Innovations in the evaluation of rural development projects: the experience of the International Fund for Agricultural Development's Independent Office of Evaluation

Innovation in the way evaluations are conducted is gaining importance. A host of factors are responsible for this, including: an emphasis on collecting stronger evidence for accountability purposes, advancements in technology and computational power, and a change in perspective concerning the role of beneficiaries in programme evaluations. This article presents three examples of innovations used by the Independent Office of Evaluation (IOE) of the International Fund for Agricultural Development (IFAD). These include geo-spatial analysis as a means for data triangulation, genetic matching methods in quasi-experimental ex-post impact evaluations, and SenseMaker - a qualitative participatory technique that involves programme beneficiaries. As the results obtained from these three methods have been satisfactory, this article suggests that expanding the use of these successful innovations, in addition to forging partnerships with experts and sharing experiences, are all key ingredients to mainstreaming innovations in evaluations.

Oscar Garcia and Hansdeep Khaira, Independent Office of Evaluation, International Fund for Agricultural Development

The need for innovation

While the evaluation community has relied on tried and tested methods of evaluation, an increasing need is being felt to go beyond the "comfort zone" and push the frontiers of evaluation method - to aim for innovation in the way evaluations are conducted. This need has its genesis in several factors.

First, the measurement and assessment of development outcomes is gaining ever-increasing traction, not in the least because of the need to measure and report progress on the Sustainable Development Goals as well as the donor community's emphasis on understanding how their funds are being put to use, i.e. the accountability proposition. This means that better and stronger evidence is needed to attribute programme effects. Second, there is a felt need for cost-effectiveness in evaluations; this emanates from shrinking or static budgets within evaluation offices who, at the same time, are being asked to undertake evaluations of ever-increasingly complex programmes. Third, a paradigm shift in the way beneficiaries of development programmes are viewed – not merely as participants of programmes but also as stakeholders – is giving rise to the use of participatory methods in the design and implementation of programmes, but also in the measurement of their outcomes.

In recognition of the above factors, the IFAD's IOE has made innovation an important part of its evaluations, especially in the past five or so years. It has attempted to take advantage of the leaps in technological advancements and the increasing computational power of computers to incorporate innovations in methodology. This article shares the learning of IOE from the perspective of three innovative methods that it has used.

Innovations in IOE evaluations

Geo-spatial analysis

Introduction. The wide-scale availability of satellite data has made geo-spatial analysis[2] an exciting proposition to consider for use in evaluations of rural development initiatives. IOE, in collaboration with the Environment and Climate Division of IFAD, used geo-spatial analysis in an evaluation concerning an agriculture support project in Georgia. The project aimed to rehabilitate aging and dilapidated irrigation water canals so that water could reach the farmers for irrigation of their crops. The development goal of this intervention was to increase farmers' incomes and food security through increases in farm production which they could use for increasing their own consumption while selling the excess in markets. The geo-spatial analysis was used on a pilot basis as a complement to a household survey that was carried out to measure project effects, thus strengthening the triangulation of information sources.

Methodology. The methodology consisted of using time-series satellite imagery to compute temporal variations (before and after the intervention) of the Normalised Difference Vegetation Index (NDVI)[3] for comparing the change in vegetation cover between beneficiary farm plots and comparison/control group farm plots, before and after the rehabilitation of irrigation schemes. The rationale being that the intervention would cause

- a change in production area (either expansion of the production area or increase in production within the same area) from before to after the intervention, compared with similar areas not affected by project interventions. The output would be an estimate of the magnitude and significance of the difference in "greenness" change in land cover between the intervention area and control areas.

The method used for analysing the data was the before/after control/impact (BACI) index.[4] The concept is similar to the difference-in-difference method.[5] The analysis was performed on freely available satellite images using Google Earth, a cloud-based open platform: 250-m NASA MODIS NDVI product (8 days) from 2003 (before-intervention year) to 2016 (after-intervention year). The methodology was completely automatized by developing an algorithm in open source statistical software R.

The first step consisted of analysing the time-series dataset (2003 to 2016) and calculating a multi-annual vegetation development profile, allowing a determination of the vegetation growth period and then classifying the area according to different vegetation development patterns. Only the cluster classes present in the area of intervention (similar land cover and vegetation development patterns) were considered eligible for the analysis.

The second step consisted of assessing the similarity between pixels in the project ad control areas. Similarity was defined as the complement of the Root Mean Squared Error (RMSE) between the fractional compositions and one, i.e. similarity s=1–RMSE. Values close to one indicated nearly identical overall composition of a control and a treatment site. Twenty control areas with higher RMSE were considered for the calculation of the BACI contrast. In the last step, the impact of the intervention was calculated as the BACI effect that represented the differential change between project and control areas, when compared before and after the intervention.

The BACI analysis provides two important statistics: the significance level (P-value) of the BACI effect test and the BACI contrast. The BACI contrast is calculated as the difference between project and control subjects, and between the periods of comparison.

BACI contrast = (μCAa – μCAb) – (μPRJa – μPRJb)[6]

"Since evaluators are not expected to be experts in new methods, evaluation offices need to forge partnerships with the private sector, research institutions and universities who may be the originators of, or experts in, innovative methods."

By convention, a negative BACI contrast indicates that the variable has increased more (or decreased less) in the intervention site with respect to controls in the time period ranging from before to after the project implementation. The BACI contrast is expressed in the same units of the variable of interest, here NDVI. In order to highlight the magnitude of the contrast with respect to the initial conditions, it was normalised by the mean of the impact area NDVI before the intervention took place and expressed it as a percentage. This derived variable is referred to as "relative contrast".

Results. The results showed a statistically significant negative BACI contrast (i.e. improvement in NDVI of project areas with respect to control areas after the intervention) in 7 out of 14 samples (four had a significant 0.05 p-value). Focusing on the sites for which a significant BACI effect was detected, the average relative contrast is -1.24 per cent. Considering NDVI as a rough approximation of the fractional vegetation cover, these numbers translate into a relatively small (1.24 per cent) improvement in the vegetation development of beneficiary farms with respect to the control farms. Importantly, these findings were consistent with findings from the household survey which showed a minor increase in farmland used for crop production by beneficiaries.

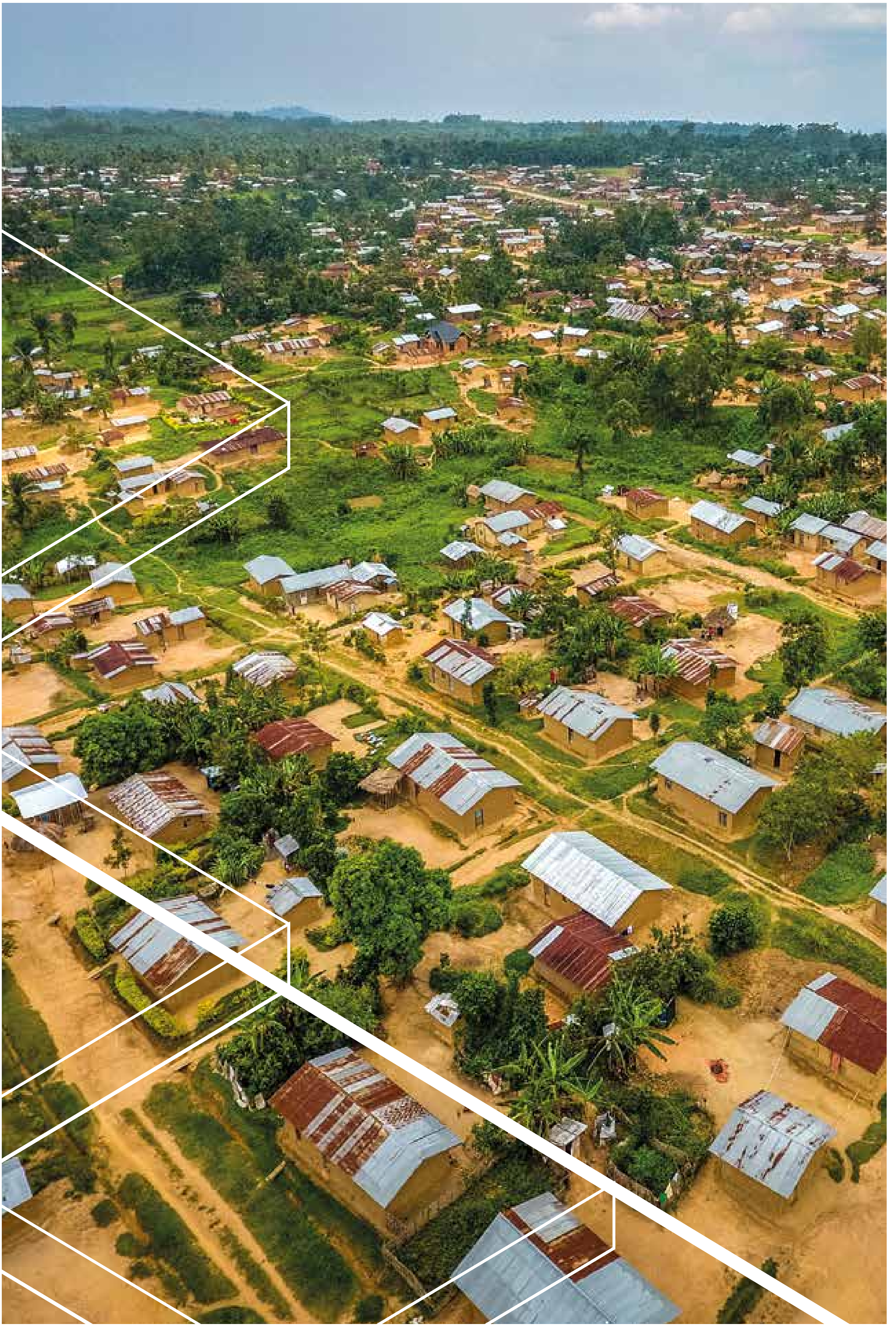
Opportunities and limitations. The pilot study showed that geo-spatial analysis offers the advantage of triangulating survey data with geo-spatial data (or the other way around). In addition, several other advantages derive from it: i) the ability to reach remote, hard-to-access or dangerous areas; ii) easier identification of control groups (identifying villages at the same altitude, with roughly the same number of inhabitants, distance from regional centre, etc.); iii) cost-effectiveness due to increasing availability of open-source data, software and storage space (cloud-based).

There are, however, certain caveats and limitations to be borne in mind: i) the initial set-up cost of acquiring expertise (in-house) can be high, as evaluation offices need to partner with universities, research institutions and the private sector; ii) it cannot be used for all types of interventions and there are limitations; iii) some field work in the form of obtaining exact geographic coordinates is required.

Genetic matching

Introduction. Impact evaluations that make use of econometric techniques to better attribute project effects are an area that evaluators have been tapping into. The most commonly employed technique is Propensity Score Matching (PSM), particularly when there has been no random assignment of beneficiaries, i.e. programs have relied on self-selection, which is often the case in development projects. A propensity score is the probability that a unit (household, for example) with certain characteristics will be assigned to the treatment group (as opposed to the control group). The scores can be used to reduce or eliminate selection bias in observational studies by balancing covariates (the characteristics of participants) between treatment and control groups. When

Figure 1: Satellite image of the project areas (PRJ) and control areas (CA) for one of the irrigation schemes rehabilitated by the project



the covariates are balanced, it becomes much easier to match participants with multiple characteristics.

Although commonly used, PSM is not without its criticisms. For instance, there is no consensus on how exactly matching should be done and how to measure the success of the matching procedure. Further, it is argued that the true propensity score can never be known in observational studies, thereby casting doubt that the propensity score estimates are accurate. Although Rosenbaum & Rubin recommended iteratively checking the propensity score for balance, this can be quite challenging. In order to overcome this limitation, and taking advantage of the growth in computational power, IOE used a different method for its impact evaluation in Georgia, called the Genetic Matching method (Genmatch)[7]. This uses an algorithm to maximize the balance of observed covariates across matched treatment and control units and eliminates the need to manually and iteratively check the propensity score. Such a method is possible due to the increasing popularity of computationally intensive simulation and machine learning methods.

Method. The Genmatch method uses a combination of PSM and Mahalanobis[8] distance methods, the two methods used to match treatment and control groups on a set of characteristics. It matches samples on their weighted Mahalanobis distances calculated from the distance matrix that includes propensity scores and other functions of the original covariates. Genmatch adopts an iterative approach of automatically checking and improving covariate balance measured by univariate paired t-tests and/or univariate Kolmogorov-Smirnov (KS) tests. In every iteration, weights used in the distance calculation are adjusted to eliminate significant results from the univariate balance tests from the end of the last iteration. The iterative process ends when all univariate balance tests no longer yield progress in increasing p-values. The aim is to maximise the p-value associated with the covariate which represents the greatest difference between the two samples.

In the case of IOE's impact evaluation, the Genmatch was used for sampling (matching treatment and control clusters, or villages, on covariates). A Genmatch algorithm was used to calculate weights for each covariate and a matching algorithm was then used to identify the most similar communities to the treated communities at the time of the project baseline, prior to treatment.

Results. Although genetic matching generally outperforms PSM, to test whether it did in the case of the impact evaluation, match balance was tested for the samples matched on propensity scores. The results demonstrated that Genmatch led to greater balance on covariates as compared to PSM.

Opportunities and caveats. The main advantage of Genmatch is that it directly optimizes covariate balance. This avoids the manual process of checking covariate balance in the matched samples and then re-specifying the propensity score accordingly. By using an automated process to search data for the best matches, Genmatch is able to obtain better levels of balance without requiring the analyst to correctly specify the propensity score. It makes use of the current advances in computational power. Open source software that implements GenMatch and a variety of other matching algorithms is available for the R programming environment.

The advantage of any new matching method is limited because of the selection on observable assumptions. The plausibility of the assumptions must be carefully scrutinized in each application using evidence beyond the statistical method. In observational studies, key identifying assumptions cannot be tested by simulations nor proven mathematically. Therefore, more validation studies based on real data are needed to improve observational methods in practice and to clarify the conditions in which these methods are appropriate.[9]

SenseMaker

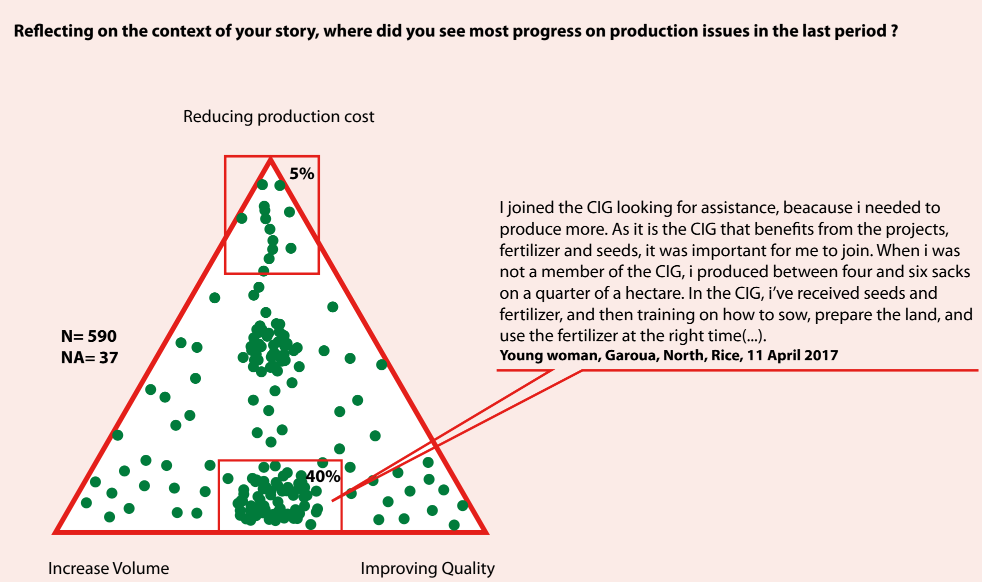
Introduction. As part of the Country Strategy and Programme Evaluation (CSPE) conducted in the Republic of Cameroon, the IOE evaluation team[10] piloted an innovative approach to confirm the linkages between the support provided by projects and changes in living conditions as perceived by beneficiary households. The study targeted two projects financed by IFAD focused on cassava, rice and onion value chains. It sought to fill the evidence gap regarding the contribution made by project outputs to the changes measured in IFAD impact domains, such as agricultural productivity, incomes and food security.

The approach was based on a participatory methodology called SenseMaker that involves the collection of a large number of brief stories from beneficiaries, recounting one or more notable changes perceived as a result of their participation in producer organizations supported by an IFAD project. The short stories were then analysed by the respondents themselves through a separate interpretation questionnaire. This lent the analysis greater legitimacy by reducing the bias associated with an external expert's interpretation of the data.

Method. SenseMaker is based on the collection and indexing of micro-narratives. These anecdotes, experiences or stories are self-signified by the storytellers. This means that respondents assign meaning to their own stories (self-interpretation) immediately after they have shared their anecdotes, experiences or stories through a set of questions (signifiers) rather than an external intermediary interpreting the narratives (common in qualitative approaches). SenseMaker implies the collection of a large number of stories (300+ to thousands) to gain multiple perspectives on the domain of interest. The signification (indexing) of the fragments allows for quantitative pattern analysis backed with explanatory narratives.

Results. In total, 590 stories were collected and self-interpreted from twenty Producer Organizations (PO). The

Figure 2: Respondents' perception in relation to the production of their crop



opening question to beneficiaries was: Since you have become a member of the common interest group (CIG), can you tell us about an important positive or negative change related to the production, processing or selling/marketing of your crop (onion, rice or cassava) and how this has affected you and your family? Then, an analysis of the responses using dedicated SenseMaker software made it possible to uncover trends embedded in the stories, by positioning the large number of stories on specially designed charts (using visual combinations of multiple choice questions, triads, dyads and stones). The software analysis brought up additional questions which were further explored during four participatory workshops with beneficiaries who took part in the survey.

The beneficiaries interviewed were generally satisfied with services provided by their PO, which were mainly focused on training, processing and storage of produce. At the same time, sustainability issues emerged from the stories, mainly linked to internal tensions in the POs, inadequacy of processing equipment and storage facilities, and limited availability of seeds and fertilizers after project completion.

Opportunities and limitations. The experience of IOE using a participatory tool involving respondents such as the SenseMaker was positive. The method allowed for quick collection and analysis of qualitative data (cost and time efficient), it provided evidence-based "hard" and "soft" data, and it fit in the evaluation process and adapted to a mixed-methods approach. However, the use of this method also brought forth certain limitations, such as the unavoidability of researcher biases (framework design, theory of change, sampling), the need for technical support at first use and a requirement for the close supervision of data collection.

Conclusions

This article has provided some successful examples of innovations used by IOE in its evaluations. From this exercise, we can conclude that once an innovation is successful, it should be made to "stick", and its use should be expanded to other evaluations when relevant. The geo-spatial analysis and the SenseMaker have been successfully used by IOE in evaluations of other country programmes as well. A second takeaway is that, since evaluators are not expected to be experts in new methods, evaluation offices need to forge partnerships with the private sector, research institutions and universities who may be the originators of, or experts in, innovative methods. Finally, is it important for the evaluation community to share its experiences derived from the use of innovative methods to foster collective learning, as this article set out to do. eVALU

References

Carbon, M. and Ahmedou, H. (2017). Contribution study of two ''value chain'' projects. Unpublished manuscript, Country Strategy and Programme Evaluation, Republic of Cameroon.

Diamond, A. and Sekhon, J. S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies, The Review of Economics and Statistics 2013 Vol. 95, 932-945.

Ho, Daniel E., Kosuke I., Gary K., and Elizabeth A. Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis 15, no. 03 (2007): 199-236.

IFAD Independent Office of Evaluation (2014). Impact Evaluation of the Agricultural Support Project in Georgia. International Fund for Agricultural Development, Rome.

IFAD Independent Office of Evaluation (2018). Country Strategy and Programme Evaluation in Cameroon. International Fund for Agricultural Development, Rome.

Meroni et al. (2017). Remote sensing monitoring of land restoration interventions in semi-arid environments with a before–after control-impact statistical design, International Journal of Applied Earth Observation and Geoinformation.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. Biometricka. Apr. 1.,

Endnotes

1. Geospatial analysis is the gathering, display, and analysis of imagery, GPS, metadata, satellite photography, remote sensing, and historical data of particular areas to identify impact. For example, orbiting satellites can probe the build-up of infrastructure (e.g. number of tents in a refugee camp, roads, irrigation canals), or assess the different wavelengths emitted from the Earth's surface to indicate changes in vegetation, drought (for instance to appraise the risk of famine), or flooding.

2. Normalized Difference Vegetation Index (NDVI) is an index of plant "greenness" or photosynthetic activity.

3. Presented in the research paper: Remote sensing monitoring of land restoration interventions in semi-arid environments with a before–after control-impact statistical design, Meroni et al. 2017.

4. The project effect or impact is calculated as the difference in the value of the variable of interest before intervention and after intervention for both treatment and control or non-treatment groups. The resulting values for both the groups are then subtracted from each other to give the final outcome or effect size.

5. where μ is the site-specific spatial mean of the variable selected to represent the impact (here NDVI); CAa, PRJa stand for control area and project area after project intervention, respectively; CAb and PRJb stand for control area and project area before intervention, respectively.

6. Presented by Alexis Diamond and Jasjeet S. Sekhon in The Review of Economics and Statistics, 2013. Mr Diamond provided methodological guidance to this IOE impact evaluation.

7. The Mahalanobis distance is a measure of the distance between a point P and a distribution D. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. The Mahalanobis distance between two objects is defined as:
d (Mahalanobis) = [(xB – xA)T * C -1 * (xB – xA)]0.5
Where: xA and xB is a pair of objects, and C is the sample covariance matrix.

8. Ibid.

9. The CSPE team comprised Mr. Michael Carbon, IOE Senior Evaluation Officer, and lead evaluator, and Hamdi Ahmedou, IOE consultant.

Authors' profile

Oscar Garcia is the Director of the Independent Office of Evaluation at the International Fund for Agricultural Development (IFAD) in Rome. He is responsible for evaluating the performance of IFAD, with an active portfolio of 230 rural development projects amounting to US$ 6.9 billion with operations in 99 countries. Before joining IFAD, he served as head of the advisory services for Green Economy at the United Nations Environment Programme –UNEP. Prior to that, Mr. Garcia was senior evaluation advisor at UNDP Evaluation Office in New York, overseeing programmatic and thematic evaluations in Africa, Asia and Latin America. He was also Director General for Trade Policies at the Bolivian Ministry of Economic Development. He holds a Master degree in organizational change management, New School University, New York, an MBA from the Bolivian Catholic University in association with the Harvard Institute for International Development and a BSc in Economics, University of Santa Catarina, Brazil.

Hansdeep Khaira is an Evaluation Officer in the Independent Office of Evaluation (IOE) of the International Fund for Agricultural Development (IFAD). Since 2016 he has worked on a variety of evaluations including country, project and impact evaluations. Previously, he worked in IFAD's Programme Management Department on monitoring, evaluation and performance analysis of development projects. Before joining IFAD, Mr. Khaira worked with the Food and Agriculture Organization in policy research and analysis, statistics and training related to agriculture and food security. His work experience covers over 35 developing countries in sub-Saharan Africa, Near East and North Africa, Asia, Latin America and the Caribbean. He holds a Master degree in Agricultural Economics from the University of London's School of Oriental and African Studies (SOAS) and a Master degree in Business Administration from Mumbai University, India.