

Development of Four-Column Data Storage Model for Data-Manipulation of Greenhouse Gases and Soil Properties

Bence Matyas^{1,2}, György Matyas³, Mihaly Szendrei⁴, Ankit Singla^{2,5}, Yuhua Kong⁶, Janos Katal¹, Agnes Zsuposne Olah¹ & Kazuyuki Inubushi²

¹ Centre for Agricultural and Applied Economic Sciences, University of Debrecen, Debrecen, Hungary

² Graduate School of Horticulture, Bioresource Science, Chiba University, Matsudo, Chiba, Japan

³ Oracle APEX hosting, Individual entrepreneur, Hungary

⁴ Prec Cast Ltd., Sátoraljaúhely, Hungary

⁵ Department of Microbiology and Fermentation Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh, India

⁶ College of Forestry, Henan Agricultural University, Henan, China

* Correspondence: Kazuyuki Inubushi, Graduate School of Horticulture, Bioresource Science, Chiba University, Matsudo, Chiba, Japan. E-mail: inubushi@faculty.chiba-u.jp

Received: June 6, 2014 Accepted: July 2, 2014 Online Published: October 12, 2015

doi:10.5539/jfr.v4n4p115

URL: <http://dx.doi.org/10.5539/jfr.v4n4p115>

Abstract

Studies on greenhouse gas emissions and climate change require huge amounts of data. Generally, these data can be inducted directly from measuring devices or from different systems like geographical information systems and other specific systems, which are used for experimental data storage and management. Database technology has come a long way over the years, however even with these advances in technology and the availability of supercomputers processing large amounts of data from different data storage methods is still a challenge. In the present study, one physical data table was generated for data-manipulation and data storage using a web-developed platform. Data storage with four-column usage assures the possibility for continuous refinement during the research processes. It is because of the determination of new variables after the entity identification process.

Keywords: Data-manipulation, non-relational data storage, four columns data storage, greenhouse gases, physical data table

1. Introduction

Apart from the water vapour, carbon dioxide (CO₂), methane (CH₄) and nitrous oxide (N₂O) are important greenhouse gases (GHG) contributing 60, 20 and 6% towards global warming, respectively (IPCC, 2007). The global temperatures have increased 0.8°C over the past century and are predicted to increase another 1.1-6.4 degrees over the next century (Peters et al., 2013). The concentrations of CO₂, CH₄ and N₂O in the atmosphere has increased from pre-industrial levels of 280 ppm, 715 and 270 ppb to about 379 ppm, 1,774 and 319 ppb, respectively, in 2005 (IPCC, 2007). Although the concentration of CH₄ and N₂O is much lesser than CO₂, both gases have 25 and 298 times higher global warming potential than CO₂, respectively (IPCC, 2007). Ravishankara et al. (2009) reported that N₂O can also lead to depletion of the stratospheric ozone layer. The emissions of CO₂ resulting from the soil respiration and vegetation are the principal sources from which this gas enters into the atmosphere (Raich & Schlesinger, 1992). Rice is considered as one of the most important anthropogenic sources that accounts for 10-15% of the global CH₄ emission to the atmosphere (Cheng et al., 2008). Agricultural soil is a major source of N₂O (Hayakawa et al., 2009; Singla et al., 2013, 2014a; Singla & Inubushi, 2015). Several reports are available in the published literature which studied the emissions of these GHG under various soil types, cropping, irrigation and fertilizer management (Cabrera et al., 1994; Glatzel et al., 2004; Kong et al., 2013; Singla & Inubushi, 2014a, b; Singla et al., 2014b). It will be useful and time-saving if such large amount of data could be stored in one project with one platform usage. The development of a data-manipulation model could meet such requirement and provide a useful platform for data storage. Generally, data mining projects consist of three essential tasks: data collection, data preparation, and data modelling (Pyle,

1999; Westerman, 2001). The generation of such models also ensures possibilities for storing different data in one data table with one platform usage (Imhoff et al., 2003). Significant benefits of such system models manifest when many different types and huge amounts of data are required to be compiled in one project. Users could be able to input data as usual data tables (relational) while using data models. The integrated mechanisms of such models ensure that the system stores input data into one physical data table (non-virtual). The objective of the present study was to develop input correlations between soil attributes, their GHG production potential, and geographical information system (GIS) data of sampling parameters under one platform.

2. Method

2.1 Data Sites Used in the Study

The seven sampling sites from central Japan (34° 49'~36°36' N, 139°00'~54' E) and eight sites from eastern Hungary (47°55'~48°12' N, 21°23'~67' E) were selected for data storage. The physical parameters of soil sampling sites in both countries were taken as: land use, latitude/longitude, altitude, soil orders (FAO taxonomy), annual mean temperature (°C), annual mean precipitation (mm), and the soil moisture (%). The chemical and microbial properties of soils from both countries were taken as: pH (H₂O), EC (d S m⁻¹), total C {mg g⁻¹ dry soil(ds)}, total N (mg g⁻¹ds), C/N ratio, microbial biomass C (MBC) (µg g⁻¹ ds), microbial biomass N (MBN) (µg g⁻¹ ds), soluble organic C(SOC) (µg g⁻¹ ds), nitrate-N(NO₃⁻-N) (µg g⁻¹ ds), ammonium-N(NH₄⁺-N) (µg g⁻¹ ds), nitrite-N(NO₂⁻-N) (µg g⁻¹ ds), and MBC/ total C (%). The concentrations of CO₂, N₂O, and CH₄ were measured using gas chromatographs (GC-14B, Shimadzu, Japan) equipped with a thermal conductivity detector, electron capture detector and flame ionization detector, respectively (Singla & Inubushi, 2013). Equations in multiple regression models for the relationship between GHG production/consumption potential and the soil properties were stored in the database (Table 1). All the statistical analyses were completed using SPSS Statistics 20 (IBM, New York, USA).

Table 1. Multiple regression models for the relationship between the cumulative greenhouse gas (GHG) production/consumption (Y) and the soil properties (X)

Country	#	Equations	Adjusted R ²	p
Japan (n = 7)	#1	Y1 = 1130.0 X1 – 32.0	0.860	0.002
	#2	Y1 = 4.3 X4 + 298.2	0.666	0.016
	#3	Y1 = 0.435 X3 + 296.9	0.548	0.035
	#4	Y3 = 31.48 X10 – 355.8	0.510	0.043
Hungary (n = 8)	#5	Y1 = 1750.3 X2 – 574.0	0.942	< 0.001
	#6	Y1 = 291.0 X7 – 321.1	0.699	0.006
	#7	Y1 = 5.663 X3 – 301	0.877	< 0.001
	#8	Y2 = 32.7 X2 – 7.4	0.936	< 0.001
	#9	Y2 = 0.102 X3 – 1.9	0.803	0.002
	#10	Y2 = 5.1 X7 – 2.0	0.605	0.014
	#11	Y2 = 0.018 Y1 + 3.4	0.958	< 0.001
	#12	Y3 = 6.262 X10 – 44.98	0.609	0.014
	#13	Y3 = 3.3 X8 – 31.2	0.477	0.035
Japan and Hungary (n = 15)	#14	Y1 = 108.9 X7 + 55.01	0.408	0.006
	#15	Y2 = 13.64X2 + 0.867	0.255	0.032
	#16	Y3 = 1.137 X6 – 0.169 X5 – 22 X9 + 26.2	0.811	< 0.001

Y1: cumulative carbon dioxide (CO₂) production; Y2: cumulative nitrous oxide (N₂O) production; Y3: cumulative methane (CH₄) consumption; X1: nitrite-nitrogen (NO₂⁻-N); X2: ammonium nitrogen (NH₄⁺-N); X3: microbial biomass carbon (MBC); X4: microbial biomass nitrogen (MBN); X5: soluble organic carbon (SOC); X6: total carbon (TC); X7: total nitrogen (TN); X8: soil water content; X9: the ratio of MBC to TC; X10: soil carbon-nitrogen (C:N) ratio. #1 to #13 were analyzed by single regression method. #14, #15, and #16 were analyzed by stepwise regression method.

2.2 Database

Oracle Application Express (Apex) was chosen for geographical and experimental data storage. Apex provided opportunity to easy access because users could share works (by hierarchical levels) in same web platform without any software installation. Data of both countries were stored in Apex data tables at the first stage. The structure of entity identification was determined with key usage (Figure 1). The data tables which were used under GHG work space are available at web-link: <http://apex.oracle.com/pls/apex/f?p=4500:1000:8427548650555>

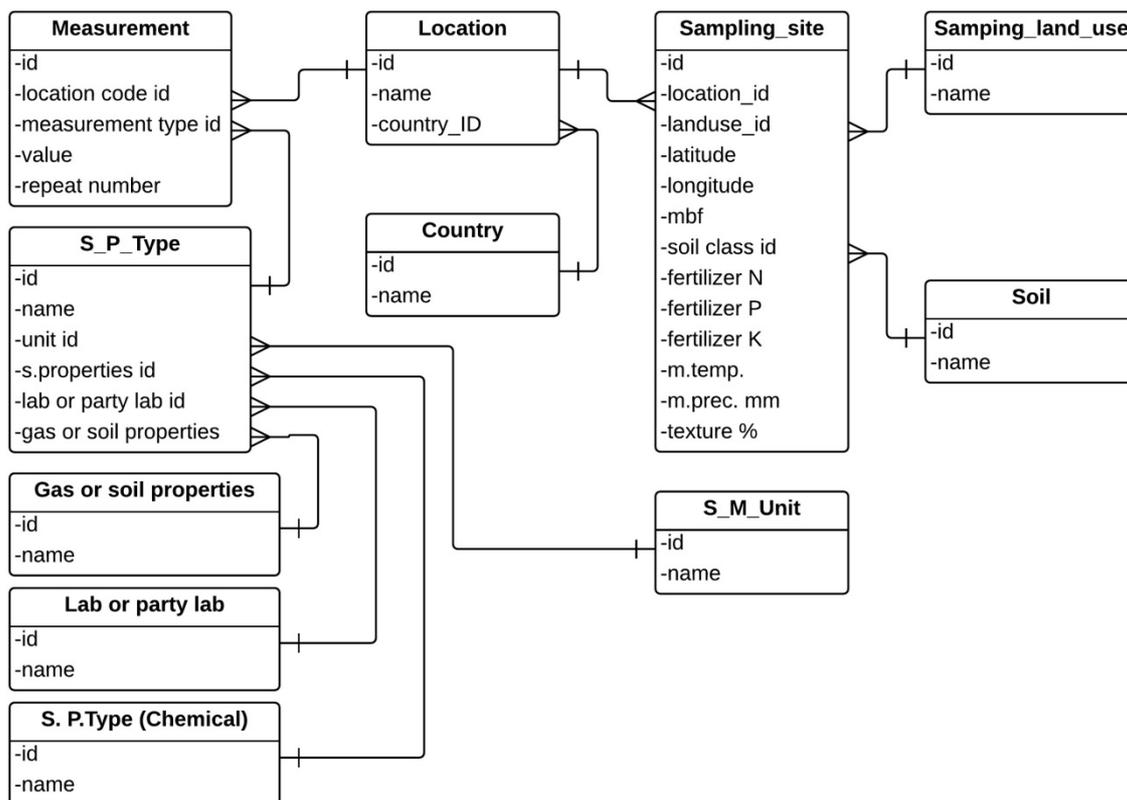


Figure 1. Measurement, Country, Location, Sampling_site, Sampling_land_use, Soil, S_P_Type, S_M_Unit, P. Type, Lab or party, Gas or soil properties tables were created to store data with APEX platform. ID columns were used to create relationships between these data tables. Number values could be entry with columns of ID, value, repeat number, latitude, longitude, altitude (mBf), ferti N, ferti P, ferti K, m.temp., m.prec., texture usage. Text could be entry with columns of name. Boolean options could be chosen with columns of lab or party, and gas or soil properties

2.3 Data-Manipulation Model

The developed data storage model called Joker Tao (JT) was used for storing data of different systems like data of GIS and input experimental data in one data table. After the data entry process in Apex data tables, data of CO₂, N₂O and CH₄ production/consumption potentials and the soil properties under different land-use types in central Japan and eastern Hungary (Kong et al., 2013) were stored in one physical data table with four column usage (Figure 2). The data lines arose manually or due to events (prompts from the model). Different data with attributes and correlations were stored in uniform storage format. Each of the attributes was key indexed. A multitude of huge amounts of data lines generated a qualitative leap in this system model. When data lines were created, it was done automatically which could be logical. Attributes of data storage, data maintenance and properties function both as cognitive and feature. At the same time, the above explanatory properties monitor and enforce the feedback environment. The most elementary entity of a set is an object, which is determined by three factors: attribute of describing uniqueness, attribute of specified and characterized entity, and value of attribute of specified and characterized entity (Table 2).

Table 2. Set theory of entity identification with the developed four columns data storage system model

Baseline			A U B (union)			A ∩ B (intersection)		
ID	Attribute	value	ID	Attribute	value	ID	attribute	Value
1	Cardinality	1	1	Cardinality	1	4	cardinality	1
1	set A	Yes	1	set A	yes	4	of set B	yes
2	Cardinality	1	2	Cardinality	1	4	of set A	yes
2	set A	Yes	2	set A	yes	A \ B		
3	Cardinality	1	3	Cardinality	1	ID	attribute	Value
3	set A	Yes	3	set A	yes	1	cardinality	1
4	Cardinality	1	4	Cardinality	1	1	of set A	Yes
4	set B	Yes	4	set B	yes	2	cardinality	1
4	set A	Yes	4	set A	yes	2	of set A	Yes
23	Cardinality	1	23	Cardinality	1	3	cardinality	1
23	set B	Yes	23	set B	yes	3	of set A	Yes

The record of this set theory shows how identification process was completed. Data of Baseline and Union main columns are same, because the event of base consist A and B sets determination. Cardinality was used for storing number data type as value. Intersection and minus procedure were used to represent query of dataset. Relationship between sets of A and B was determined how the model interprets the input data, because the possibility was guaranteed all data could be entity and attributes at the stage of data entry.

3. Results

Structure of data storage was determined with attributes of conceptual-functional name, functional logic and physical functionality. Tables of sampling country, sampling land use, sampling location and sampling sites were created to input attributes of sampling. Tables of measurements, measurement unit, types of measurement and types of soil properties were created to input measured data (Figure 1).

The developed integrated mechanism ensures that all input data could be stored in one physical data table with four columns usage. The process of insertion from data of multi-tables into one physical data table was determined with a Java code given below:

```
public static String getEntityName() throws Exception
{
    Connection conn = broker.getConnection();
    PreparedStatement pstmt = conn.prepareStatement("select * from joker tao");
    ResultSets = pstmt.executeQuery();
    inti = 0;
    while (rs.next()) {
        i++;
    }
    System.out.println("Number of records: " + i);
    broker.freeConnection(conn);
    return "";
}
```

The developed system model operates whichever nature, real (data from physical data table) or virtual (data from virtual data tables) objects. The structure of data storing was determined vertically and not horizontally as similar to Menon (2005). Actual query determines how the system interprets the data, because every stored data could be entity and attribute in this model. Identifier of Group_ID, Uniq_ID, Field_ID, Array_Index, Seek_Value,

Field_Value were created to separate attributes of different entities:

```
public static void insertJokerRow(Integer GROUP_ID, Integer UNIQ_ID, Integer FIELD_ID, Integer
ARRAY_INDEX, String SEEK_VALUE, String FIELD_VALUE) throws Exception {
if (GROUP_ID == null) pstmt.setNull(1, 2); else pstmt.setInt(1, GROUP_ID.intValue());
if (UNIQ_ID == null) pstmt.setNull(2, 2); else pstmt.setInt(2, UNIQ_ID.intValue());
if (FIELD_ID == null) pstmt.setNull(3, 2); else pstmt.setInt(3, FIELD_ID.intValue());
if (ARRAY_INDEX == null) pstmt.setNull(4, 2); else pstmt.setInt(4, ARRAY_INDEX.intValue());
if (SEEK_VALUE == null) pstmt.setNull(5, 12); else pstmt.setString(5, SEEK_VALUE);
if (FIELD_VALUE == null) pstmt.setNull(6, 12); else pstmt.setString(6, FIELD_VALUE);
pstmt.execute();
}
```

Data.txt file was created to complete the insertion process. The method reads input data from this text file and inserts them into one data table. Identifier of Group_ID, Uniq_ID, Array_Index, and Field_Value were used for entity identification:

```
public static void readFile() throws Exception
{
File f = new File("data.txt");
BufferedReader br = new BufferedReader(new FileReader(f));
while (br.ready()) {
String line = br.readLine();
int GROUP_ID = Integer.parseInt(line.substring(0, 10));
int UNIQ_ID = Integer.parseInt(line.substring(11, 21));
int ARRAY_INDEX = Integer.parseInt(line.substring(22, 32));
String FIELD_VALUE = line.length() > 32 ? line.substring(33, line.length()) : "";
insertJokerRow(Integer.valueOf(GROUP_ID), Integer.valueOf(UNIQ_ID), null,
Integer.valueOf(ARRAY_INDEX), null, FIELD_VALUE);
}
br.close();
}
```

Thanks to the above Java code the all data tables were converted to one physical data table. Concept of data storage was determined with columns of ID, Name, Attribute and Value (Table 3).

Table 3. Relationship determination (variable of Genus_proximus_id was used to hierarchical entity identification. Value of 1 in the fourth row means that the entity which was determined with number 2 ID records are in relation to entity which were determined with value of 100 ID records)

ID	ATTRIBUTE	INDEX	Value	explanation
1	1	1	100	group membership
1	2	1	Genus_proximus_id	
1	13	1	74	
2	1	1	100	
2	2	1	name	
2	13	1	71	71=alpha
3	1	1	100	
3	2	1	Sub_group_ID	sub group membership
3	13	1	74	code
4	1	1	100	Basic attribute

Associated attributes of this system model means that the system could initiate the search with a parallel processing mode to complete queries. Every single A and B search path is further engraved (marked) and path number increases.

$$\text{Path number} = \text{path number} + 1.$$

The model logs if path number is a large amount. The more frequent searches are priorities. The mentioned searching mode is completed with large amounts of data rows and cognitions usage. One data table was used to identify attributes (columns), entities (rows) and relationships between stored data. Relationships between stored data were determined with variables of ID and *genus_proximus_idusage* in four columns physical data table. Every record is a different entity in a relational data storage models (Menon, 2005), but a record with same ID values is the entity in this system model (Figure 2) like in some specific non-relational database systems (Imhoff et al., 2003; Sadalage & Fowler, 2013). The developed model operates data by the same operation (proceeding) which leaves the data type out of consideration.

4. Discussion

Differently, the relational data storage models with multi-table using (Imhoff et al., 2003), only one physical data table was used for data-manipulation. In this case, there is no need to create relationships between data tables or use horizontal column expanding, which could slow down the time of the query. More data rows were used in this model than in relational multi-level database systems, but path number usage could shorten the searching line during the time in which query is completed. Data storing with four-column usage assures the possibility for continuous refinement during the researching processes, because new variables could be determined after the entities identification process. Data attributes, soil sampling parameters, values of measurements and values of GHG production/consumption potentials were stored in one physical data table which could retrieve these data using integrated mechanisms in one platform. It offers a tool and opportunity for the researchers and developers to coordinate parallel teamwork and continuous data entry processes.

We created a customized code table, which allows any data whether entity, attribute, data connection or formula, to each be stored and managed under one physical data table. JT also handles relational data models differently; it allows data lines with the same ID values to identify a single entity. Similarly, non-relational models are also unique in JT; the inputted data models are not stored in an unstructured concept, making it easier to manage a huge amount of data without creating several applications. When compared to NoSQL models the speed of the JT system lies in its use of vertical data expansion and the elimination of the sequential search approach, which slows queries. JT is self-learning as all equations can be stored within the code table while maintaining the characteristics of the balance method similar to data stored in the physical data table.

The present study showed that the development of four columns data storage model using one physical table to store data from multi tables will be useful for research and development projects in any discipline.

Acknowledgements

This study was financially supported by the Hungarian-Japanese intergovernmental S&T cooperation program for the year 2009–2011. The corresponding author is thankful to György Mátyás, the idea owner of JT. The first version of JT was developed (R. number: INNO-1-2008-0015 MFB- 00897/2008) in an EU application. The authors are also thankful to the Call Tec Consulting Ltd. which was the first company in Hungary to obtain highest Oracle ranking.

References

- Cabrera, M. L., Chiang, S. C., Merka, W. C., Pancorbo, O. C., & Thompson, S. (1994). Nitrous oxide and carbon dioxide emissions from pelletized and non pelletized poultry litter incorporated into soil. *Plant and Soil*, 163, 189-195. <http://dx.doi.org/10.1007/BF00007967>
- Cheng, W., Sakai, H., Hartley, A., Yagi, K., & Hasegawa, T. (2008). Increased night temperature reduces the stimulatory effect of elevated carbon dioxide concentration on methane emission from rice paddy soil. *Global Change Biology*, 14, 644-656. <http://dx.doi.org/10.1111/j.1365-2486.2007.01532.x>
- Glatzel, S., Basiliko, N., & Moore, T. (2004). Carbon dioxide and methane production potentials of peats from natural, harvested and restored sites, Eastern Quebec, Canada. *Wetlands*, 24, 261-267. [http://dx.doi.org/10.1672/0277-5212\(2004\)024\[0261:CDAMPP\]2.0.CO;2](http://dx.doi.org/10.1672/0277-5212(2004)024[0261:CDAMPP]2.0.CO;2)
- Hayakawa, A., Akiyama, H., Sudo, S., & Yagi, K. (2009). N₂O and NO emissions from an Andisol field as influenced by pelleted poultry manure. *Soil Biology and Biochemistry*, 41, 521-529. <http://dx.doi.org/10.1016/j.soilbio.2008.12.011>

- Imhoff, C., Galem, N., & Geiger, J. G. (2003). *Mastering data warehouse design: relational and dimensional techniques*. Indianapolis, Indiana: Wiley Publishing, Inc..
- IPCC. (2007). Changes in atmospheric constituents and in radiative forcing. In *Climate change, the physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*, Cambridge University Press, Cambridge, UK.
- Kong, Y., Nagano, H., Katai, J., Vago, I., Olah, A. Z., Yashima, M., & Inubushi, K. (2013). CO₂, N₂O and CH₄ production/consumption potentials of soils under different land-use types in central Japan and eastern Hungary. *Soil Science and Plant Nutrition*, 59, 455-462. <http://dx.doi.org/10.1080/00380768.2013.775005>
- Menon, R. M. (2005). *Expert Oracle JDBC Programming*. New York: Springer-Verlag.
- Peters, G. P., Andrew, R. A., Boden, T., Canadell, J. G., Ciais, P., Quéré, C. L., ... Wilson, C. (2013). The challenge to keep global warming below 2°C. *Nature Climate Change*, 3, 4-6. <http://dx.doi.org/10.1038/nclimate1783>
- Pyle, D. (1999). *Data preparation for data mining*. San Francisco, CA: Morgan Kaufmann,
- Raich, J. W., & Schlesinger, W. H. (1992). The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus*, 44(B), 81-99. <http://dx.doi.org/10.1034/j.1600-0889.1992.t01-1-00001.x>
- Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N₂O): the dominant ozone-depleting substance emitted in the 21st century. *Science*, 326, 123-125. <http://dx.doi.org/10.1126/science.1176985>
- Sadalage, P. J., & Fowler, M. (2013). *NoSQL Distilled: A brief guide to the emerging world of polyglot persistence*. Addison-Wesley: Pearson Education.
- Singla, A., Dubey, S. K., Iwasa, H., & Inubushi, K. (2013). Nitrous oxide flux from Komatsuna (*Brassica rapa*) vegetated soil: A comparison between biogas digested liquid and chemical fertilizer. *Biology and Fertility of Soils*, 49, 971-976. <http://dx.doi.org/10.1007/s00374-013-0796-4>
- Singla, A., & Inubushi, K. (2013). CO₂, CH₄ and N₂O production potential of paddy soil after biogas byproducts application under waterlogged condition. *International Journal of Agriculture Environment and Biotechnology*, 6, 233-239.
- Singla, A., Dubey, S. K., Singh, A., & Inubushi, K. (2014b). Effect of biogas digested slurry-based biochar on methane flux and methanogenic archaeal diversity in paddy soil. *Agriculture Ecosystems and Environment*, 197, 278-287. <http://dx.doi.org/10.1016/j.agee.2014.08.010>
- Singla, A., & Inubushi, K. (2014a). Effect of biochar on CH₄ and N₂O emission from soils vegetated with paddy. *Paddy and Water Environment*, 12, 239-243. <http://dx.doi.org/10.1007/s00374-014-0950-7>
- Singla, A., & Inubushi, K. (2014b). Effect of biogas digested liquid on CH₄ and N₂O flux in paddy ecosystem. *Journal of Integrative Agriculture*, 13, 635-640. [http://dx.doi.org/10.1016/S2095-3119\(13\)60721-2](http://dx.doi.org/10.1016/S2095-3119(13)60721-2)
- Singla, A., Sakata, R., Hanazawa, S., & Inubushi, K. (2014a). Effect of biogas digested slurry based-biochar and digested liquid on N₂O, CO₂ flux and crop yield for three continuous cropping cycles of komatsuna (*Brassica rapa* var. perviridis). *Biology and Fertility of Soils*, 50, 1201-1209. <http://dx.doi.org/10.1007/s00374-014-0950-7>
- Singla, A., & Inubushi, K. (2015). Effect of slag-type fertilizer on N₂O flux from komatsuna vegetated soil and CH₄ flux from paddy vegetated soil. *Paddy and Water Environment*, 13, 43-50. <http://dx.doi.org/10.1007/s10333-013-0405-z>
- Westerman, P. (2001). *Data warehousing: using the Wal-Mart model*. (1972). *Ethical standards of psychologists*. Washington, DC: American Psychological Association. <http://dx.doi.org/10.1016/b978-155860684-5/50001-6>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).