Developing a Web-based System for Computing Pre-Harvest Residue Limits (PHRLs)

Han Sub Chang · Hey Ree Bae · Young Bae Son · In Ho Song · Cheol Ho Lee · Nam Geun Choi · Kyoung Kyu Cho · Young Gu Lee

**Abstract** This study describes the development of a web-based system that collects all data generated in the research conducted to set pre-harvest residue limits (PHRLs) for agricultural product safety control. These data, including concentrations of pesticide residues, limit of detection, limit of quantitation, recoveries, weather charts, and growth rates, are incorporated into a database, a regression analysis of the data is performed using statistical techniques, and the PHRL for an agricultural product is automatically computed. The development and establishment of this system increased the efficiency and improved the reliability of the research in this area by standardizing the data and maintaining its accuracy without temporal or spatial limitations. The system permits automatic computation of the PHRL and a quick review of the goodness of fit of the regression model. By building and analyzing a database, it also allows data accumulated over the last 10 years to be utilized.

**Keywords** Pre-harvest residue limit (PHRL) · Regression analysis · Web-based · Automatic computation · System development · Pesticide residues

1 Introduction

1.1 Research background

According to Jeong et al. (2004), pesticides are indispensable in modern agriculture; they clearly increase agricultural productivity by protecting agricultural products and improving their quality. However, compounds used in agriculture to effectively prevent a variety of diseases, insect pests, and weeds have diverse chemical structures and, depending on the environmental conditions, may remain in the environment and in agricultural products themselves (Jeong et al. 2004). Because consuming agricultural products with excessive pesticide residues may cause health problems, a Maximum Residue Limit (MRL) for agricultural products is set, managed, and supervised by the state.

According to Kim et al. (2010), the pesticide safety standards are set by calculating MRL values for each food according to the intake amount of that food. The theoretical MRL for a food is based on the acceptable daily intake of pesticides that a person receives through food, and it is based on various toxicity test scores that are submitted for safety evaluation as part of the pesticide registration process. To control the safety of pesticides, the Rural Development Administration reviews these toxicity test score reports and limits the acceptable daily intake of the pesticide. In general, the theoretical MRL value for a food is computed by multiplying 80% of the acceptable daily intake of the pesticide by the mean weight of the population and dividing by the daily intake per person (Kim et al. 2010).

For preventive agricultural product safety control, the National Agricultural Products Quality Management Service sets a pre-harvest residue limit...
(PHRL) in connection with the pesticide MRL of the Korean Food and Drug Administration so that the possibility of producing and distributing unsuitable agricultural products is reduced, thereby ensuring a safe supply of agricultural products. According to Choi et al. (2000, 2002), Kim et al. (2002), Seong et al. (2004), and Lee et al. (2008), the National Agricultural Products Quality Management Service sets the PHRL for the purpose of assuring agricultural product safety in the production stage. Furthermore, it evaluates the safety of agricultural products in the production stage according to the PHRL, investigates the amount of pesticide residues sprayed during the pre-harvest period, predicts the amount of pesticide residues at the time of harvest by computing the biological half-life and the decay constant, and prevents the distribution of unsuitable agricultural products that may exceed the MRL.

1.2 Study goals

Because collaboration among multiple specialized research institutes with expertise in the characteristics of residual toxic substances in crops is essential to setting the PHRL, the PHRL computation process uses statistical techniques to analyze data collected from approximately 25 research institutes each year. In addition, reviews and evaluations are conducted to increase the accuracy and reliability of the results. However, if general-purpose software, such as Microsoft Excel, is used for the data collection and PHRL computation, longer processing times and an extensive amount of data are required and simultaneous processing becomes impossible. This study therefore attempted to develop an automated PHRL computation system that uses statistical techniques to analyze the data collected from each research institute during the process of establishing the PHRL. Furthermore, we established a web-based database to make data access and statistical analysis possible without temporal or spatial limitations (i.e., wherever there is an internet connection).

2 Introduction of the PHRL setup

2.1 Theoretical background

The National Agricultural Products Quality Management Service analyzes whether toxic substances such as pesticides remain in agricultural products at various stages, including production, storage, and before shipment. Moreover, it applies the MRL according to the Food Sanitation Act to determine the safety of pesticides. This service allows farmers to freely postpone shipments of unsuitable products that exceed the MRL, convert the usage of these products, or discard them entirely. However, it has heretofore been impossible to prevent the distribution of unsuitable agricultural products before consumption because the pesticide safety investigation that is conducted during the stages of market distribution and sales yields its results only after sales to the final consumer. The Food Sanitation Act of 1999 did not set the MRL separately for each investigation stage; as a result, there was an unacceptable increase in the amount of unsuitable agricultural products when the MRL was unilaterally applied at the time of investigation without considering the decrease in the amount of pesticide residues after spraying. Accordingly, the National Agricultural Products Quality Management Service set a reasonable PHRL that takes into account the decrease in the amount of pesticide residues in pre-harvest agricultural products and proposed it to the Ministry of Food, Agriculture, Forestry, and Fisheries. This contributes to the production and supply of agricultural products that are safe for human consumption.

2.2 Principles of pesticide dissipation

Pesticides dissipate through several routes. Although these routes differ depending on the environment, the main routes of pesticide dissipation are enzymatic decomposition, nonenzymatic decomposition, volatilization, photolysis, and simply being washed away. In the case of agricultural products, pesticide concentrations decrease as plants grow. Productive excretion, one of the defensive mechanisms of plants against external foreign substances, also contributes to their dissipation. In general, the rate of pesticide dissipation in a given space is known to follow a first-order reaction. This means that the reaction velocity is linearly proportional to the concentration of the reactant. When pesticide A disappears in a given space through a certain route, the dissipation velocity can be expressed as follows:
In Formula (1), \( \frac{d[A]}{dt} \) represents the change in the concentration of pesticide A over time, i.e., the dissipation velocity; the dissipation velocity is linearly proportional to the concentration of the pesticide \([A]\). Here, \( k \) is the reaction constant, which represents a proportional constant. Formula (1) can be rewritten in the form of an equation by separating the two variables of concentration and time \( t \), as follows:

\[
d\frac{[A]}{[A]} = kd\ dt \quad \text{.....(2)}
\]

Integrating each side yields Formulas (3) and (4), as follows:

\[
\int \frac{d[A]}{[A]} = \int -k\ dt \quad \text{.....(3)}
\]

\[
\ln[A] = -kt + C \quad \text{.....(4)}
\]

In Formula (4), \( C \) is the constant of integration. Formula (4) signifies that \( \ln[A] \), the logarithmic function of the concentration, is a linear function of time. \( C \), the constant of integration, can be found as shown below. The initial concentration, which is the pesticide concentration when \( t = 0 \) in Formula (4), can be represented in terms of \( [A]_0 \) as \( C = \ln[A]_0 \).

\[
\ln[A]_0 = C \quad \text{.....(5)}
\]

Substituting Formula (5) into Formula (4) yields Formulas (6) and (7):

\[
\ln[A] = -kt + \ln[A]_0 \quad \text{.....(6)}
\]

\[
\ln\left[\frac{[A]}{[A]_0}\right] = -kt \quad \text{.....(7)}
\]

In the more common form that includes two points, this relationship can be expressed as follows:

\[
\ln\left[\frac{[A]_2}{[A]_0}\right] = -k(t_2 - t_1) \quad \text{.....(8)}
\]

Formula (6) is represented in exponential form as Formula (9).

\[
[A] = [A]_0 e^{-kt} \quad \text{.....(9)}
\]

We designed this study in such a way that a researcher can compute the regression coefficient after automatically specifying Formula (9) based on the data entered in the system.

### 2.3 Principles of the PHRL setup for agricultural products

The persistence of pesticides in agricultural products differs depending on the conditions of pesticide processing, crop growth, climate, and the rate of residue decrease over time. The level of pesticides remaining in pre-harvest agricultural products shows irregular fluctuations due to pesticide spraying and, after the final pesticide application, a steady decline. The time window over which pesticides are used and the number of applications suggested by the Rural Development Administration in the Pre-Harvest Interval (PHI) for agricultural chemicals are set so that at their time of shipment, agricultural products do not exceed the MRL. The PHRLs also are set so that at their time shipment, agricultural products do not exceed the MRL established by the Food Sanitation Act. The rate at which the residual amount decreases depends on the decomposition characteristics of the pesticide under consideration. The PHRL is based on the number of days from the time a sample is collected to the time of shipment, thus reflecting the decrease in the residual amount following the last pesticide application.

### 2.4 Necessity of improving the methods of data collection and analysis

Considering that the periods of growth and development of subject crops are concentrated, whereas the area of cultivation is dispersed, the National Agricultural Products Quality Management Service and several specialized research institutes collaboratively conduct PHRL studies, draw conclusions, and report in writing to the Experiment Research Institute at the National Agricultural...
The manual process used to produce such data files carries an inherent risk of invalid results stemming from errors in data entry, and an absence of process validation is inherent in the manual process. Moreover, the issue of possible duplication of the data arising from researchers and the researcher-in-charge having different Excel files in the process of sending, reviewing, and modifying the data is an additional factor that delays the final conclusions of the study. Although a process is needed in which data collected from multiple research institutes participating in the research project undergo regression analysis and are confirmed and reviewed in person by the researcher-in-charge, it results in decreased work efficiency because it is a repeatedly conducted, continuous manual process. In addition, in cases in which it is necessary to modify the experimental design according to weather conditions and/or the state of growth and development of the crop while conducting field experiments, it is possible to compute an incorrect value because of unintended mistakes occurring due to formulas or functions being deleted or modified. The accuracy of this stage of data analysis typically depends on the ability of the researcher to utilize the software in connection with the Microsoft Excel file distributed to each research institute. Thus, in the standard method of research assignment management, problems can arise from the use of Excel files for data collection and analysis.

Consequently, this study intended to increase the efficiency of the research work and improve the accuracy and reliability by managing a database based on a web-based device of statistical analysis.
instead of using Excel files for data collection and analysis management.

3 System establishments

3.1 Overview

The newly developed system is composed of five parts. In the research assignment information entry part of the system, the data to be processed for the research year under study are entered by the research institute in charge. In the basic data entry part, the results of the research conducted by each research institute are entered in person by each co-researcher. In the regression analysis, the minimum of the decay constant is computed after executing a regression analysis using a predetermined algorithm and the entered research data, and the goodness of fit of the regression model is reviewed. In the PHRL administration part of the system, the daily PHRL is automatically computed using the minimum of the decay constant found in the regression analysis. Lastly, in the authority management part of the system, analysts and administrators are granted different levels of authority over the system. A system was developed that enables all data resulting from the study to be entered and analyzed on a web server. A diagram of the developed system is shown in Fig. 2.

3.2 Research assignment information entry

Prior to the development of the system described in this study, research assignments were conducted by the research institute in charge; this institute selected the crops and pesticides intended for each year’s research assignments and instructed the collaborative research institutes by means of an official document. However, the new system permits research assignments to be initiated as soon as the research institute in charge enters the research assignment information for the year under study and the designated research institute confirms it. In the research assignment information entry part of the

Fig. 2 Diagram of the system
system, the research year, the experimental crop item, the pesticide, and the research institute are designated. In addition, the data accumulated over the last 10 years are incorporated into a database that can be accessed by research year and research institute.

### 3.3 Basic data entry

The basic data entry part of the system was designed to enable the co-researcher designated during research assignment information entry to conduct the research assignment and enter all resulting data. The basic data entered includes information concerning the residual concentration, the recovery, limit of detection (intended for validation of the analytical method), the weight (intended for identifying the state of growth and development of the products of the experimental crop), and the weather conditions. Excel file uploading was enabled in an effort to lessen the burden of entering a large amount of data concerning the concentration of pesticide residues.

### 3.4 Regression analysis

The regression analysis part of the system consists of five parts: estimation of the regression coefficient of the regression model according to the results of the regression analysis; automatic calculation of the recovery and limit of detection; the rate of change in weight of the products of the experimental crop and the corresponding graph; the weather chart and its corresponding graph; and computation of the PHRL. This part of the system purports to compute the minimum of the decay constant through regression analysis and verify the goodness of fit of the estimated regression model. Hence, a web-based algorithm (SPSS version 18.0) was established. An algorithm was designed by means through which the basic data can be entered into the system and a regression analysis can be conducted; the minimum of the decay constant is computed automatically. A regression model is suggested for each experimental section and processed both at the standard rate of spraying of the pesticide and at twice the rate. To increase the reliability of the results, the mean and the standard deviation are computed. The process was designed in such a way that elements such as the intercept (the regression coefficient of the regression equation), the slope (the regression coefficient of the regression equation), the coefficient of determination (R-squared), the studentized residual (ri), and the p-value are automatically calculated and displayed to evaluate the goodness of fit of the suggested regression model. The system was designed so that outliers that escape the range of the studentized residual set for the purpose of identifying outlying residual concentration values are marked in red. Also, the regression curve can be graphed to identify the rate of decline in the amount of a pesticide residue depending on daily fluctuations. The biological half-life of the pesticide (the time required for the concentration of pesticide to decrease to half of its initial concentration) was computed by dividing the value of ln2 by the decay constant (k).

To find the regression equation \([A] = [A]_0 e^{-kt}\), the intercept term \([A]_0\), and the slope \(-k\) (which are the regression coefficients of the regression equation) are obtained by least square estimation. The coefficient of determination (R-square) is automatically computed and identified as a measure of the goodness of fit of the estimated regression equation. To test the significance of the regression equation, its p-value was determined using an ANOVA table at the 0.05 significance level.

Because the above-mentioned regression model includes several assumptions (including homoscedasticity), these assumptions should be validated before interpreting the results of the regression analysis. For this reason, a residual analysis is necessary. In particular, because the studentized residual \((ri)\) is typically distributed randomly between -2 and +2, the system enables the researcher to remove outliers from variables by marking in red those values for which the value of the studentized residual \((ri)\) is outside the range of the absolute value of 2.

The value that this study ultimately intends to compute is the decay constant, which is necessary for proposing the PHRL. Because the regression coefficient \((-k)\) has the upper and lower bounds at the 95% confidence level, the system was designed to compute the decay constant by selecting the upper bound of the regression coefficient for the sake of the safety of consumers using the agricultural products.

### 3.5 PHRL administration

We designed the system in such a way that, once a specific assignment is designated to a co-researcher
according to what was entered in the research assignment information entry stage, the co-researcher then enters the basic data (concentrations of pesticide residues, the recovery, the limit of detection, the weather chart, and the weight) and conducts the regression analysis, and the minimum of the decay constant is automatically computed. The daily PHRL is computed automatically according to the computed minimum of the decay constant. The computation formula is as follows: \( \text{PHRL} = \text{MRL} \times e^{(\text{decay constant} \times \text{date before shipment})} \). In this formula, the decay constant is a proportional constant that represents the decline in the amount of pesticide residues remaining in an agricultural product after the pesticide is sprayed on the crop; it depends on the type of crop, the pesticide, and the time elapsed.

The system was also designed in such a way that when the appropriate PHRL does not exist, the minimum of the decay constant for the primary categories, the secondary categories, and the total of the agricultural products to be applied with the next ranking standard is used. In addition, because the PHRL is set according to the pesticide’s MRL announced by the Korean Food and Drug Administration, a renewal function for the PHRL was set up so that the PHRL is changed automatically when the MRL of a pesticide is changed. In this manner, the function remains updated in accordance with the latest MRL, and the setup of the PHRL for the current research year can be accessed and checked.

### 3.6 Authority management

The authority management part of the system was designed to distinguish the authorities over executing the research assignments that are granted to different researchers. Authorities to enter, modify, and analyze the data were constructed to be separate so that the results cannot be modified arbitrarily. Moreover, the roles of the collaborative research institutes and the research institute in charge are distinguished by differentiating functions. In addition, the research period was designed to be added and deleted according to the yearly changes in the collaborative research institutes. The flow within the system from researcher to administrator is shown in Fig. 3.

#### Fig. 3 System flow chart


<table>
<thead>
<tr>
<th>User</th>
<th>Data Input</th>
<th>Analysis</th>
<th>Maximum Residue Limit(MRL)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researcher</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need the update?</td>
<td>No</td>
<td>Project registration or not?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Free Registration or not</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Administrator</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for authority</td>
<td>No</td>
<td>Free Registration or not</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free Registration or not</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Free Registration or not</td>
<td>No</td>
</tr>
</tbody>
</table>

### 4 Results

#### 4.1 Comparison with existing methods

The results of a study of the residual concentration of the pesticide Fenarimol in green chili peppers conducted by the Experiment Research Institute at the National Agricultural Products Quality
Management Service in 2010 were used to compare the existing method with the newly established system. The results showed a difference in the third decimal place between the regression coefficient intercept computed by the standard data processing method using Microsoft Excel (0.3912) and the intercept computed by the newly established system (0.3944). However, because the intercept term does not affect the setup of the PHRLs, this is not problematic. This result seems to be due to the fact that the rounding method used in Excel differs somewhat from that used in the newly designed system. When computing the biological half-life of Fenamidol and its PHRL, the standard method presents the inconvenience of having to re-calculate according to a fixed formula in a different Excel sheet. Additionally, neither the studentized residual (ri) nor the p-value, parameters that must be determined to confirm the goodness of fit and the significance of the logarithmic function regression equation, can be computed in Excel. Hypothesis validation of a regression model in an Excel sheet requires the use of both the t distribution and F distribution tables, and this requires the preparation of six Excel sheets for each pesticide. Because PHRLs are set for approximately 100 pesticides each year, the number of Excel sheets the experiment research institute in charge needs to retain and check in a year is therefore approximately 600. The number of Excel sheets concerning the 681 PHRLs announced in the 2010 standard is in the thousands. However, the newly established system can resolve all of these issues; using the new system, the researcher need only enter the experimental results in order for all the calculations and regression analyses to be executed automatically.

4.2 Web server publishing

According to Choi and Han (2004), the National Agricultural Products Quality Management Service has been developing and managing the Laboratory Information Management System (LIMS) since 1999, pushing ahead with the investigation and analysis of agricultural product safety with efficiency and credibility. For this reason, the system developed in this study published a web server to be enacted in the LIMS.

In the web server system, the research year and the research institute can be designated, and the experimental crop item and the pesticide can be set according to the experimental design of the year concerned. Fig. 4 shows a typical research assignment information entry and access screen showing research assignment information from 1999 to the present.

![Image of research assignment information entry and access screen]

In this system, the co-researcher is able to enter research results, such as the concentrations of pesticide residues, after identifying the research assignment information entered by the researcher-in-charge. Fig. 5 shows the basic data on residual concentration entered by the researcher.
The system was designed so that the recovery and the limit of detection used to validate the analytical method are entered, and the entered data is automatically processed according to the computation formula set at the time of the regression analysis.

In the weight and weather chart areas, the individual weight of the experimental crop products, humidity, temperature, precipitation, and other experimental parameters are entered according to the number of days elapsed since pesticide spraying. A typical entry screen is shown in Fig. 6.

4.3 Regression and data analysis

As described above, the system was designed so that after all basic data are entered, the regression analysis is executed, and the results are computed automatically. The pesticide dissipation trend can be seen in the table data, which show the elements necessary to evaluate the goodness of fit of the automatically computed model, the minimum of the decay constant, and the regression curve graph according to the exponential regression equation. The elements for evaluating the goodness of fit of the regression model are computed automatically as a result of the regression analysis. The regression curve graph is shown in Fig. 7.
The recovery, the limit of detection, the weather chart (including temperature, humidity, precipitation, and other factors), and a graph of the daily individual growth of the crop can be checked using the data entered previously. To establish and evaluate a database for the cultivation environment and the weather chart, which is an important part of carrying out the research assignment, the temperature and humidity can be graphed as shown in Fig. 8. In addition, because the crop’s growth rate is sometimes the largest factor in pesticide dissipation, daily fluctuations in growth can be graphed in order to identify changes in the rate of growth.
When standard Microsoft Excel software was used to manage the research results, a file that includes separate function formulas was used in order to compute the daily PHRL. When the system developed in this study is used, the PHRL is automatically computed for each day according to the algorithm defining the minimum of the decay constant computed from the regression analysis using the above-mentioned pesticide residue data. In the latter case, a PHRL renewal function was designed so that it can always be updated in accordance with the latest MRL for the food announced by the Korean Food and Drug Administration. Moreover, the system was designed so that the partial and complete yearly data from the PHRL Administration can be checked and used by anyone authorized to utilize and search the data. A screen showing the result of accessing the PHRLs is shown in Fig. 9.

5 Conclusions

The PHRL, which is computed automatically by the system described in this work, is a preventive standard that sets the residual amount of each pesticide permitted at a certain time before harvest so that the residual amount of pesticide at the time of harvest of the agricultural product does not exceed the MRL announced by the Korean Food and Drug Administration. About 100 new standards are created each year. The final proposal is made after estimating the regression coefficient of the regression model using regression analysis of results for individual pesticides and confirming the linearity and homoscedasticity of the computed regression equation. The PHRL, which is proposed scientifically, makes significant contributions both to consumer health and to preserving the producers' income. The development of the web-based system described here for automatic computation of the PHRL is significant in that the first and only attempt in the world was made in Korea.

As a result of developing and establishing the system described in this paper, the PHRL is automatically computed, and the goodness of fit and significance of the regression model are quickly confirmed through standardization and flawless maintenance the data. This results in increased efficiency and improved reliability of the PHRL research. The data accumulated over the last 10 years was also incorporated into a database and analyzed; it is therefore now possible to utilize it in setting the direction of future studies.

A number of advantages in executing and managing research assignments are associated with the establishment of a web-based system for automatic PHRL computation. First, access through a web server using a computer with an internet connection permits analysis of the data as soon as it is entered without temporal or spatial limitations; this shortens the data collection time, facilitates sharing, and maintains data security by setting access authority. Second, duplicate Excel file generation due to repeated reviewing of data becomes controllable, and discordances caused by duplicate data are prevented. Third, the integrity and standardization of the research database is ensured through an error test of the data during the regression phase of the analysis. Fourth, because the regression analysis is carried out using a web-based statistical analysis system,
assessments of goodness of fit and significance for regression models such as regression analysis, studentized residuals, coefficient of determination, p-values, and other measures can now be calculated automatically following a pre-defined process. Within this process, a warning message is displayed for data that do not conform to the preset standard, enhancing the reliability of the analysis.

In general, a regression model makes several assumptions, such as linearity, homoscedasticity, independence, and normality. Although this study established a system to test linearity and homoscedasticity, testing procedures for independence and normality were not designed, and this omission needs to be remedied. Further refinement of the system will also be necessary following modification of the experimental design for future collaborative research with related institutes such as the Rural Development Administration.

Acknowledgments

The authors acknowledge SPSS Korea co. Ltd. And especially Mr. Byoung Moo Lee, Mr. Hyun Sub Cha, Mz. Hyo Young Yang for the development of statistical techniques.

References


The Global Knowledge Linkage Structures of the Agricultural Sector Pertinent to Information Technology: A Triple Helix Perspective

Md. Dulal Hossain · Junghoon Moon · Young Chan Choe

Abstract The development of informatization impacts all sectors, including agriculture. Agricultural informatization builds the knowledge linkage structures of agricultural innovation systems globally. This study investigated the global knowledge linkage structures in agricultural innovation pertinent to information technology (IT) for agricultural research and development (R&D) investments and activities. We explored the longitudinal trend of systemness within the networked research relationships in the triple helix (TH) of the university, industry and government (UIG). We collected data from publications in the Science Citation Index (SCI), the Social Sciences Citation Index (SSCI), and the Arts and Humanities Citation Index (A&HCI) to analyze the TH network dynamics. We also performed a scientometrics analysis to quantitatively identify the knowledge and insights of global agricultural innovation structures. These results could be informative for individual countries. Our findings reveal that the global knowledge linkage structures in the agricultural sector that are pertinent to IT fluctuate widely and fail to increase the capacity of agricultural innovation research due to a neglect of the network effects of the TH dynamics of UIG.

Keywords Triple helix · Agricultural innovation · IT · Global knowledge linkage · University · Industry · Government

1 Introduction

Globally, agriculture remains one of the dominant economic sectors, and agricultural research related to IT plays a vital role in building most of the world’s economies. Agricultural research has enormous economic and social benefits, and its cost-benefit ratios can exceed those of nearly any other form of investment (Chambers and Ghildyal 1985). Innovations constitute an important economic output, and one prerequisite for innovation is knowledge linkage (Chatziparadeisis 2003). Innovation systems exist within firms and across the interfaces between institutional agents such as universities, industries, and government agencies. The knowledge production, diffusion, transfer, acquisition, absorption/assimilation, and use of these networks add value that should be measured (Chatziparadeisis 2003) to map the network effects of university, industry and government (UIG) in an attempt to build the knowledge linkage structures of innovation systems.

Numerous co-authorship studies in the international literature have explored the TH models of individual countries or groups of countries (Leydesdorff and Sun 2009; Park et al. 2005; Shapiro 2007). However, little is known about the global
trends in knowledge linkage structures of innovation in relation to the TH dynamics of UIG. In particular, the measure of the agricultural innovations related to IT is unexamined by researchers using the TH model. These gaps in the literature are the prime motivators for the research we present here.

The TH model of UIG relations (Etzkowitz and Leydesdorff 2000) permits us to study both the bilateral and the trilateral interactions among these three institutional domains in an innovation system. Therefore, in this study, we follow the design of Leydesdorff and Sun’s (2009) study by operationalizing relationships in terms of the co-authorships in the literature contained in the SCI, SSCI and A&HCI (Leydesdorff and Sun 2009). In this paper, we address the following research questions: What are the quantitative features and characteristics of the global agricultural innovations that are pertinent to IT? How often do researchers working in different global institutions collaborate? What pattern do these co-authors relationships reveal? Based on the results, specific characteristics of the global agricultural research system that are pertinent to IT will be further discussed.

2 Theoretical framework and literature review

The foundation of our theoretical framework is built on two elements: the scientometrics approach and the TH model. The scientometrics analysis provides knowledge and insights about the quantitative features of global agricultural innovation structures related to IT. Through the use of a TH model, which studies the relationship between academia, business, and government, the knowledge linkage structure of any R&D system can be examined (Etzkowitz 2008). Extensive literature shows that the TH model including these three independent UIG actors has the capability to capture both the dynamics within the three helices and new developments at the network level by demonstrating mutual information exchanges between the helices (Etzkowitz and Brisolla 1999; Leydesdorff 2003; Leydesdorff et al. 2006; Leydesdorff and Fritsch 2006; Park et al. 2005; Shapiro 2007).

Fig. 1 Institutional and functional differentiation in the knowledge-based age (Adapted from Leydesdorff and Scharnhorst 2003)

<table>
<thead>
<tr>
<th>University</th>
<th>Government</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science, Technology &amp; Arts &amp; Humanities</td>
<td>Science Citation Index (SCI-Extended)/Social Science Citation Index (SSCI)/Arts and Humanities Citation Index (A&amp;HCI)</td>
<td>Innovation in all fields</td>
</tr>
<tr>
<td>Quantitative features and characteristics provided by the Web of Science</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The literature also notes that this methodology can be used to analyze innovation systems other than nationally defined ones, such as sectorial, technological, and regional innovation systems (Carlsson 2006). In this study, we map the global knowledge linkage structures of the agricultural innovation systems that are related to IT (Fig. 1).

Agricultural innovation systems have changed with the rapidly changing global food and agriculture system. IT has played a crucial role in these changes by contributing agricultural innovations through the improvement of economic growth. Agricultural IT innovations are important mechanisms by which agriculture adapts to future climate change (Parry 1990). A more systems-oriented understanding of how innovation occurs in an agricultural economy is critical to promote dynamism, responsiveness, and competitiveness in agriculture and to enhance productivity (Spielman and Kelemework 2009). Nevertheless, policymakers, investors, donors, and practitioners are unable to promote policies and
investments that foster innovation in agriculture without adequate measurements of the properties and performance of agricultural innovation systems. In these innovation systems, universities are the primary producers of knowledge, industries are the proxy for the economic factors, and the relevant level of government is responsible for interfacing with and organizing both university and industry functions at the systems level (Leydesdorff 2006).

The knowledge linkage structures within the agricultural sector differ across countries, and agricultural IT innovations have particularly changed with the advent of the Internet. For example, in the US, the proportion of farmers who had access to the Internet increased from 13% in 1997 to 29% in 1999. In the case of Australia, 58% of farms had computer access and 34% had Internet access by June 2000 (Rolfea et al. 2003). In Korea, the rate of Internet use in rural areas has exponentially increased since 1998. For example, only 0.6% of those who worked in the agriculture industry used the Internet; this percentage increased to 29.4% as of 2006 and continues to increase. Rural broadband accessibility in Korea was at 81% in 2002 and increased to 100% in 2007 (Ministry of Food, Agriculture, Forestry and Fisheries, Korea).

3 Method and materials

3.1 Measuring TH dynamics

We used the entropy statistics introduced by Boltzmann and Shannon’s information theory (Shannon 1948) for the measurement of TH dynamics. Entropy is used to measure the uncertainty or disorder of a given set of elements (Grupp 1990; Saviotti 1988). When variation is considered as a relative frequency or a probability distribution ($\sum P_i$), the Shannon-type information or the uncertainty contained in the distribution ($H$) is defined (Shannon 1948; Shannon and Weaver 1949) as follows:

$$H_i = -\sum P_i \log_2 (P_i). \quad (1)$$

Equivalently, for a two-dimensional distribution, $H_{ij}$ is represented by the following:

$$H_{ij} = -\sum \sum P_{ij} \log_2 (P_{ij}). \quad (2)$$

Fig. 2 A TH configuration with negative and positive overlap between the three subsystems (Adapted from Park and Leydesdorff, 2010)

![Diagram](image.png)

This uncertainty is the sum of the uncertainty in the two dimensions, reduced by their mutual information; that is, the two variations overlap in their covariation and condition each other in the
remaining variations. Using information theory, mutual information can be written using the T of transmission between two distributions as follows:

\[ H_{ij} = H_i + H_j - T_{ij} \]  (3)

\[ T_{ij} = H_i + H_j - H_{ij}. \]  (4)

If the two distributions are completely independent (i.e., the covariation is zero), then \( T_{ij} \) is zero; otherwise, \( T_{ij} \) is positive (Theil 1972, p. 59f). According to the Abramson (1963, p. 129) derivation (Abramson 1963), the mutual information in three dimensions, using "u" for university, "i" for industry, and "g" for government, can be defined analogously as follows:

\[ T_{uig} = H_u + H_i + H_g - H_{ui} - H_{ug} - H_{ig} + H_{uig}. \]  (5)

Depending on the relative sizes of the contributing terms, the resulting indicators can be negative or positive (or zero), where a negative value indicates that uncertainty is reduced at the network level. McGill (McGill 1954) named this possible reduction of uncertainty “configurational information” (Jakulin and Bratko 2004; Yeung 2008). The reduction cannot be attributed to one of the contributors; hence, it cannot provide systemic network effects. The bilateral terms contribute to the reduction of uncertainty, while uncertainty in three dimensions adds to the uncertainty prevailing at the network level.

In Fig. 2, in the left-hand configuration, the system is not trilaterally centralized, whereas the common overlap between the three spheres in the right-hand picture supports a robust integration of both bilateral and trilateral relations. Central integration of UIG relations is considered desirable from a policy perspective to establish a competitive advantage at the national or regional level (Etzkowitz and Leydesdorff 2000; Mirowski and Sent 2007). This view of information in three dimensions allows us to measure the balance between the dynamics of integration and differentiation at the systems level by examining the relative frequencies of relationships between the partially overlapping sets. Overall, mutual information can be considered an information-theoretic analog of covariation, which reduces the uncertainty on both sides. Unlike covariance analysis between three or more variates, information-theoretic measures are dimensionless and allow for comparisons between (quasi) experimental results that differ in their metrics (Garner and McGill 1956, p. 228).

### 3.2 Data collection

We collected data from the Web of Science (http://apps.isiknowledge.com), provided by the ISI of Thomson-Reuters. All papers were from the SCI, SSCI and A&HCI (1990-2011), and at least one IT-related study from the agriculture sector was collected from each database. This study focuses on global trends within publications in IT-related agricultural sectors. Therefore, the data consist of global bibliometric information on scientific papers with topics related to IT and agriculture.

To limit the search to papers relevant to IT and the agricultural sector, we entered the following search query: TS=(Agri* or Agro* not Agrou*) AND TS=(Information and Technology* or Information Communication Technology or ICT or Communication and technology* or Information System* or Computer Science or Computer and System* or Telecommunication* or Telematics or Informatics or computing or software and system* or GIS or Information Management Systems or Management Information System or MIS or Broadband Connectivity or Internet* or data mining or computer networks or mobile phone or smart phone or smart and technology* or Informatization or Ubiquitous or Supply chain Management or Farm to table or Food traceability system or e-business or e-commerce or e-government). TS is the query for topics. We picked “agri* or agro* not agrou*” as the search words because agri* comprises all words beginning with “agri-” (e.g., agriculture, agricultural, etc.), agro* comprises all words beginning with “agri-” (e.g., agriculture, agricultural, etc.), agro* comprises all words beginning with “agro-” (e.g., agrochemical, agronomics, etc.), and agrou* comprises words beginning with “agrou-” (e.g., aground), which are not related to agriculture.

To identify all research pertinent to IT, the above-mentioned keywords for IT were used. We only collected data starting from 1990, although Web of Science provides bibliometric information on papers published from 1898. We limited the date range because there were few papers published before 1990 with topics related to IT and agriculture.
4 Data analysis

4.1 Descriptive statistics: Scientometrics approach

The scientometrics approach was first used for the data analysis to provide knowledge and insight on global agricultural innovation, including its quantitative features and characteristics. A total of 8,672 bibliometric data were collected from the period 1990-2011. A total of 8,059 papers in SCI, 1,371 papers in SSCI, and 75 papers in A&HCI that were pertinent to IT and the agricultural sectors were published in the ISI databases (Accessed on July 25, 2011).

Globally, yearly publications in the SCI, SSCI and A&HCI are increasing, with a sharp increase since 1996, as shown in Fig. 3.

**Fig. 3** Number of globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and pertinent to IT by publication year

![Graph showing yearly publications](image)

Fig. 4 illustrates the papers published globally by category; 93.22% were articles, and the remaining papers were published in other document formats, including notes, reviews, and editorial materials.

**Fig. 4** Globally published papers in the SCI, SSCI and A&HCI relevant to the agricultural sector and IT by document type

![Graph showing document types](image)
Fig. 5 shows the papers published globally by language: 95.95% were in English, and the remaining papers were published in other languages, such as German, Portuguese, and Spanish. Research publications in the subject area of agriculture were the largest share (36.19%), followed by environmental sciences ecology (28.16%) and others, as shown in Fig. 6.

**Fig. 5** Globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and IT by language

![Language Distribution](image1)

**Fig. 6** Globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and IT by subject area (first 50)

![Subject Area Distribution](image2)
Approximately 40% of the research articles by global scientists came from just 50 institutions. The United States Department of Agriculture, Agriculture Research Service (USDA ARS) published the highest number of papers (3.55%) of all the institutions, as shown in Fig. 7. Of the records identified, 7,293 records (84.01%) did not contain data on the funding agency. The remaining articles were funded by the European Union (0.98%), the National Natural Science Foundation of China (0.67%), and the National Science Foundation of the US (0.60%), as shown in Fig. 8. Regarding the publication of globally published papers in the SCI, SSCI and A&HCI on IT in the agricultural sector by country, the U.S. published the highest numbers of papers (34.28%), whereas Korea published only 1.04%, as shown in Fig. 9.

Table 1 provides the digital economy rankings, which assess the quality of a country’s IT infrastructure and the ability of its consumers, businesses and governments to use IT to their benefit. When we compare the published papers and digital economy rankings by country, we find dissimilar results regarding the ranking of the countries. This variation may be because the countries with the largest digital economies focused on industrialization rather than on digital agriculture.
Fig. 8 Globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and IT by funding agency (first 50)

![Chart showing number of papers by funding agency]

Fig. 9 Globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and IT by country (first 50)

![Chart showing number of papers by country]

Table 1 Digital economy rankings and scores, 2010

<table>
<thead>
<tr>
<th>Country</th>
<th>Ranking</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norway</td>
<td>1</td>
<td>1000</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
<td>950</td>
</tr>
<tr>
<td>Finland</td>
<td>3</td>
<td>900</td>
</tr>
<tr>
<td>Australia</td>
<td>4</td>
<td>850</td>
</tr>
<tr>
<td>Germany</td>
<td>5</td>
<td>800</td>
</tr>
<tr>
<td>USA</td>
<td>6</td>
<td>750</td>
</tr>
<tr>
<td>France</td>
<td>7</td>
<td>700</td>
</tr>
<tr>
<td>Japan</td>
<td>8</td>
<td>650</td>
</tr>
<tr>
<td>China</td>
<td>9</td>
<td>600</td>
</tr>
<tr>
<td>India</td>
<td>10</td>
<td>550</td>
</tr>
</tbody>
</table>

... (continues)
<table>
<thead>
<tr>
<th>2010 rank</th>
<th>Country</th>
<th>2010 score (of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sweden</td>
<td>8.49</td>
</tr>
<tr>
<td>2</td>
<td>Denmark</td>
<td>8.41</td>
</tr>
<tr>
<td>3</td>
<td>United States</td>
<td>8.41</td>
</tr>
<tr>
<td>4</td>
<td>Finland</td>
<td>8.36</td>
</tr>
<tr>
<td>5</td>
<td>Netherlands</td>
<td>8.36</td>
</tr>
<tr>
<td>6</td>
<td>Norway</td>
<td>8.24</td>
</tr>
<tr>
<td>7</td>
<td>Hong Kong</td>
<td>8.22</td>
</tr>
<tr>
<td>8</td>
<td>Singapore</td>
<td>8.22</td>
</tr>
<tr>
<td>9</td>
<td>Australia</td>
<td>8.21</td>
</tr>
<tr>
<td>10</td>
<td>New Zealand</td>
<td>8.07</td>
</tr>
<tr>
<td>11</td>
<td>Canada</td>
<td>8.05</td>
</tr>
<tr>
<td>12</td>
<td>Taiwan</td>
<td>7.99</td>
</tr>
<tr>
<td>13</td>
<td>South Korea</td>
<td>7.94</td>
</tr>
<tr>
<td>14</td>
<td>United Kingdom</td>
<td>7.89</td>
</tr>
<tr>
<td>15</td>
<td>Austria</td>
<td>7.88</td>
</tr>
<tr>
<td>16</td>
<td>Japan</td>
<td>7.85</td>
</tr>
<tr>
<td>17</td>
<td>Ireland</td>
<td>7.82</td>
</tr>
<tr>
<td>18</td>
<td>Germany</td>
<td>7.80</td>
</tr>
<tr>
<td>19</td>
<td>Switzerland</td>
<td>7.72</td>
</tr>
<tr>
<td>20</td>
<td>France</td>
<td>7.67</td>
</tr>
<tr>
<td>21</td>
<td>Belgium</td>
<td>7.52</td>
</tr>
<tr>
<td>22</td>
<td>Bermuda</td>
<td>7.47</td>
</tr>
<tr>
<td>23</td>
<td>Malta</td>
<td>7.32</td>
</tr>
<tr>
<td>24</td>
<td>Spain</td>
<td>7.31</td>
</tr>
<tr>
<td>25</td>
<td>Estonia</td>
<td>7.06</td>
</tr>
</tbody>
</table>

Source: Economist Intelligence Unit, 2010

4.2 Analysis of TH dynamics
The inputs for the TH analysis are the numbers of papers with only university authors, only industry authors, only government authors, only authors who are either from university or industry, and so on. Collaboration patterns are measured by co-author relationships based on the SCI, SSCI and A&HCI data, as shown in Fig. 10. With some variation, collaboration patterns increased noticeably from 1996 to 2010.

**Fig. 10** Globally published papers in the SCI, SSCI and A&HCI on the agricultural sector and IT by source title (Journal) (first 50)

Globally, university addresses are involved in a high percentage (70%) of the published papers in most countries, including OECD countries. The percentage of articles that exclusively included an industrial address was consistently the lowest during the study period. This result was not affected by university-industry relations because this percentage also decreased during the study period. The linear regression in Fig. 11 is drawn to assist the reader with the interpretation. In summary, these statistics indicated that there were decreasing triple-helix relationships in IT-related agricultural sector publications from 1991 to 2000 because of a decline in university-industry relationships, indicated by co-authored publications. However, the number of papers incorporating both industrial and government addresses remained low, both with and without university participation.
Fig. 11 Number of papers published globally in the SCI, SSCI and A&HCI on the agricultural sector and IT with bilateral and trilateral relations between TH sectors within the global economy.

Fig. 12 illustrates IT-related agricultural research publications with mutual information on the bilateral relations between the TH sectors, as measured by co-author relationships. The UG research collaboration (Tug) showed the highest value, with 1.47 mbits of bilateral relations in 1991. While Tug values decreased in 1993 (0.23 mbits), UI collaborations began to blossom this year (0.74 mbits). The transmission value of UI co-author relationships (Tui) increased to 0.96 mbits in 1990. Mutual information between university and industries (Tui) has been stagnant, and there was more active scientific cooperation between the government and industry publications (Tig) than between university and industry publications (Tui) during 1995-2010.
**Fig. 12** Mutual information measured in bilateral relations between TH sectors in the SCI, SSCI and A&HCI publications on the agricultural sector and IT, globally

**Fig. 13** Global publication rates of papers and synergy effects among TH sectors on the basis of co-author relationships in the SCI, SSCI and A&HCI on the agricultural sector and IT
The TH indicator (Tuig) can also be negative due to a synergetic effect of TH relations, whereas the mutual information in the bilateral relations is positive by definition. Fig. 13 provides the longitudinal trend, expressed with two-year moving averages. This trend shows an interesting path between the three institutional domains. The information values of Tuig during 1990-2007 ranged from −92.23 mbits in 1990 to a minimum of −103.41 mbits in 2007. The mutual information between the three TH agencies (Tuig) remained relatively steady during the 1998-2007 period. The TH dynamics of UIG relationships varied considerably. This variation generally accords with changes in government research policies and institutional regulatory frameworks worldwide.

5 Discussion and conclusion

In this knowledge-based society, research fields and industrial structures are internationally organized, raising a number of control problems at the national level. Government intervention can no longer be expected to steer these developments. Global trends in the knowledge linkage structures of innovation within agricultural R&D systems are a complex construct of integrating and differentiating mechanisms. On the basis of UIG subsystems, differentiation is enforced within each of the helices of the TH model. In particular, academics wish to publish, industries wish to gain financially from collaborations, and policy makers wish to represent the public interest. Our results reveal that because of this differentiation, integration in TH relations became less central not only to policy making, but also to the dynamics of the knowledge linkage structures of innovation itself.

This study found that approximately 40% of research articles by global scientists came from just 50 institutions. The United States Department of Agriculture, Agriculture Research Service (USDA ARS) published the highest number of papers (3.55%) among all institutions. In regard to funding agencies, most articles were funded by the European Union (0.98%); the National Natural Science Foundation of China (0.67%) and the National Science Foundation of the U.S. (0.60%) were the second and third, respectively. By country, the U.S. published the largest numbers of papers (34.28%) on this topic, whereas Korea published only 1.04%. When we compare these results with the digital economy rankings, we find dissimilar country rankings for agricultural innovations compared to positions on IT. For example, Korea ranked first in the digital opportunity index (DOI) in 2007 and 13th in the digital economy rankings in 2010. However, Korea ranked 25th in terms of knowledge linkage structures in the agricultural sector related to IT. This variation may be because the highest-ranked digital economy countries focus on industrialization using IT rather than on digital agriculture. We also found that incentives at the institutional level were not equally distributed for all sectors. Consequently, the social sciences and humanities consistently lagged behind the natural and life sciences.

The implications of this study are produced by mapping the dynamics of the R&D knowledge linkage structures of agricultural innovations that are relevant to IT through science, technology, and innovation networks using the TH indicators. Our findings reveal that the global knowledge linkage structures of the agricultural sector and IT fluctuated frequently and failed to increase the research capacity for agricultural innovations because of a neglect of the network effects of the TH dynamics of UIG. From this study lesson, individual countries can compare their network effects in the TH dynamics of UIG with the global trends and can correct imbalances by reforming their agricultural R&D policies and emphasizing IT engagement. Greater collaboration between government institutions, universities and industry clusters can improve knowledge production and diffusion and increase commercial applicability by establishing agricultural innovations that engage with IT. We recommend that countries’ national R&D research policies should be based less on strict quantitative performance measures and more on a balanced approach between bibliometric indices and the informed judgment of peers with expertise and academic maturity. We believe that including the network effects of the TH dynamics of UIG in policy will improve the knowledge linkage structures of agricultural research pertinent to IT by improving the agriculture economics of individual countries, ultimately leading to global development.

References


