

# 海洋観測データの良否識別を目的とした条件付確率場における素性関数の自動設計の試み

An Attempt to Design Feature Functions in Conditional Random Field for Error Detection of Ocean Observation Data

上川路 洋介\*<sup>1</sup> 松山 開\*<sup>1</sup> 福井 健一\*<sup>2</sup> 細田 滋毅\*<sup>3</sup> 小野 智司\*<sup>1</sup>  
Yosuke Kamikawaji Haruki Matsuyama Ken-ichi Fukui Shigeki Hosoda Satoshi Ono

\*<sup>1</sup>鹿児島大学大学院 理工学研究科 情報生体システム工学専攻

Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University

\*<sup>2</sup>大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

\*<sup>3</sup>独立行政法人 海洋研究開発機構

Japan Agency for Marine-Earth Science and Technology

This paper proposes a method for error detection on observation data of globally-covered ocean monitoring system Argo. The proposed method utilizes Conditional Random Field (CRF) to assign quality labels with automatically-designed feature functions that aim to detect all types of observation errors, whereas previous work focused on specific types of the errors. To composite complex feature functions considering dependency between explanatory variables, decision tree is utilized in the proposed method. Experimental results showed that the proposed method especially detects observation errors extending over many observation layers.

## 1. はじめに

海洋の長期的な変動を把握することを目的として、海洋の水温と塩分の自動計測が可能なアルゴフロートを投入することにより、全球海洋データのリアルタイムな観測を行う全球海洋監視システム「アルゴ」が2000年より稼働している。アルゴフロートにより自動観測されたデータには観測エラーを含むことがあるため、品質管理によって観測の信頼度を表す品質管理フラグが観測値ごとに割り当てられる。品質管理は2段階で行われる。まず、品質管理手法 [Argo 02] により品質管理ラベルが自動的に割り当てられる。しかし、水温や塩分などは自然変動の影響により変化するため、観測エラーとの切り分けが困難であり、既存の自動品質管理では観測エラーの見落としや誤検出が発生している。このため、第2段階として専門技術者が目視で確認を行い、手動で補正を行わなければならない負担となっている。しかし、専門技術者ごとの補正基準の相違や心的要因による判断の揺らぎ、補正スキルを持ち合わせていない国の存在などにより、全球データの品質の均一性が担保できない点も課題となっている [細田 13]。

著者らは、海洋データのエラー検出・指標決定の問題を系列ラベリング問題として捉え、条件付き確率場 (Conditional Random Field: CRF) により自動識別する方式を提案している [Ono 15]。上記の方式では、専門技術者へのヒアリングにより素性関数を設計し、さらにサポートベクタマシン (Support Vector Machine: SVM) [Cortes 95] を用いて閾値を決定することで、高精度で観測エラーの検出を行えることを示したものの、対応できる観測エラーの種類に限られる点に問題があった。

本研究では、観測エラーの種類を限定せずに検出が可能となるよう、CRFの素性関数を自動的に構築することを試みる。観測エラーを検出するためには、単純で多数の素性関数を用意するのではなく、複数の説明変数に関する条件式を構築する必要があることが先行研究で示されている [Ono 15]。このため、提案方式では決定木学習を利用することで、複数の条件式の組

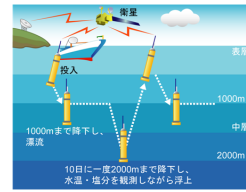


図 1: 観測サイクル

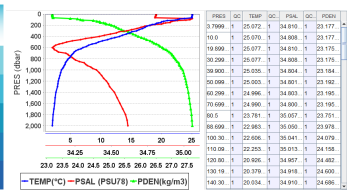


図 2: プロファイルの例

み合わせからなる素性関数を作成する。作成した条件式を用いて点推定を行うよりも、CRFの素性関数として利用し系列推定を行うことでより高い精度で観測エラーを検出できることを示す。

## 2. 全球観測システム：アルゴ

### 2.1 概要

全球海洋監視システムアルゴは、2000年より全球観測データのリアルタイムな取得を目的として開始された [Roemmich 01]。このプロジェクトでは、全球アルゴ観測網を実現するために、アルゴフロートと呼ばれる自動昇降可能な海洋観測ロボットを海へ投入し、海洋内部の水温や塩分の自動観測を行っている。観測データが衛星を介して地上局に送られると、品質管理が施され、インターネットを通じて公開される。現在、このプロジェクトには全世界で30カ国以上が参加し、アルゴフロートは3,500台以上が常に稼働している。現在までに100万点を超えるプロファイルの蓄積に成功しており、従来知り得なかった地球規模の変動が捉えられ、気候変動のメカニズム解明に向けて研究が進められている。

### 2.2 海洋観測

アルゴフロートによる観測サイクルを図1に示す。アルゴフロートは海に投入されると、海流の影響が弱い水深1,000[m]で漂流する。観測時になると、水深2,000[m]付近まで降下し、水温と塩分を観測しながら海面まで浮上する。1回の浮上によって生成される観測データはプロファイルと呼ばれ、プロファイルには各観測層における水温値および塩分値が記録され

る。アルゴフロートは、この観測サイクルを 10 日間隔で自動的に行う。プロファイルの例を図 2 に示す。縦軸は圧力 [dbar] であり、これは水深 [m] とおよそ同義である。青色のグラフは水温 [°C]、赤色のグラフは塩分 [PSS-78]、緑色のグラフは水温と塩分から算出される密度 [kg/m<sup>3</sup>] を表す。

### 2.3 品質管理

アルゴにおける品質管理では、アルゴフロートによって観測されたプロファイルに対して観測値の信頼度を決定する。品質管理フラグの種類として、1(正)、2(おそらく正)、3(おそらく誤)、4(誤) の 4 段階の信頼度が原則として使用される。

品質管理には、即時品質管理 (Real-time quality control: RQC) と遅延品質管理 (Delayed-mode QC: DQC) の 2 種類がある。RQC はプロファイルが観測されてから 24 時間以内に観測データを公開することを目的として、簡易的に行われる品質管理である。リアルタイムなデータの公開を優先するため、技術者による目視確認が行われないことが多い。一方、DQC は研究や解析用のデータとして提供することを目的として行われる高精度な品質管理である。専門技術者により観測されたプロファイルや近傍で観測されたプロファイルとの比較などを含めて目視確認が行われ、必要に応じて補正が行われる。

### 2.4 観測エラーの分類と先行研究の問題点

アルゴフロートによる観測データのうち、塩分濃度に関するエラーの分類例を表 1 に示す。なお、表 1 の分類は、海洋学的な知見に基づく分類ではなく、本研究において便宜上行った分類であることに留意されたい。また、表 1 に示したプロファイル数は、北太平洋域で日本によって遅延品質管理が行われたプロファイルの数を表している。

表 1 における密度逆転とは、深度とともに単調増加するはずの密度が逆転し、連続する観測層において深い層の方が密度が低い値が観測されたエラーを指す。同値エラーとは、アルゴフロートが観測する際に、電圧低下により観測が行われず、直前に観測した値がそのままコピーされる観測不良を指す。外れ値とは、観測した値が正常な範囲を外れている場合を指す。これらの観測エラーでは、単層にのみ観測エラーを表す品質管理ラベル 4 を割り当てることが多い。

先行研究 [Ono 15] では、\*印が付与された観測エラーのみを対象として、専門技術者へのヒアリングに基づいて手動で素性の設計を行った。対象とするエラーに対しては高い精度で品質管理ラベルを割り当てることができた。一方で、事前に対象とする観測エラーを人手により明確化し、各観測エラーに対応する素性を設計する必要がある。さらに、観測エラーを定義、分類することは専門技術者であっても困難であることに加え、対象とする海域毎に素性を個別に作成するか、閾値などの調整が必要となる。

## 3. 提案手法

### 3.1 基本アイデア

本研究では、海洋観測データに対する品質管理フラグの良否判別を系列ラベリング問題と捉え、以下のアイデアに基づく系列ラベリング方式を提案する。

方針 1: 条件付き確率場 (Conditional Random Field: CRF) [Lafferty 01] を用いる。隠れマルコフモデル (Hidden Markov Model: HMM) [Seymore 99] と比較して、CRF は様々な素性を同時に考慮できる点に特徴がある。提案手法では先行研究と同様に CRF を利用することで、多様な属性を参照しつつ、前後のラベルの組み合わせを考慮した良否識別を行うことができる。

表 1: 観測エラーの分類

エラーの種類	プロファイル数	内訳	プロファイル数
密度逆転	1,253	負側密度逆転 *	1,039
		正側密度逆転 *	203
		正負両側	11
同値 *	220		
密度逆転と同値が混在 *	89		
外れ値 *	352		
水圧エラー	1,282		
浅層エラー	2,156		
全体層エラー	1,494	ぎざぎざ	63
		オフセット *	660
		センサ経時変化	223
		外的要因	3
		その他	545
10 層以下の連続エラー	1,655		
原因不明	134		
ラベル 9	328		
合計	8,963		

方針 2: 決定木学習を用いて CRF の素性関数を自動的に設計する。先行研究では、密度逆転等の明確に定義が可能な観測エラーのみを対象としており、各観測エラーに対して専門技術者に聞き取りを行い、素性関数を手動で設計した。その結果、観測エラーの有無が観測値単体ではなく水深などに依存することがわかり、また、アルゴフロートの観測データは実数であるため、素性関数は複数の等式または不等式の組み合わせで表現すべきであることがわかった。一方で、先行研究の方法では観測エラーの種類毎に手動で素性関数を設計する必要があり、素性関数を設計する労力を要する。このため、本研究では、決定木学習を行うことで CRF の素性関数の自動設計を行う。決定木学習により、複数の属性を考慮した複雑な素性関数を自動的に構築することが可能となる。また、決定木を直接用いる点推定ではなく、CRF に基づく系列推定を行うことで、連続する観測層にまたがる観測エラーを精度良く推定することが可能となる。

### 3.2 CRF による系列ラベリング

CRF は系列ラベリング問題に用いられる識別モデルであり [Lafferty 01]、系列データやラベルの前後の依存関係の特徴を素性として記述できるため、多様な属性を参照しつつ、前後のラベルの組み合わせを考慮した良否識別を行うことができる。CRF ではラベル間の依存関係に応じて様々なモデルが存在するが、本稿では入力データと 1 つ前のラベルに依存する Linear-chain モデルを前提に説明する。

系列の長さが  $T$  である入力データ  $\mathbf{x} = (x_1, x_2, \dots, x_T)$  が与えられたとき、出力ラベル  $\mathbf{y} = (y_1, y_2, \dots, y_T)$  となる条件付き確率  $P$  を式 (1) のようにモデル化する。

$$P(\mathbf{y}_t, \mathbf{y}_{t-1} | \mathbf{x}) = \frac{1}{Z_{\mathbf{x}}} \exp \left( \sum_{t=1}^T \sum_k \lambda_k f_k(\mathbf{x}, y_t, y_{t-1}) \right) \quad (1)$$

ここで、 $f_k$  は素性関数、 $\lambda_k$  は素性関数  $f_k$  に対する重み、 $z_{\mathbf{x}}$  は  $\sum_{\mathbf{y}} P(\mathbf{y}_t, \mathbf{y}_{t-1} | \mathbf{x}) = 1$  を保証する正規化係数である。素性関数  $f_k$  は識別器を学習させる際に、識別の情報として与える特徴量であり、式 (2) のように定義される。

$$f_k(\mathbf{x}, y_t, y_{t-1}) = \begin{cases} \phi_k & \text{if condition} = \text{true} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

ここで、 $\phi_k$  は素性値と呼ばれる実数値であり、素性に依って任意に設定できる。素性関数は入力  $\mathbf{x}$  および  $t$  番目と  $(t-1)$  番目の出力ラベルに依存しており、問題に応じて設計する。

学習では入力データ  $\mathbf{x}$  と出力ラベル  $\mathbf{y}$  を 1 組とする学習データ  $\mathbf{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$  を与え、最尤推定によりパラメータである  $\lambda_k$  を求める。条件付き確率  $P$  の最尤推定をとった目的関数  $L(\mathbf{D})$  を最適化する。

$$L(\mathbf{D}) = \sum_{\mathbf{D}} \log P(y_t, y_{t-1} | \mathbf{x}) - \sum_k \frac{\lambda_k^2}{2\sigma^2} \quad (3)$$

ここで、式 (3) の第 2 項は過学習を防ぐための正規化項である。目的関数  $L(\mathbf{D})$  を最適化する重みパラメータ  $\lambda_k$  を推定するために、最急勾配法により重みを更新する。

学習した識別器を使って、入力データ  $\mathbf{x}$  に対する出力ラベル  $\mathbf{y}_{\text{out}}$  を決定する際は、式 (4) に示す最大化問題を解くことにより決定することができる。

$$\mathbf{y}_{\text{out}} = \underset{\mathbf{y}}{\operatorname{argmax}} P(\mathbf{y} | \mathbf{x}) \quad (4)$$

### 3.3 決定木学習による素性関数設計

決定木は予測モデルの一種であり、決定木を学習データから構築することで分類や回帰を行うことができる。決定木では、根を起点とし、(根を含む) 節に対応する説明変数の値によって分岐を繰り返し、到達した葉によって目的変数の予測値を得る。分類問題では、葉はクラスを表し、枝がそのクラスに至るまでの特徴の集合を示す。

学習を行う際は、学習データにおいて目的変数を最もよく分類する説明変数の分岐を生成し、根とする。上記の分類の基準として、例えば、以下に示すような Gini 不純度が利用される。

$$I_g = 1 - \{P(c_1 | v = a_1)^2 + P(c_2 | v = a_2)^2\} \quad (5)$$

ここで説明変数  $v$  が値  $a_1, a_2$  をとる場合に目的変数が  $c_1, c_2$  となるグループの生起確率をそれぞれ  $P(c_1 | v = a_1), P(c_2 | v = a_1)$  とする。次に、分割条件の良し悪しを評価するために、分割前後の  $I_g$  の変化を求める。上記のような Gini 指標を全変数について算出し、その値がもっとも低くなる変数を根として採用する。以上の処理を再起的に繰り返すことで木を構築する。これ以上改善されない状態まで、または、あらかじめ決められた終了条件を満たすまで節の追加を繰り返す。

## 4. 評価実験

### 4.1 実験設定

提案方式の有効性を検討するため、北太平洋域での観測プロフィールを用いて実験を行った。観測データには 1 (正) から 4 (誤) までの品質管理ラベルが付与されるが、ラベル 2 および 3 は曖昧さを含むラベルであるため、問題を単純化するために、ラベル 1 および 2 を 1 に、ラベル 3 および 4 を 4 にそれぞれ統一することで 2 値のラベルを付与する問題とした。なお、ラベル 9 (観測ミス) については本実験の対象外とした。実験に使用したプロフィールは、北太平洋域で観測され、日本で遅延品質管理が行われたプロフィールとした。上記のうち、観測エラーを含む 8,223 プロフィールと、ランダムに選択された観測エラーを含まない 822 プロフィールとを使用した。観測エラーを含むプロフィールのうち、90%にあたる 7,401 プロフィールを学習に使用した。学習に使用しない検証用の 822 プロフィールは、観測エラーを含まない 822 プロフィールとあわせて、検証用として用いた。

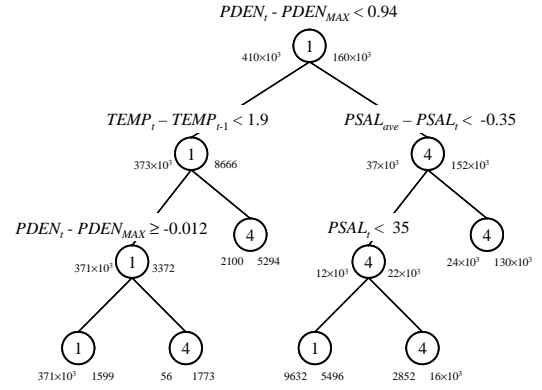


図 3: 学習により得られた決定木

### 4.2 決定木学習による CRF 用素性の自動設計

前節で示した学習用プロフィールを用いて、決定木を構築した。Classification And Regression Trees (CART) [Olshen 84] を用いて学習を行い、図 3 に示す決定木を得た。図 3 において、 $PSAL_t$ ,  $TEMP_t$  および  $PDEN_t$  は観測層  $t$  における塩分濃度、水温および密度を表す。 $PDEN_{max}$ ,  $PSAL_{ave}$  は  $(t-1)$  層までの最大密度、 $PSAL_{ave}$  はプロフィール全体における塩分濃度の平均値を表す。

得られた決定木の各葉ノードについて、葉ノードから根ノードに至るノードの集合を素性の条件式とすることで、表 2 に示す素性集合を得た。なお、正常な観測値に関する基本的な素性が必要であるため、 $f_1$  のみは決定木とは関係なく手動で設計した。

### 4.3 自動設計された素性を用いた観測エラーの推定

提案手法である自動設計された素性集合を用いた CRF の有効性を検証するため、自動品質管理 (Real-time QC)、および、先行研究において手動で設計した素性集合を用いた CRF との比較を行った。本実験で用いた CRF においては、Linear-chain モデルを利用し、学習率を 0.5 とした。重みの更新は更新値の総和が  $10^{-10}$  以下になるまで繰り返すものとした。また、CRF を用いた系列推定のかわりに、各層毎に品質管理ラベルを推定する点推定を行う方式との比較も行った。点推定を行う方式では、各素性の条件が成立する場合に、対応するラベルを割り当てることとした。自動設計された素性集合を用いた点推定は、決定木を直接適用する方式に相当する。

先行研究において手動で設計された素性を表 3 に示す。塩分の品質管理フラグを識別対象とし、密度逆転 (正側, 負側)、同値エラー、その他観測不良、オフセットエラーの 5 種類の観測エラーを検出する。WOA は入力プロフィールと比較可能な World Ocean Atlas 05 のデータ値で、標準偏差値を利用した。また、一部の素性において条件式に SVM を導入し、学習により閾値を決定するものとした。

学習に利用していない 1,644 プロフィールにおける、観測層単位での識別結果を表 4 に示す。TP は観測エラーに対して正しくラベル 4 を割り当てた層の総数、TN は観測エラーを含まない層に対して正しくラベル 1 を割り当てた層の総数、FP は観測エラーを含まない層に対して誤ってラベル 4 を割り当てた層の総数、および、FN は観測エラーを含む層に対して誤ってラベル 1 を割り当てた層の総数である。従来手法の素性集合は、密度逆転や同値エラーを対象とした素性であるため、他の種類の観測エラーを検出することができずに、TP が低下した。一方、提案手法は従来手法や Real-time QC と比較して大

表 2: 決定木から得られた CRF 用素性集合

番号	条件	素性値
$f_1$	$2 < PSAL_t \leq 41$	$1/T$
$f_2$	$PDEN_t - PDEN_{max} < 0.94 \wedge TEMP_t - TEMP_{t-1} < 1.9 \wedge PDEN_t - PDEN_{max} \geq -0.012$	1.0
$f_3$	$PDEN_t - PDEN_{max} < 0.94 \wedge TEMP_t - TEMP_{t-1} < 1.9 \wedge PDEN_t - PDEN_{max} < -0.012$	1.0
$f_4$	$PDEN_t - PDEN_{max} < 0.94 \wedge TEMP_t - TEMP_{t-1} \geq 1.9$	1.0
$f_5$	$PDEN_t - PDEN_{max} \geq 0.94 \wedge PSAL_{ave} - PSAL_t < -0.35 \wedge PSAL_t < 35$	1.0
$f_6$	$PDEN_t - PDEN_{max} \geq 0.94 \wedge PSAL_{ave} - PSAL_t < -0.35 \wedge PSAL_t \geq 35$	1.0
$f_7$	$PDEN_t - PDEN_{max} \geq 0.94 \wedge PSAL_{ave} - PSAL_t \geq -0.35$	1.0

表 3: 先行研究において使用した素性

No.	内容	素性の条件	素性値
$f_1$	正常	$2 < PSAL_t \leq 41$	$1/T$
$f_2$	不良	$PSAL_t < 2 \wedge 41 < PSAL_t$	1.0
$f_3$	同値	$PSAL_t = PSAL_{t+1} \wedge 800 < PRES_t$	1.0
$f_4$	水誤	$PRES_t \leq PRES_{t-1}$	1.0
$f_5$	負密	$SVM(PDEN_t - PDEN_{max} \wedge PRES_t) = 4$	1.0
$f_6$	正密	$SVM(PDEN_{min} - PDEN_t \wedge PRES_t) = 4$	1.0
$f_7$	オフ	7 割の層が WOA と標準偏差の 3 倍以上の差がある	1.0
$f_8$	オフ	7 割の層が過去データと 0.05 以上の差がある	1.0
$f_9$	オフ	過去アルゴデータの塩分ラベルがすべてエラー	0.1

表 4: 実験結果 (全観測エラー)

手法	TP	FN	TN	FP	F 値
Real-time QC	10,314	8,201	128,243	426	0.705
点推定+手動設計素性	1,170	17,345	128,010	659	0.115
決定木 (点推定+自動設計素性)	15,760	2,755	109,543	19,126	0.590
CRF (系列推定)+手動設計素性 [Ono 15]	1,151	17,364	128,053	616	0.113
CRF (系列推定)+自動設計素性 (提案手法)	18,235	280	127,547	1,122	0.962

幅に TP を改善できており、それらの手法で検出できなかった観測エラーを検出できたことがわかる。特に、複数の層に渡って連続して発生している観測エラーの検出が可能となり、表 4 に示す数値が大幅に改善している。

また、対象とする観測エラーを、先行研究 [Ono 15] と同様に密度逆転、同値エラー、および外れ値のみに限定した場合の実験結果を表 5 に示す。この結果より、密度逆転や同値エラーについては、従来手法よりも、提案方式における検出性能が低下していることがわかる。

最後に、CRF を用いて系列推定を行った方式と、点推定方式とを比較する。表 4 においては、提案手法と決定木とを比較すると、系列推定を行うことで、観測エラーの推定性能が大幅に向上することが確認できる。一方、対象とする観測エラーを限定した場合 (表 5)、系列推定を行うよりも点推定を行う方が精度が高い。このことから、密度逆転および同値エラーは、発生頻度が高いものの単層またはごく少数の連続する層のみにラベル 4 が割り当てられるため、プロフィール全体にラベル 4 を割り当てると比較して重要度が低いとみなされていることがわかる。

以上のように、提案方式において、発生頻度が極端に少ない観測エラーや、単層のみで例外的に生じる観測エラーについては学習が適切に行われず、検出精度が低下した。素性値の設定や学習データのアンダー/オーバーサンプリング等の工夫が必要と考える。

## 5. おわりに

海洋観測データに対して CRF を用いて品質管理フラグを割り当てる方式において、素性関数を決定木学習により自動構築する方式を提案した。本方式により、観測エラーの種類を事前に定義、分類することなく素性関数を設計できる。評価実験により、構築した素性関数を用いて CRF により系列ラベル推定を行うことで、特に複数の観測層にわたる観測エラーに対して

表 5: 実験結果 (密度逆転, 同値エラー, 外れ値のみ)

手法	TP	FN	TN	FP	F 値
Real-time QC	169	140	35,725	303	0.433
点推定+手動設計素性	273	9	36,326	49	0.904
決定木 (点推定+自動設計素性)	232	50	36,345	30	0.853
CRF (系列推定)+手動設計素性 [Ono 15]	256	21	36,343	32	0.909
CRF (系列推定)+自動設計素性 (提案手法)	228	54	36,218	157	0.684

高い精度で品質管理フラグを割り当てることを確認した。

一方で、従来から対象としていた密度逆転等の観測エラーでは、従来手法よりも精度が劣ることも確認された。また、決定木学習によって得られた素性関数は、専門技術者の直観的な理解とは差異がある点も課題である。今後は、多層に渡る観測エラーと比較的少数の層で発生する観測エラーとの双方に対応する素性の設計を試みる。

## 謝辞

本研究は独立行政法人海洋研究開発機構 地球環境変動領域アルゴデータ班の協力のもとに実施した。また、本研究の一部は、倉田記念日立科学技術財団 倉田奨励金の支援を受けた。ここに記して感謝の意を表する。

## 参考文献

- [Cortes 95] Cortes, C. and Vapnik, V.: Support-vector networks, *Machine learning*, Vol. 20, No. 3, pp. 273–297 (1995)
- [細田 13] 細田 滋毅, 全球海洋監視システム「アルゴ」, *人工知能学会第 27 回全国大会*, 3K1-OS-08a-1, 2013.
- [Lafferty 01] Lafferty, J. D., McCallum, A., and Pereira, F. C. N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pp. 282–289, San Francisco, CA, USA (2001), Morgan Kaufmann Publishers Inc.
- [Olshen 84] Olshen, L., Stone, C. J., et al.: Classification and regression trees, *Wadsworth International Group*, Vol. 93, No. 99, p. 101 (1984)
- [Ono 15] Ono, S., Matsuyama, H., Fukui, i. K., and Hosoda, S.: Error detection of oceanic observation data using sequential labeling, in *Data Science and Advanced Analytics (DSAA)*, 2015. 36678 2015. *IEEE International Conference on* (2015)
- [Roemmich 01] Roemmich, D., Boebel, O., Desaubies, Y., Freeland, H., Kim, K., King, B., Le Traon, P.-Y., Molinari, R., Owens, B. W., Riser, S., et al.: Argo: The global array of profiling floats (2001)
- [Seymore 99] Seymore, K., McCallum, A., and Rosenfeld, R.: Learning hidden Markov model structure for information extraction, in *AAAI-99 Workshop on Machine Learning for Information Extraction*, pp. 37–42 (1999)
- [Argo 02] Argo Data Management Team: Report of the Argo Data Management Meeting, in *Proc. Argo Data Management Third Meeting, Marine Environmental Data* (2002)