

技術論文

重要水処理微生物を同定するための新規統計手法の開発

Development of Statistical Method for Identification of Microorganisms Responsible for Wastewater Treatment

福島 寿和* 中川 淳一 川野 秀一 押木 守
Toshikazu FUKUSHIMA Junichi NAKAGAWA Shuichi KAWANO Mamoru OSHIKI

抄 録

生物学的排水処理プロセスである活性汚泥法は100年超の歴史があり、一部の鉄鋼排水処理にも適用されている。しかしながら処理を担う微生物はブラックボックスとして取り扱われており、処理不安定性の一因となっている。遺伝子解析により、数千～万種の微生物が混在する複雑なプロセスであることがわかってきたが、ビッグデータを紐解き、水処理との因果関係を明らかにすることは依然困難である。そこで、我々は微生物学-数学の学融合により、新規統計手法を開発し、重要水処理微生物を同定することに成功した。今後、今回発明した手法を用いて解析することで、生物学的排水処理プロセスの安定化・効率化に貢献することが期待される。

Abstract

Activated sludge process, which is a biological wastewater treatment process, has a history of over 100 years and has been also applied to a part of wastewater from steelmaking industry. However, the microorganisms responsible for the wastewater treatment have been regarded as a black box, which is one of the reasons for the instability of the process. Next generation sequencing technology has revealed that the microbial community in the process is very complex. However, the relationship between wastewater treatment and microorganisms is still unknown due to the complex community, so called “big-data”. Therefore, we developed novel statistical methods by transdisciplinary approaches, namely integrating microbiology and mathematics, and successfully identified important microorganism responsible for wastewater treatment. This method is expected to contribute to the further stabilization and efficiency of biological wastewater treatment processes.

1. 緒 言

生物学的排水処理法の代表である活性汚泥法 (activated sludge process) は、100年以上前の1914年に実用化され、世界中の都市下水、工場排水などの排水処理に幅広く適用されている¹⁾。活性汚泥法の基本となる標準活性汚泥法の処理フローを図1に示す。あらかじめ微生物を導入した曝気槽と呼ばれる好氣的(酸素が溶存している)反応槽に排水を導入すると、微生物によって有機物を主とした汚濁成分が分解、除去される。後段の沈殿槽で固体である微生物と液体である処理水を固液分離することで清浄な処理水が得られる。この処理を担う微生物を一般に活性汚泥と呼ぶ。鉄鋼業においても、コークス製造工程でコークス炉より生成される通称“安水”と呼ばれる排水の処理に活性汚泥法

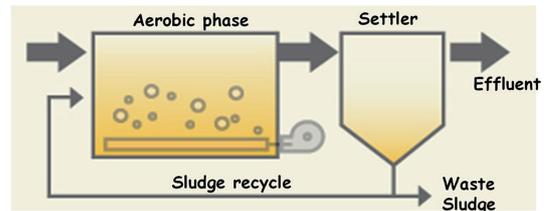


図1 標準活性汚泥法の処理フロー
Schematic of conventional activated sludge process

が適応されている。その歴史は古く、40年以上前の文献²⁾でコークス炉排水を対象とした活性汚泥法が紹介されている。

上述の通り、活性汚泥法には長年の歴史があり、広く実用化されている我々にとって身近な生物プロセスの一つであるが、処理を担う微生物の生理、生態は一部しか解明されていない。図2に示した概念図の通り、活性汚泥法の

* 先端技術研究所 環境基盤研究部 主幹研究員 博士(環境学) 千葉県富津市新富 20-1 〒293-8511

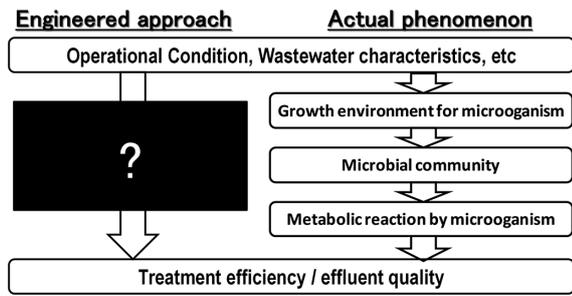


図2 生物学的処理プロセスの工学的制御方法および実際に起こる現象

Conceptual flow diagrams of biological wastewater treatment

プロセス構成や排水性状が変わると処理水質が変わるが、そのメカニズムをブラックボックスとして扱うため、処理効率の更なる向上や改善策は経験に頼らざるを得ない。実際には、図2右側の通り、条件が変わると適応できる微生物の種類が変わり、処理効率や性能が変わる。このように、排水処理を担う微生物生態を理解することで、より高度な排水処理プロセスの操業指針を検討することが理論的には可能となる。

次世代シーケンス解析技術の進歩により、遺伝子情報を取得することで生物学的排水処理プロセス中の微生物群集を網羅的に解析できるようになった。近年では解析コストも安くなり、誰もが気軽に解析できる技術になりつつある。一方で、当該技術によって数千～万種の微生物が活性汚泥として混在していることがわかってきた。そのような膨大且つ複雑な微生物情報から、水処理との因果関係を解明することは依然として困難である。例えば、排水処理プロセス中に存在している微生物が水処理に大きく寄与していると考えられるが、何%以上存在すれば優占種と言えるのか？どの微生物がどの汚濁成分の処理に寄与しているのか？これらの疑問に対しては、研究者の経験に基づく主観的判断に頼らざるを得ない。

このような現状を鑑み、日本製鉄(株)では生物学-数理学の学融合により、次世代シーケンスから得られる遺伝子ビッグデータから重要水処理微生物を客観的に同定することを目標に検討を行い、新規統計手法³⁾の開発に至った。その詳細について以下に報告する。

2. 本 論

新規統計手法³⁾を開発するため、まず検討用の実験データを構築した。水温などの環境条件を制御するため、実験室規模の水処理プロセスを運転し、水質および微生物データを構築した。この際、人為的に運転条件を変えることで、あえて処理水質および微生物群集を変動させた。得られたデータを用いて統計解析手法の検討を行い、統計解析結果より予測される水質データと実験データを比較することで、手法の妥当性を検討した。

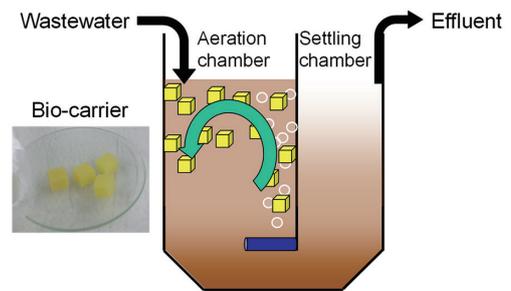


図3 実験室規模流動床担体プロセス
Schematic diagram of laboratory-scale moving bed biofilm reactor (MBBR)

2.1 実験データの構築

2.1.1 排水処理データ

実験室規模流動床担体プロセス (Laboratory-scale Moving Bed Biofilm Reactor; 以下 MBBR) を運転し、排水処理データを構築した。活性汚泥法では沈殿槽で沈降した活性汚泥を返送する必要があるが、当該プロセスでは微生物を担体に附着させることで曝気槽に保持できる。本報では、運転を簡便にするため、活性汚泥法ではなく、MBBRを採用した。図3の通り容量3.4Lの曝気槽と2.1Lの沈殿槽から構成されている。運転開始時に、曝気槽の体積比約33%の1cm³のスポンジ担体と、微生物の接種源としてコークス炉排水の排水処理プロセスより採取した活性汚泥を適量投入した。排水はコークス炉排水を模擬し、天然海水:工業用水=60:40となるよう混合した水で処理対象となるフェノール、チオシアンおよびチオ硫酸を任意の濃度となるよう調製した。そのほか、アンモニア、炭酸塩およびりんを添加した。エアポンプにより曝気槽に流量4L/分で空気曝気をすることで酸素を供給し、またスポンジ担体を巡回させた。水温は温度コントローラーと水中ヒーターにより30℃に制御した。pHは常時測定し、7.5以上になるよう水酸化ナトリウム溶液を必要に応じて添加した。

運転期間中、排水の流入速度を調整することで、水理的滞留時間 (Hydraulic Retention Time; HRT) を変動させ、処理負荷を変動させた。処理水質を定期的に分析し、各汚濁成分の1日当たりの処理速度 (mg/L/day) を算出した。図4に得られた処理水質データより算出した各処理速度の一例を示す。このMBBRではチオシアンを汚濁成分としており、HRTの短縮、すなわち処理負荷を高くすることによって、1日当たりの処理速度が上昇した。模擬排水中に含まれるアンモニアの一部はアンモニア酸化細菌によって亜硝酸に酸化される。このため、アンモニアの微生物反応については亜硝酸生成速度として算出した。

2.1.2 微生物データ

MBBRより定期的に採取したスポンジ担体試料を用いて、微生物群集解析を日鉄環境(株)にて実施した。実験概要を以下に示す。採取したスポンジ担体を4分割した後、

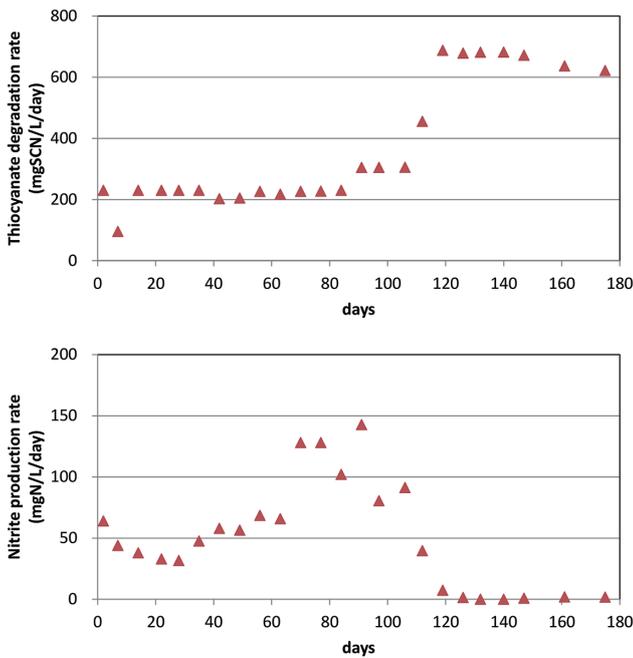


図4 汚濁成分の処理速度(亜硝酸は生成速度)の一例
Examples of estimated degradation/production rate in MBBR

Extrap Soil DNA Plus ver.2を用いて担体に付着した微生物からDNAを抽出し、精製した。得られたDNAを用いて、細菌および古細菌の16S rRNA遺伝子V4-5領域をPCR(Polymerase Chain Reaction)により増幅させた。続いてPCR産物をMiSeq System(イルミナ社)を用いて塩基配列を解読し、QIIME(Quantitative Insights Into Microbial Ecology)⁴⁾により97%以上の類似性をもつ配列を同一OTU(Operational Taxonomic Unit)として、同一微生物種と便宜上みなした。得られたOTUを基に微生物群集を決定した。

微生物群集解析結果の一例を図5に示す。図は各微生物種(OTU)の相対割合を示している。目視で認識できる微生物種はせいぜい数十種程度であるが、実際には数千種が存在していた。このように、水処理プロセスに存在する数千の微生物種について検討や考察することは事実上不可能である。一方で、どの程度の存在割合であれば重要な微生物とみなすかは、水処理プロセスや処理対象物質により異なり、研究者の主観に頼らざるを得ない。微生物群集解析はブラックボックスとして捉えてきた微生物を理解する最初のステップとなり、図5の通り、存在割合が変動していることは、水処理との因果関係が予想される。誰もが客観的に重要水処理微生物を同定できる手法があれば、このジレンマを少しでも打破できると考え、新規統計手法を開発する動機となった。

2.2 統計解析

上述の通り、排水組成や運転条件を制御した実験室規模MBBRでも数千の微生物種が混在しており、その中から水処理で特に重要な微生物を選定することは難しいことがわ

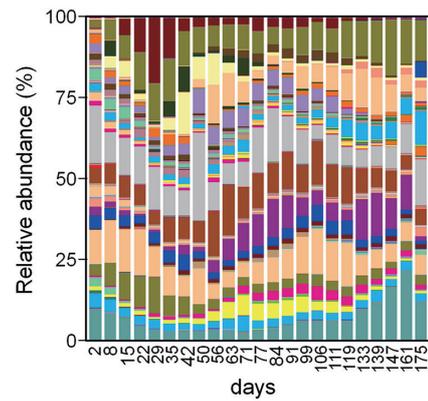


図5 微生物群集結果の一例
Example of microbial community in MBBR

かった。

水処理と微生物との関係を解析する方法としてまず回帰分析が考えられる。水処理性能、例えば汚濁成分の処理(分解)速度を目的変数 y として、処理を担う微生物種を説明変数 x としてその関係性を分析する。この際、通常回帰分析では、説明変数である微生物種すべてに係数が付与されてしまうことが問題となる。例えば、チオシアン処理を考えた場合、実験室規模MBBR内に存在するすべての微生物がチオシアン分解能を有することはなく、別の処理反応に関与する微生物も多数存在する。しかしながら、回帰分析では分解に関与しない微生物にも何らかの係数が付与されてしまうため、分解微生物は過小評価、分解に寄与しない微生物は過大評価となってしまう。加えて、水処理データ数(=採取した試料数)に対して、微生物種数が100倍以上多くなるため、正確な係数を付与することはさらに困難となる。すなわち、汚濁物質が複数種あり、膨大な数の微生物種が混在する水処理プロセスに対して、回帰分析を適用することは困難である。このため、著者らは重要水処理微生物を同定するための新規統計手法を開発した。開発したLasso+Bootstrap法およびその改良手法について、以下より説明する。

2.2.1 Lasso + Bootstrap 法

スパース推定⁵⁾の一つであるLasso(Least absolute selection and shrinkage operator)推定に基づき、各汚濁成分の分解の大部分を担う“重要水処理微生物”の同定を試みた。スパース(sparse)とは“まばらな”や“疎”を意味する英単語で、本報における微生物データのように、存在している種数は膨大であるが、実際に処理に大きく貢献している微生物種数は少ないと考えると、このアプローチが当てはまる。スパース推定は回帰係数の推定と変数選択を同時に行う方法で、ここでは、重要水処理微生物を選択する。Lasso推定に加えて、微生物データの確率的変動(ばらつき)を考慮に入れるためにBootstrap法⁶⁾を組み込んだ安定性解析を実施した。なお、Bootstrap法は微生物種の

類縁関係を示す系統樹の安定性解析にもよく用いられるため、微生物学の専門家にも馴染み深い手法である。

解析の流れは以下の通りである。

1. 得られた実験データに対して、Bootstrap 標本を発生させる（標本数は任意だが、ここでは 1000 組）。
2. 各標本について Lasso 推定を適用し、選択された回数が 1 以上の微生物種の出現回数を数える。
3. 任意の出現回数（ここでは 600 回）を超える微生物種を抽出する。
4. 抽出された微生物種を用いて線形回帰モデルを構築し、交差検証法による R^2 値を計算する。
5. R^2 値および実測した水質データと線形回帰モデルの比較に基づき、結果を考察する。

結果の一例を図 6 に示す。遺伝子解析により 3752 種の微生物が検出されたが、Lasso+Bootstrap 法により各汚濁成分の分解に大きく関与しているとされた微生物は 3~6 種となった。この同定された重要水処理微生物種を用いた線形回帰モデルによる交差検証法の結果、図 6 に示す通り、実際の水処理速度と類似した値、変動を示した。異なる条件で運転した 2 つの実験室規模 MBBR を対象に、延べ 6 種の汚濁成分除去について Lasso+Bootstrap 法を適用した結果、 R^2 値は 0.61~0.77 と大きく、優れたモデルを構築できていると考えられた。また、抽出された重要水処理微生物種と系統学的に近縁な種の生理学的特徴を文献調査したところ、多くの微生物は情報が得られなかったが、亜硝酸生成に対して重要水処理微生物と同定された種の一部はアンモニアを亜硝酸に酸化するアンモニア酸化細菌であることがわかり、微生物学的視点からも妥当な解析結果と考えられた。ただし、ここで注意が必要なのは、同定された微

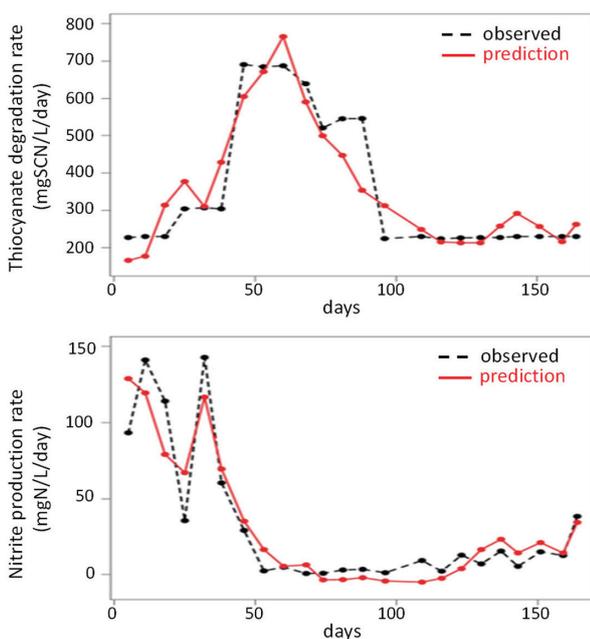


図 6 Lasso + Bootstrap 法による解析結果
Analysis results by Lasso+Bootstrap method

生物種は汚濁成分の除去速度の変化と関係がある微生物であり、分解していることを直接的に示すものではない。汚濁成分の分解以外、例えば分解微生物と共生関係にあるなどの理由で重要水処理微生物として同定された可能性もある。このような課題はあるものの、数種の同定された微生物種であればその後の生物学的検討は現実的なものとなり、新たな発見が得られると期待できる。

2.2.2 SPCR 法

前述の通り、Lasso+Bootstrap 法によって重要水処理微生物の同定に成功し、実測データとの比較により精度の高い解析技術であることを示すことができた。一方で、同定された微生物の多くは実際に分解能を有するのか、どのような環境（pH や水温など）が適しているか、など生理学的特徴を文献調査により把握することができなかった。実験的に証明することは、微生物を分離培養するなど難易度が高く、すべての微生物種の特徴を明らかにすることは、やはり現実的ではない。加えて、微生物種同士の競合や共生といった関係も考えられるため、事態はより複雑である。

そこで、微生物と環境因子や微生物間の関係性を捉え、その関係性を用いた回帰モデルを構築するために、スパース主成分回帰モデル（Sparse Principal Component Regression ; SPCR）⁷⁾の適用可能性を検討した。主成分回帰とは 1 段階目に主成分分析、2 段階目に回帰分析を行うが、2 段階目で説明変数となる主成分スコアの設計が、目的変数に全く含まれていないことが課題となる。一方の SPCR 法は 1 段階法による主成分回帰モデリング手法と捉えることができ、我々は当該手法を用いて環境-微生物や微生物間の関係性を捉え、その関係性を用いた回帰モデルの構築を検討した。

ここでは、1 段階目の微生物種の絞り込みとして、前述の Lasso+Bootstrap 法により 1000 の Bootstrap 標本のうち、1 回でも抽出された微生物種をすべて選抜した。対象とするデータによって異なるが、この操作により数千種の微生物データから数十種が選抜された。続いて、選抜された微生物種に対して SPCR を適用した。この際主成分の数は 1~5 として、同定された微生物種に対して実施した交差検証法により最も R^2 値の高い主成分の数を決定した。

結果の一例を図 7 に示す。図の例では Lasso+Bootstrap 法での R^2 値が 0.61 であったのに対して SPCR 法では 0.67 とより高くなった。このように、SPCR 法を適用することで、すべてのデータではないが、多くについては精度が高まることがわかった。また、主成分数はデータによって異なった。異なる主成分軸で抽出された微生物は異なる環境を好む可能性、同じ軸では何らかの共生関係の可能性、など環境因子と微生物、または微生物間の関係性が明らかになる可能性がある。現時点では得られた結果に対し、考察することが困難であり、これ以上の議論はできないが、今後

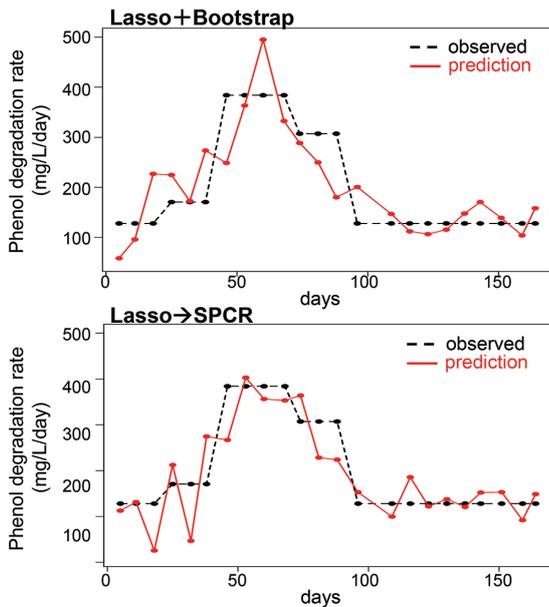


図7 Lasso + Bootstrap 法および SPCR 法による解析結果の比較

Comparison of analysis results by Lasso+Bootstrap method and SPCR method

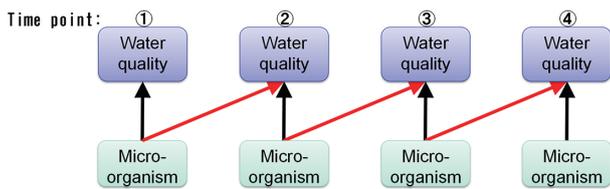


図8 時系列情報を含めた解析の概念図
Conceptual diagram of time series analysis

SPCR による解析結果をより深く考察，検討することで，水処理微生物の新たな生理学的知見が得られる可能性がある。

2.2.3 時系列情報を含めた解析

上述の Lasso+Bootstrap 法および SPCR 法を用いた解析には時系列要素はなく，同一時点の水質と微生物データを比較している。しかしながら実際の水処理プロセスでは，水処理微生物が増減したのち，その表現型として水処理の悪化や改善が現れるため，これらの変化には時間差がある可能性がある。また，その逆に pH などの環境因子の変動と水処理微生物の増減にも時間差があるケースも報告されている^{8,9)}。このような可能性を踏まえ，時系列情報を含めた解析を検討した。

具体的には図8に示すように，ある時点の水質は同一時点および1時点前の微生物が関与する可能性があるとして，両時点の微生物データと水質データを開発した統計手法を用いて検討した。この結果，一例として時系列がない解析での R^2 値が 0.77 であったのに対して，時系列情報を含めることで，0.90 に向上するなど，多くのデータについて解析精度が高まることがわかった。また，微生物種によって

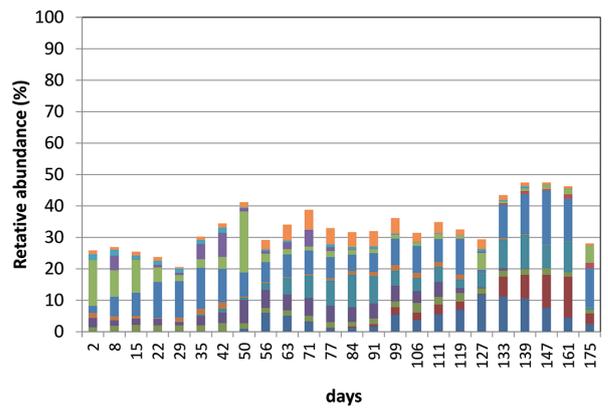


図9 Lasso + Bootstrap 法により同定された重要水処理微生物の存在割合

Relative abundances of microorganisms identified by Lasso+Bootstrap method

は抽出されたデータが同一時点もしくは1時点前と異なっていた。これは増殖速度の違いなどを表していると考えられる。今回のデータは1週間に1回採取したものであるが，より高頻度，例えば1日1回データを取ることができれば，より詳細な考察が可能になるかもしれない。

2.3 考察

今回開発した Lasso+Bootstrap 法をはじめとする統計解析手法により数千を超える微生物種から特に水処理との関係性が示唆される重要水処理微生物を数種程度に絞り込むことができた。数種であれば，それら微生物のより詳細な解析や遺伝子定量による微生物量のモニタリング¹⁰⁾も現実的に可能となる。

ここで視点を変えて，同定された重要水処理微生物の全微生物種に対する存在割合について考察する。図9に示した通り，同定された微生物種は全体の2~4割程度にしか満たなかった。半分以上の微生物は存在しているが，汚濁物質の処理には関わっていないことになる。今回の検討対象は人工的に調製した排水を処理する実験室規模 MBBR であるため，未知の汚濁成分が入っていることは考えられない。このため，死んだ微生物を餌として分解している，水処理以外の役割を担っている微生物が相当数存在しているなど，興味深い可能性を示唆している。

また，今回の微生物データは DNA を対象とした解析結果であるが，抽出した DNA には活性が低いもしくは死んだ微生物由来の DNA も含まれる。このため，死んだ微生物由来の DNA をあらかじめ排除する¹¹⁾，活性と関係性がある RNA を対象にする¹²⁾など，微生物データを得るための実験方法を工夫することで，より精度の高い結果が得られる可能性も考えられた。

本報では Lasso+Bootstrap 法に加え，開発した SPCR 法や時系列情報を含めた解析法を紹介した。現時点でこれら手法の優劣は付けられず，目的に応じて手法を選定，もし

くはすべての手法で解析し、得られた結果を総合的に考察することが望ましい。

このように、次世代シーケンスから得られる遺伝子ビッグデータを用いて、一歩進んだ解析が可能になった。ただし、本報で議論した通り、冒頭で指摘した“ブラックボックス”を完全に解明するためにはまだまだ多くの研究が必要である。今後、本報のようなアプローチにより、微生物データからより多くの知見を得て、それらを蓄積していくことが重要である。

3. 結 言

膨大な種類の微生物が混在する排水処理プロセスの測定データに対して、今回開発した Lasso+Bootstrap 法などの統計手法を適用することで、客観的に重要水処理微生物を選定することができた。今後、統計手法を用いたビッグデータ解析により、微生物群集データが生物学的排水処理プロセスの安定化・効率化に貢献することが期待される。

参考文献

- 1) Seviour, R. et al.: Microbial Ecology of Activated Sludge. 1st ed. London, IWA Publishing, 2010, 688p
- 2) 勝見俊三 ほか：燃料協会誌. 57 (4), 227 (1978)
- 3) 日本特許出願公開 2020-036579. 2020年3月12日
- 4) Caporaso, J.G. et al.: Nature Methods. 7, 335 (2010)
- 5) 川野秀一 ほか：スパース推定法による統計モデリング. 初版. 東京, 共立出版, 2018, 168p
- 6) Efron, B. et al.: An Introduction to the Bootstrap. 1st ed. New York, Chapman and Hall/CRC, 1994, 456p
- 7) Kawano, S. et al.: Computational Statistics & Data Analysis. 89, 192 (2015)
- 8) Wu, Y.J. et al.: Journal of Bioscience and Bioengineering. 115 (4), 424 (2013)
- 9) Fukushima, T. et al.: Water Science and Technology. 62 (6), 1432 (2010)
- 10) Fukushima, T. et al.: Journal of Water and Environment Technology. 5 (1), 37 (2007)
- 11) Vela, J.D. et al.: Water Research. 138, 241 (2018)
- 12) Albertsen, M. et al.: Plos One. 10 (7), e0132783 (2015)



福島寿和 Toshikazu FUKUSHIMA
先端技術研究所 環境基盤研究部
主幹研究員 博士(環境学)
千葉県富津市新富20-1 〒293-8511



川野秀一 Shuichi KAWANO
電気通信大学 大学院情報理工学研究科
准教授 博士(機能数理学)



中川淳一 Junichi NAKAGAWA
前 先端技術研究所 数理科学研究部
首席主幹研究員 博士(数理科学)



押木 守 Mamoru OSHIKI
北海道大学 大学院工学研究院 環境工学部門
准教授 博士(環境学)