す。表示の設定条件として、蝶の種名「ツマグロヒョウモン」と年「ALL(2009-2011)」を指定すると、対象の種名の存在が確認された件数が1件以上の区市町村の区域は赤く、0件の場合は白く塗られた。マウスで区市町村をクリックすると吹き出しが現れ、その区域での存在確認件数を表示し、その件数のリンクによって、該当するデータを表示できる。このマップにより、対

## 文 献

- [1] Convention on Biological Diversity (CBD), "Report of the Tenth Meeting of the Conference of the Parties to the Convention on Biological Diversity", <http://www.cbd.int/doc/meetings/cop/cop-10/official/cop-10-27-en.pdf>, pp. 118-120, 2010.
- [2] J. Silvertown, "A new dawn for citizen science", *Trends in Ecology & Evolution*, Vol. 24, No. 9, pp. 467-471, 2009.
- [3] 藤原宣夫, 日置佳之, 須田真一, "-MBR方式による- 住民参加の生きもの調査ガイドブック", 国土技術政策総合研究所資料, No. 139, 2003.
- [4] 安川雅紀, 前角達彦, 須田真一, 中原美理, 鷺谷いづみ, 喜連川優, "蝶モニタリングのためのデータ管理システム", 第3回データ工学と情報マネジメントに関するフォーラム (DEIM2011), C8-6, 2011.
- [5] East and Southeast Asia Biodiversity Information Initiative (ESABII), "ESABII Database", <http://www.esabii.org/>, 2010.
- [6] The Natural Geography In Shore Areas Project (NaGISA), "NaGISA Database", <http://www.esabii.org/>, 2011.
- [7] The UK Butterfly Monitoring Scheme (UKBMS), <http://www.ukbms.org/>, 2011.
- [8] FusionCharts Technologies LLP, "FusionCharts v3 - Stunning Charts in Flash & JavaScript (HTML5) for Web & Enterprise", <http://www.fusioncharts.com/>, 2011.
- [9] 国土交通省国土計画局, "位置参照情報ダウンロードサービス", <http://nlftp.mlit.go.jp/isj/>, 2007.
- [10] Data Integration & Analysis System (DIAS), <http://www.editoria.u-tokyo.ac.jp/dias/>, 2011.