

ランダムフォレストと変数重要度分析を用いた 河川流域での水温変化特性の支配要因の検出

溝口 裕太¹・赤松 良久²・宮本 仁志³・中村 圭吾⁴

¹正会員 土木研究所 水環境研究グループ河川生態チーム (〒305-8516 茨城県つくば市南原 1-6)
E-mail: mizoguchi-y577bt@pwri.go.jp (Corresponding Author)

²正会員 山口大学大学院教授 創成科学研究科 (〒755-8611 山口県宇部市常盤台 2-16-1)

³正会員 芝浦工業大学教授 工学部土木工学科 (〒135-8548 東京都江東区豊洲 3-7-5)

⁴正会員 土木研究所 水環境研究グループ河川生態チーム (〒305-8516 茨城県つくば市南原 1-6)

流域規模の水温特性を支配する要因を明らかにするために、観測水温と平衡水温に基づく指標である修正熱感度 ($dM_m T_w / dM_m T_{eq}$) もしくは熱平衡偏差 ($M_a T_w - M_a T_{eq}$) を目的変数、流域と河道に関する 33 の特徴量を説明変数とする回帰木モデルを構築し、それを用いて変数重要度の分析を行った。その結果、修正熱感度については、主流長、主流勾配など河道特性と、集水面積、最低標高など集水域の地形に関する要素が重要度の高い変数として検出された。また、熱平衡偏差については、森林、草地、農地など集水域の地被に関する変数や、河川水の昇温を駆動する日射を受ける斜面方位の重要性が示された。

Key Words: water temperature characteristics, basin and channel characteristics, random forest, variable importance analysis

1. 序論

気候変動は将来の河川における水温レジームに変化をもたらす¹⁾。この気候変動下での水温予測には、気候変化に対する水温の応答性を明らかにすることが重要であり、その応答感度は河川地先の特徴に依存することが明らかになりつつある。例えば、河川流域に関係する応答要因では、集水域面積、標高、地被などが挙げられ、一方、河道関連の要因では、勾配、水系次数、基底流出、河畔林による遮光などが挙げられる^{2,3,4)}。これらから、河川の水温変化に影響を及ぼす河川地先の特徴は、それぞれの流域、河道の特性に強く拘束されるため、一つの流域で得られた知見を単純に他流域へ適用することは困難である。これまで、気温と水温の統計的関係^{2,6)}や熱輸送モデルに基づく水温予測^{3,7)}などの研究が個々に進められてきたが、気候変動下における今後の研究展開では、特徴の異なる流域での研究を丁寧に重ね、その成果の包括的な理解を介して、一般性がある知見に発展させることが重要となる。しかしながら、この観点からの研究は国際的にも事例が少なく、我が国においても水温研究が活発であった1970、80年代以降の報告は限られているのが現状である。

そこで、本研究では、Random Forest (以下、RF)⁸⁾による回帰木モデルの構築と、変数重要度分析を組み合わせ、流域規模の水温特性の説明に重要な流域、河道の特徴を検出することを目的とする。ここでは、河川水温と、間接的に水温を支配すると考えられる水文流出に関する複数の既往研究事例^{2,5,9)}を参考にして、流域、河道に関する説明変数を設定した。また、目的変数には、前報¹⁰⁾で流域規模の水温特性を捉える指標として提案した修正熱感度と熱平衡偏差 (2(2)に詳述) を採用した。このような、水温指標と流域、河道に関する説明変数との要因分析は、気候変動の影響を受けやすい流域の特定や河川地先のもつ感熱特性の明確化につながると期待される。

2. 材料と方法

(1) 現地観測

中国地方の一級水系である高津川、佐波川および小瀬川を対象流域とした。現地観測の詳細は前報¹⁰⁾に譲るが、観測点は流域全体を網羅するように設定し、高津川30、佐波川19、小瀬川17地点である(図-1)。対象期間は、2017年3月1日から2019年2月28日の2ヵ年とし、本研

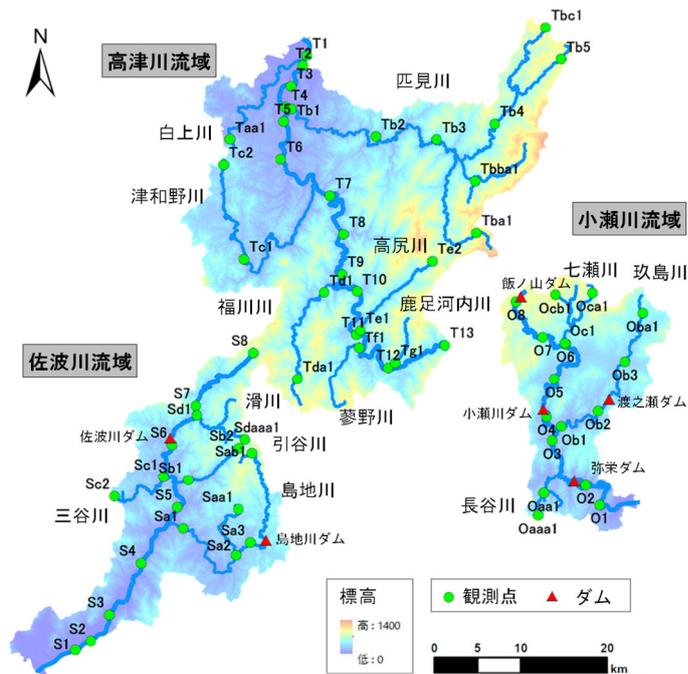


図-1 対象流域と観測点（溝口ら¹⁰⁾を参考に作成

究では、月、年平均値を議論の基本とすることから、15分間隔の観測値から日平均値を求め、それぞれを算出した。なお、観測データの欠損が最長5日程度かつ、近隣の観測点における日平均水温に、欠損期間内の急激な変化が認められない場合には、欠損期間前後のデータから求めた平均値を、補間データとして利用した。

(2) 水温特性

RFの目的変数とする2つの指標の計算方法と空間的な変化特性は前報¹⁰⁾に譲るが、熱平衡偏差 ($M_a T_w - M_a T_{eq}$) は年平均観測水温 $M_a T_w$ と年平均平衡水温 $M_a T_{eq}$ の差を、修正熱感度 ($dM_a T_w / dM_a T_{eq}$) は12対の月平均観測水温 $M_m T_w$ と月平均平衡水温 $M_m T_{eq}$ から求まる回帰直線の傾きと定義した。また、熱平衡偏差は大気の熱環境からの独立性、修正熱感度は大気の熱環境の変化に対する応答性を示す指標である。したがって、熱平衡偏差の独立性が低く(値が0に近い)、修正熱感度の応答性が高い(値が大きい)場合には、気象が水温形成の支配的な要因だと考えられ、そういった特徴を持つ観測点は下流域に広く認められる。ロワール川流域の事例では、気温と観測水温から求まる熱感度 TS (Thermal Sensitivity, dT_w / dT_a , T_w : 水温, T_a : 気温) を目的変数の一つに採用²⁾しているが、気温5-20℃の範囲外を含む場合は、非線形性が強いことが報告されている¹²⁾。他方、本研究で用いた平衡水温は、蒸発冷却などの状態変化を含む、水面を介した大気との熱交換プロセスが考慮されている。そのため、冷水魚(特に稚魚)の温度耐性など生態系保全に重要な高温域、また、気候変動の影響を受けやすい融雪期の水温変化など低温域を含む、広範な温度範囲での線形性が成り立つ

表-1 流域および河道に関する特徴量

流域-基本			
集水面積	流域集中度	形状係数	河川密度
平均傾斜	平均標高	最低標高	最高標高
起伏量	平坦地	急傾斜地	
流域-斜面方位			
北	北東	東	南東
南	南西	西	北西
流域-地被・土壌			
森林	針葉樹林	広葉樹林	草地
裸地	農地	黒ボク土	
河道		気象	
主流長	主流勾配	降水量	気温
構造物-ダム			
集水面積	距離	貯水量	

¹³⁾ため、修正熱感度は高い優位性を持つと言える。

(3) 流域および河道の特徴量

既往の河川水温と、間接的に水温を支配すると考えられる水文流出に関する研究事例²⁵⁾⁹⁾を参考に、流域、河道に関する特徴量を表-1のように設定し、回帰木モデルを構築するための説明変数とした。まず、流域特性である観測点(図-1)ごとの集水域は、DEM(10mメッシュ)¹⁴⁾を入力データの基本とした水文解析(ESRI社製ArcGIS 10.7)から求めた。また、流域集中度(集水面積と等しい円の円周/流域周囲長)、形状係数(集水面積/主流長²)、河川密度(河川延長/集水面積)は、同解析から求まる流域周囲長、主流長、河川延長を用いて計算した⁵⁹⁾。その他、平均、最低、最高標高と、起伏量(最低と最高の標高差)、平均傾斜はDEMに基づき算出し、平坦地と急傾斜地は集水域に占める傾斜がそれぞれ10°以下と、30°以上の面積割合とした⁹⁾。水温に及ぼす流域の方位の影響が報告⁹⁾されているため、DEMに基づき斜面方位を45°ごとの8つ(北、北東、東、南東、南、南西、西、北西)にわけ、それぞれが集水域に占める割合を求めた。また、地被に関しては、森林比、農地比は土地利用細分メッシュデータを、針葉樹林、広葉樹林、草地、裸地は自然環境保全基礎調査、黒ボク土は土地分類基本調査の成果を参照¹⁵⁾¹⁶⁾し、それぞれの面積割合を求めた。河道特性には、水文解析から求まる主流長、主流勾配(起伏量/流路長)を用いた⁹⁾。また、前報¹⁰⁾では、佐波川と小瀬川流域に建設されたダムが、水温特性の流程変化に及ぼす影響を認めたため、集水面積(ダム集水面積/観測点集水面積)、距離(構造物から観測点

までの距離/主流長), 貯水量(ダム貯水量/観測点集水域の年降水量)を無次元の特徴量として考慮した。気象に関する年降水量と年平均気温は, 農研機構メッシュ農業気象データ(The Agro-Meteorological Grid Square Data, NARO) (<https://amu.rd.naro.go.jp/>)を参照^{17,18)}し, 集水域ごとに平均値を計算した。なお, メッシュ農業気象データは, アメダスなどで観測される気象値が空間補間され, 約1km四方の3次メッシュデータとして提供されるものである。空間補間の具体的な方法と精度については, 既往研究¹⁷⁾を参照されたい。次節に詳述するRFの構築には, 観測水温の長期にわたる欠損がなく, 目的変数を1年ごと(2017年3月~2018年2月, 2018年3月~2019年2月)に求めることができた101のデータセットを用いた。また, 説明変数のうち気象については, 目的変数と同じ期間の年降水量と年平均気温を, その他の説明変数は, 期間に寄らず同じ値を用いた。

(4) 統計解析

a) ランダムフォレスト

目的変数(修正熱感度, 熱平衡偏差)と説明変数(流域, 河道特性)との関係を求めるために, CART(Classification and Regression Trees)に基づく, アンサンブル機械学習法の一つであるRFを用いた。RFでは, それぞれのブートストラップ・サンプルについて, 異なる無数の分類回帰木モデルが構築される。そして, 本研究のような回帰問題であれば, 構築されたすべてのモデルの予測値の平均が, 最終的な出力結果として求まるものである。また, RFの特徴には, (a)説明変数の種類が, サンプルサイズより多くても使用できること, (b)説明変数として予測に寄与しない多数の変数を含んだ場合にも, その性能は低下しにくいこと, (c)アンサンブルによる過学習の回避などがある¹⁹⁾。これらは, 流域や河道に関する多数の特徴量から, 水温特性を支配する重要な要素を探索する上での優位性である。なお, 回帰木モデルのノード数は10, 作成するモデル数は200とした。

また, 構築された回帰木モデルの精度を確認するために交差検証を行った。ここで採用した5分割交差検証では, (i)全データから重複がないようランダムにデータを抽出し, 5つのデータセットを作成, (ii)その4/5を訓練データにモデルを構築, (iii)残りの1/5を検証データとしてモデルの精度を評価するものである。また, 検証用のデータセットに重複がないよう(ii), (iii)を繰り返し, 5つのモデルを作成した。なお, RFの実装には, pythonの機械学習のモジュールであるsklearn0.22.1を使用した。

b) 変数重要度

本研究では, 環境データを対象とした機械学習での有用性が認められた²⁰⁾, wrapper法の一つであるPIMP(Permutation Importance)²¹⁾を用いて, 重要度の高い説明

変数を検出した。そもそも, PIMPは, 構築した回帰木モデルが有する本来の精度と, そのモデルに, あえてランダムに並び替えた説明変数を入力し, 計算される予測精度を比較することで, 精度を左右する相対的に影響が大きな変数を求める方法である。その具体的な手順は, 次の通りである。(ア)前項で求めた回帰木モデルを入力するランダムに並び替えた説明変数のデータセットを準備する。なお, 1種類の説明変数を対象とし, その他の説明変数と目的変数は正しい組み合わせとする。(イ)データセットである(ア)を回帰木モデルに入力して予測精度を求め, 本来の回帰木モデルの精度と比較する。(ウ)説明変数の並び替えパターンが予測精度に与える偶発性を減らすために(ア), (イ)を50回繰り返す²¹⁾。(エ)すべての種類の説明変数について, (ア)から(ウ)を繰り返し, 予測精度の減少を意味するMDA(Mean Decrease Accuracy)を求める。一般に, 回帰木モデルの予測精度が低下する(MDAが大きい)ほど, モデルの性能を左右する説明変数であり, 重要度が高い変数だと捉えることができる。なお, PIMPの実装には, RFと同様にsklearn0.22.1を使用した。

3. 結果と考察

(1) 流域および河道の特徴量

RFを含む機械学習は, 内挿問題への適用を基本とする。したがって, 訓練データである図-2に示す, 流域と河道に関する特徴量の範囲は, 本研究で構築するRFの適用限界の目安となる。まず, 流域特性である集水面積は400km²を下回る観測点が多数を占め, それを上回るのは高津川流域の下流に位置するいくつかの観測点のみである。また, 形状係数のエラーバーの範囲(以下, 範囲)は0.07-0.43にある。平坦地と急傾斜地の範囲は, 1-21%と8-69%である。平均, 最低, 最高標高および起伏量は, それぞれ246-877m, 2-753m, 771-1,344mおよび116-1,337mとなった。斜面方位は, 北, 北東の割合がやや小さいものの8-17%であり, 概ね似通った範囲となった。地被は, どの集水域も概ね80%以上を森林に覆われているが, 針葉樹, 広葉樹の内訳は集水域ごとに大きく異なる。草地, 裸地および農地は, 0-18%の範囲にある。佐波川流域に2つ, 小瀬川流域に4つのダムが建設されているが, その下流に位置する観測点(データセットは30)を除き, ダムに関する3つの特徴量は0となる。河道特性である主流長の範囲は1-74km, 主流勾配は0.01-0.21, また, 気象要因の年降水量は1,367-2,602mm, 年平均気温は8.0-14.5°Cである。

(2) 変数重要度

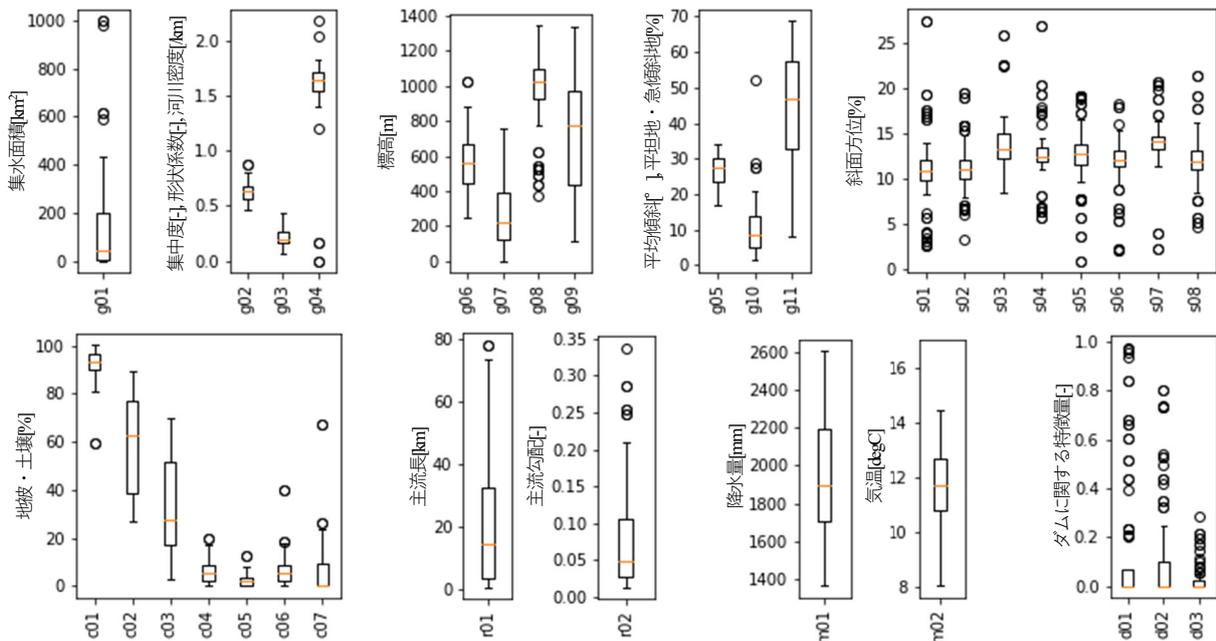


図-2 流域および河道に関する特徴量（エラーバーの上下限值：第1, 3四分位数±1.5×四分位範囲, ○：外れ値, g01：集水面積, g02：流域集中度, g03：形状係数, g04：河川密度, g05：平均傾斜, g06, g07, g08：平均, 最低, 最高標高, g09：起伏量, g10：平坦地, g11：急傾斜地, s01：北, s02：北東, s03：東, s04：南東, s05：南, s06：南西, s07：西, s08：北西, c01：森林, c02：針葉樹林, c03：広葉樹林, c04：草地, c05：裸地, c06：農地, c07：黒ボク土, r01：主流長, r02：主流勾配, m01：降水量, m02：気温, d01：集水面積, d02：距離, d03：貯水量を意味し, 図-3, 4も同様とする。）

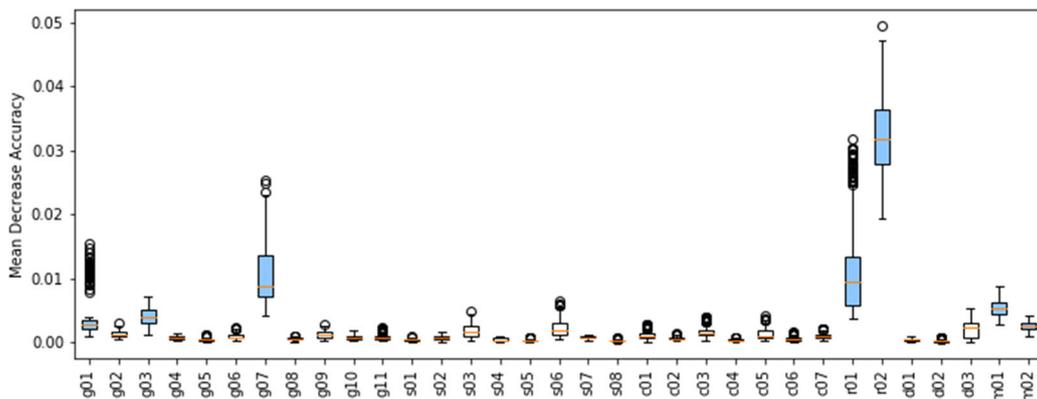


図-3 修正熱感度 (dM_w/dM_{eq}) に関する変数重要度 (MDAの中央値が全説明変数の中央値平均を上回るものを青く着色)

はじめに、検証データに基づく回帰モデルの再現性を論じる。交差検証のために構築した5つのモデルから求まる予測値に基づく、修正熱感度と熱平衡偏差の R^2 は共に0.63, RMSEは0.04, 0.74となった。また、それぞれのモデルは、精度にばらつきを持っており、例えば、熱平衡偏差の R^2 の範囲は0.35-0.74となった。これは、回帰モデルを構築する際のランダム性に依存し、訓練、検証データに含まれる水温特性と特徴量に偏りがあるためだと考えられる。データセットを時空間的に充実させることで、引き続き、モデルの予測精度を高める必要がある。具体的な結果は後述するが、本研究で検出された重要な変数と、既往研究で言及されたものとの整合を確認しており、構築した回帰モデルは、変数重要度分析を実行する上での精度を有すると考える。次に、変数重要度分析の結果を図-3, 4に示す。その特徴を概

観すると、修正熱感度は、河道特性や集水域の地形に関する変数が検出される一方、熱平衡偏差は集水域の地被に関するものが多く検出された。次に、それぞれについて詳述する。

a) 修正熱感度 (dM_w/dM_{eq})

MDAの中央値が平均を上回った重要度の高い変数は、主流長, 主流勾配, 集水面積, 最低標高, 形状係数, 降水量, 気温となった(図-3)。類似する熱感度TSを用いたロワール川流域における水温研究²⁾では、重要度の高い代表的な変数として、主流長, SF (shading factor), BFI (base-flow index), 観測点の標高, 8月の流量, 観測点付近の河川の勾配を挙げている。変数の種類, 種数や, 分析方法が異なり、単純な比較はできないが、主流長, 主流勾配(河川の勾配), 最低標高(観測点の標高)は、本研究と同様に重要性が示されている。それぞれの変数

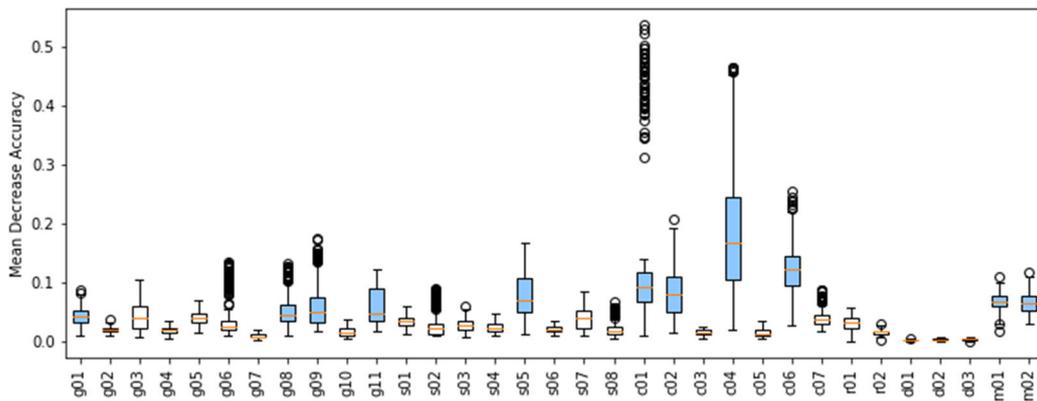


図-4 熱平衡偏差 ($M_bT_w - M_bT_{eq}$) に関する変数重要度 (MDAの中央値が全説明変数の中央値平均を上回るものを青く着色)

の特徴は、次の通りである。主流長（源流から観測点の距離）は、これが大きいほど、水塊と大気との熱交換の時間が長い²⁵⁾ことを意味する。また、主流勾配は、流速の大小²⁵⁾を意味し、緩やかな勾配であれば、同一距離を流下しても熱交換の時間は相対的に長い。最低標高は、高地ほど雪解け水の影響を強く受ける²⁾と考えられる。

この他、降水量は流量を代替する指標、形状係数は主流長あたりの流量の大きさを示す指標⁹⁾である。流量（降水量）の増加にともない、水塊は大気による加熱・冷却の影響を受けにくくなる一方で、集水域の形状が細長い（形状係数が小さい）ほど、主流長あたりの集水面積が小さいために流量は少なく⁹⁾、水塊は大気による影響を受けやすいと考えられる。また、前報¹⁰⁾では、ダム堤体下流に位置する観測点において、その上下流と比べて、修正熱感度が著しく小さくなる事例を報告している。既往研究でも、ダム等の貯水池が TS に与える影響を指摘²²⁾しているが、本研究で導入した3つの特徴量では、修正熱感度に対するダムの影響を明瞭に捉えることはできなかった。これは、RF の構築において、ダムの影響を読み取ることのできるデータセットが数例に留まったことに依存する結果だと推察される。

b) 熱平衡偏差 ($M_bT_w - M_bT_{eq}$)

熱平衡偏差については、森林、針葉樹林、草地、農地、斜面方位（南）、集水面積、最高標高、起伏量、急傾斜地、降水量、気温が、MDA の中央値が平均を上回る重要度の高い変数となった（図-4）。このうちの多くの変数は、観測水温の年平均に影響を及ぼすと考えられる集水域の地被に関連する特徴量である。Home・Hubbart²³⁾は、低茎植生、牧草地および耕作地の面積割合の増加が、河川水温を上昇させる一方、森林が水温低下を招く要因であることを明らかにしており、本研究では、それらに類する草地、農地および、森林、針葉樹林が重要度の高い変数として選ばれた。なお、対象流域の農地は、その大半が水田であるため、その知見²³⁾の単純な適用はできないが、夏季の水田での昇温効果の高さが報告⁹⁾されていることから、河川水温の上昇に、水田排水が

影響を与えていることは十分に考えられる。この他、複数の先行研究が昇温メカニズムを駆動する日射の重要性を指摘^{23,24,25)}しているが、いずれの方位の中で、年間日射量が最も大きい南向き斜面が重要な変数として検出された。平衡水温を構成する一つの要素である気温は、標高に依存することが知られている。したがって、起伏量は、集水域における平衡水温の変化量を、急傾斜地は変化速度を代替する指標だと捉えることができ、熱平衡偏差は気象の空間変化の影響を受けやすいと考えられる。

4. 結論

本研究では、前報¹⁰⁾で提案した河川水温の管理指標である修正熱感度と熱平衡偏差に関して、それらの空間変化に支配的な流域、河道特性を明らかにすることを試みた。また、一般性の高い知見を得るために、既往の河川水温と、間接的に水温を支配すると考えられる水文流出に関する複数の既往研究事例から、流域、河道に関する33の特徴量を整理した上で、それらを説明変数に回帰モデルを構築し、変数重要度分析を実行した。機械学習に必要なデータセットの時空間的な充実による、さらなるモデルの精度向上が求められるものの、流域規模の水温特性を支配する流域、河道に関する特徴量の絞り込みと、その妥当性についてメカニズムの観点から精査した。以下に、本研究で得られた主な知見をまとめる。

- 1) 修正熱感度は、それと類似する水温特性（TS）を用いた先行研究で重要性が指摘されている、主流長、主流勾配、最低標高を含む、集水面積、形状係数、降水量、気温が、重要度の高い変数として検出された。
- 2) 熱平衡偏差を説明する変数として、森林、針葉樹林、草地、農地、斜面方位（南）、集水面積、最高標高、起伏量、急傾斜地、降水量、気温が検出された。これらは、観測水温と平衡水温の年平均値への関与を通じて、熱平衡偏差に影響するものと推察された。

謝辞：本研究は、土木研究所、山口大学、芝浦工業大学の3者の協働により実施された。また、本研究の一部は、(独)環境再生保全機構の環境研究総合推進費(JPMEERF20202004)により実施した。ここに記して謝意を表す。

参考文献

- 1) Huguet, F., Parey, S., Dacunha-Castelle, D. & Malek, F.: Is there a trend in extremely high river temperature for the next decades? A case study for France, *Nat. Hazards Earth Syst. Sci.*, 8, pp.67–79, 2008.
- 2) Beaufort, A., Moatar, F., Sauquet, E., Loicq, P. & Hannah, D.M.: Influence of landscape and hydrological factors on stream-air temperature relationships at regional scale, *Hydrol. Process.*, 34, pp.583–597, 2020.
- 3) 浦野仁志, 宮本仁志, 前羽洋, 戸田祐嗣: 河川流域の流水水温形成における基底流出水温の影響について, *土木学会論文集B1*, 69, pp.I_1681-I_1686, 2013.
- 4) Hannah, D. M. & Gamer, G.: River water temperature in the United Kingdom: Changes over the 20th century and possible changes over the 21st century, *Prog. Phys. Geogr.*, 39, pp.68–92, 2015.
- 5) 牧野富彦, 宮沢哲男: 豊川水系黄柳川の河川水温について, *水温の研究*, 26, pp.13–20, 1983.
- 6) Mantua, N., Tohver, I. & Hamlet, A.: Climate change impacts on streamflow extremes and summertime stream temperature and their possible consequences for freshwater salmon habitat in Washington State, *Clim. Change*, 102, pp.187–223, 2010.
- 7) MacDonald, R. J., Boon, S., Byrne, J. M., Robinson, M. D. & Rasmussen, J. B.: Potential future climate effects on mountain hydrology, stream temperature, and native salmonid life history, *Can. J. Fish. Aquat. Sci.*, 71, pp.189–202, 2014.
- 8) Breiman, L.: Random Forests, *Mach. Learn.*, 45, pp.5–32, 2001.
- 9) 末永博, 溝口裕太, 田中規夫, 井原和彦, 又吉健太: 河川合流点での水位・流量ピーク時刻の集中を支配する降雨および流域特性の分析と集水域間の比較, *河川技術論文集*, 25, pp.121–126, 2019.
- 10) 溝口裕太, 赤松良久, 宮本仁志, 中村圭吾: 平衡水温を用いた流域スケールの河川水温管理指標の提案, *河川技術論文集*, 27, pp.289–294, 2021.
- 11) 河野蒼仁, 赤松良久, 乾隆帝: 中国地方一級水系における河川水温の時空間構造変化, *土木学会論文集B1*, 74, pp.I_463-I_468, 2018.
- 12) Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., & Marshall, L.: Investigating controls on the thermal sensitivity of Pennsylvania streams, *Hydrol. Process.*, 26, pp.771–785, 2012.
- 13) Bogan, T., Mohseni, O. & Stefan, H. G.: Stream temperature-equilibrium temperature relationship, *Water Resour. Res.*, 39, 1245, 2003.
- 14) 国土地理院: 基盤地図情報ダウンロードサービス, <https://fgd.gsi.go.jp/download/menu.php>, 2020.
- 15) 国土交通省: GISホームページ, <https://nlftp.mlit.go.jp/index.html>, 2020.
- 16) 環境省 自然環境局 生物多様性センター: 第6-7回自然環境保全基礎調査植生調査報告書, <http://www.biodic.go.jp/>, 2020.
- 17) 大野宏之, 佐々木華織, 大原源二, 中園江: 実況値と数値予報, 平年値を組み合わせたメッシュ気温・降水量データの作成, *生物と気象*, 16, pp.71–79, 2016.
- 18) 小南靖弘, 佐々木華織, 大野宏之: メッシュ農業気象データ利用マニュアル, 農研機構, Ver.4, 67p, 2019.
- 19) Díaz-Uriarte, R. & Andrés, S. A. De.: Gene selection and classification of microarray data using random forest, *BMC Bioinformatics*, 7, 3, 2006.
- 20) Effrosynidis, D. & Arampatzis, A.: An evaluation of feature selection methods for environmental data, *Ecol. Inform.*, 61, 101224, 2021.
- 21) Altmann, A., Toloşi, L., Sander, O. & Lengauer, T.: Permutation importance: A corrected feature importance measure, *Bioinformatics*, 26, pp.1340–1347, 2010.
- 22) Morill, J. C., Bales, R. C. & Conklin, M. H.: Estimating stream temperature from air temperature: Implications for future water quality, *J. Environ. Eng.*, 131, pp.139–146, 2005.
- 23) Home, J. P. & Hubbart, J. A.: A spatially distributed investigation of stream water temperature in a contemporary mixed-land-use watershed, *Water*, 12, 1756, 2020.
- 24) Campbell, G. S. & Norman, J. M.: *Introduction to Environmental Biophysics*, 2nd ed., Springer, 286p, 1998.
- 25) Younus, M., Hondzo, M. & Engel, B. A.: Stream temperature dynamics in upland agricultural watersheds, *J. Environ. Eng.*, 126, pp.518–526, 2000.

(Received June 30, 2021)

(Accepted September 3, 2021)

DETECTING DOMINANT FACTORS OF CHANGE CHARACTERISTICS OF WATER TEMPERATURE IN RIVER BASINS USING RANDOM FOREST AND VARIABLE IMPORTANCE ANALYSIS

Yuta MIZOGUCHI, Yoshihisa AKAMATSU, Hitoshi MIYAMOTO
and Keigo NAKAMURA

To clarify the factors that govern the water temperature characteristics in a basin, in this study, the observed water temperature and the equilibrium water temperature were used to build regression tree models with the modified thermal sensitivity ($dM_m T_w / dM_m T_{eq}$) or the thermal equilibrium difference ($M_a T_w - M_a T_{eq}$) as the objective variable. Additionally, 33 characteristics related to the basin and channel were used as the explanatory variables. The variable importance analysis was performed using the developed models. The results revealed that the channel characteristics such as the channel length and slope, and basin characteristics such as the catchment area and elevation are the most important variables for the gradient. Additionally, with regard to the deviation, the variables related to the land use or cover in the catchment areas, such as forest, grassland, and agricultural land, and the aspect-slope that receives solar radiation are the main factors that drive an increase in water temperature.