

## Environmental metabolomics with data science for investigating ecosystem homeostasis



Jun Kikuchi <sup>a,b,c,\*</sup>, Kengo Ito <sup>a,b</sup>, Yasuhiro Date <sup>a,b</sup>

<sup>a</sup> RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

<sup>b</sup> Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

<sup>c</sup> Graduate School of Bioagricultural Sciences, Nagoya University, 1 Furo-cho, Chikusa-ku, Nagoya, Aichi 464-0810, Japan

Edited by David Neuhaus and Gareth Morris

### ARTICLE INFO

#### Article history:

Received 8 November 2017

Accepted 19 November 2017

Available online 21 November 2017

#### Keywords:

Ecosystem service  
Environmental diagnosis  
Metabolic profiling  
Macromolecular profiling  
Database  
Multivariate analysis  
Machine learning

### ABSTRACT

A natural ecosystem can be viewed as the interconnections between complex metabolic reactions and environments. Humans, a part of these ecosystems, and their activities strongly affect the environments. To account for human effects within ecosystems, understanding what benefits humans receive by facilitating the maintenance of environmental homeostasis is important. This review describes recent applications of several NMR approaches to the evaluation of environmental homeostasis by metabolic profiling and data science. The basic NMR strategy used to evaluate homeostasis using big data collection is similar to that used in human health studies. Sophisticated metabolomic approaches (metabolic profiling) are widely reported in the literature. Further challenges include the analysis of complex macromolecular structures, and of the compositions and interactions of plant biomass, soil humic substances, and aqueous particulate organic matter. To support the study of these topics, we also discuss sample preparation techniques and solid-state NMR approaches. Because NMR approaches can produce a number of data with high reproducibility and inter-institution compatibility, further analysis of such data using machine learning approaches is often worthwhile. We also describe methods for data pretreatment in solid-state NMR and for environmental feature extraction from heterogeneously-measured spectroscopic data by machine learning approaches.

© 2017 Published by Elsevier B.V.

### Contents

1. Introduction . . . . .	57
2. Basic and practical aspects of metabolomics . . . . .	58
2.1. Molecular scale of environmental metabolomics . . . . .	58
2.2. Metabolic profiling frequently used with 1D NMR . . . . .	58
2.3. Pulse sequences . . . . .	59
2.4. Databases and tools for NMR analysis . . . . .	60
2.5. Metabolite assignment and structure elucidation . . . . .	61
3. Investigation of environmental homeostasis . . . . .	63
4. Big data collection . . . . .	65
4.1. Animal systems . . . . .	65
4.2. Plant and algal systems . . . . .	66
4.3. Microbial ecosystems and geographical samples . . . . .	70
5. Data science . . . . .	73
5.1. Data pretreatment . . . . .	73
5.2. Data integration of heterogeneous measurements . . . . .	74
5.3. Multivariate analysis . . . . .	75

\* Corresponding author at: RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.

E-mail address: [jun.kikuchi@riken.jp](mailto:jun.kikuchi@riken.jp) (J. Kikuchi).

5.4. Machine learning.....	76
6. Conclusions and future perspective .....	79
Acknowledgements .....	80
References .....	80

## 1. Introduction

A natural ecosystem can be viewed as the interconnection between environmental and metabolic systems. For example, biologically-driven carbon flows not only through ecosystems on a global level but also through biochemical pathways at the organismal and population levels (Fig. 1). The significant increase in carbon emission from soil over the last 40 years is reported to have made a small contribution to global climate change [1]. A more direct contribution of human activity to these changes results from resource extraction by humans. For example, overfishing can completely restructure patterns across trophic levels in coastal ecosystems [2]. Advances in agricultural/fishery science are providing approaches to overcome food shortages arising from population growth in developing countries [3]. However, agricultural/fishery research in industrialized countries today emphasizes food processing and value-added products rather than the alleviation of food shortages in developing countries [4,5]. Considering the remarkable diversity in food culture across the world, an understanding of natural ecosystems in a variety of environments is important [6].

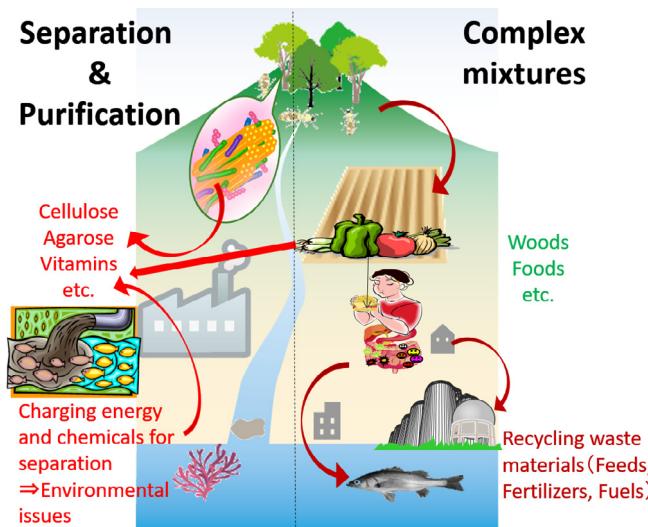
Based on reports of the direct and indirect effects of agriculture, forestry, and fisheries on natural ecosystems that produce biomass resources, we proposed the application of a recent analytical paradigm to ecosystem research [7–16]. Considerable efforts are being made to make a transition from oil refineries to bio-refineries [17]. Efforts are now under way to replace these underground resources of raw materials for industry with compounds and minerals obtained from renewable plant biomass [18]. Similarly, there is a major focus on the identification of renewable enzymes (reactive

catalysts) for enhancing biomass production [19,20]. For example, chemical engineering is concerned with the quantitative monitoring of reactions and yields of intermediates during the conversion of raw materials into biomass products [21,22].

Solution nuclear magnetic resonance (NMR) has long been used to determine the primary and stereochemical structures of synthetic [23,24] and naturally-occurring small molecules in organic and natural chemistry [25–29]. NMR analysis can determine the macromolecular structure for peptides [30,31], proteins [32,33], polysaccharides [34–36], and other molecular assemblies [37–39]. Methodological advances in multidimensional NMR analysis over the past four decades have extended its use to biological fields, in which it has made significant contributions. One example is the assignment of NMR signals in biomacromolecular complexes [40,41]. In particular, the structural analysis of high molecular-weight proteins has become possible using multi-dimensional NMR assisted with stable isotope labeling [42–46]. After sequential assignment, the secondary structures of proteins can be determined using conformation-dependent information from  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts [32,33,47–50]. Furthermore, recent progress in hardware development, such as advances toward higher magnetic fields [51,52], introduction of cryogenically cooled probes [53], and cost-effective low-field magnet NMR instruments [54], can offer new applications of NMR to various samples with high molecular complexity, such as crude extracts from biological tissues and biogeochemical samples from natural environments.

Here, we emphasize the advantages of NMR for analyzing the molecular complexity of biological extracts and crude biogeochemical samples. NMR can offer to deal with various and diverse samples from polar to non-polar solvent systems for both small and macromolecules; further, for interaction studies, it can use samples of various physicochemical states such as gas [55], sol [56], gel [57], and solid [58]. Furthermore, NMR offers site-[49,50] and atom-specific information [59], assisted by stable isotope tracing experiments. In addition, uniform stable isotope labeling in biological samples can open up the availability of NMR in two and three dimensional experiments such as protein NMR [60,61]. However, stable isotope labeling approaches cannot be applied to many environmental samples that are simply collected from natural ecosystems. In such cases, simple 1D- and 2D-NMR methods, such as 2D  $J$ -resolved spectroscopy ( $J$ -res) and heteronuclear single quantum coherence (HSQC), are useful for characterization of environmental samples. In addition, the interchangeability of NMR spectra between different laboratories is a distinct merit for cross-site analytical validity studies [62–66] with careful parameter settings [66,67]. For this purpose, using solid-state NMR for macromolecular complex and low-solubility samples is uniquely advantageous for various environmental samples. Importantly, this advantage is gained without using chromatography during measurements on environmental molecular complexity. Reproducibility in quantification is also a major advantage of NMR. Because NMR-based approaches can produce a number of data with high reproducibility and inter-institution compatibility, further analysis of such data using machine learning (ML) approaches is helpful.

Recently, solution NMR has also been used for characterizations and evaluations of biological mixtures with multivariate analysis to determine disease biomarkers [68,69], or evaluate nutrients in foods [11,70,71]. Comprehensive analysis of biological



**Fig. 1.** Conceptual framework for gathering ecosystem data using approaches based on separation and purification (left) and complex mixtures (right). Based on reports on the direct and indirect effects of agriculture, forestry, and fisheries on natural ecosystems, we propose the application of a recent engineering paradigm to ecosystem research. Considerable effort is now being made to transition from oil refineries to bio-refineries. Efforts to replace underground resources as raw materials for industry with compounds and minerals obtained from renewable plant biomass are under way.

Download English Version:

<https://daneshyari.com/en/article/7844457>

Download Persian Version:

<https://daneshyari.com/article/7844457>

[Daneshyari.com](https://daneshyari.com)