Development Engineering 7 (2022) 100090

Contents lists available at ScienceDirect

Development Engineering

journal homepage: www.elsevier.com/locate/deveng

Data-powered positive deviance: Combining traditional and non-traditional data to identify and characterise development-related outperformers

Basma Albanna [a,*], Richard Heeks [a], Andreas Pawelke [b,1], Jeremy Boy [c], Julia Handl [d], Andreas Gluecker [b]

[a] *Centre for Digital Development, Global Development Institute, University of Manchester, UK*
[b] *Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) Data Lab, Germany*
[c] *United Nations Development Programme (UNDP) Accelerator Labs Network, United States*
[d] *Alliance Manchester Business School, University of Manchester, UK*

ARTICLE INFO

Keywords:
International development
Positive deviance
Non-traditional data
Mixed methods
Data science
High performers
Global south

ABSTRACT

The positive deviance approach in international development scales practices and strategies of positively-deviant individuals and groups: those who are able to achieve significantly better development outcomes than their peers despite having similar resources and challenges. This approach relies mainly on traditional data sources (e.g. surveys and interviews) for identifying those positive deviants and for discovering their successful solutions. The growing availability of non-traditional digital data (e.g. from remote sensing and mobile phones) relating to individuals, communities and spaces enables data innovation opportunities for positive deviance. Such datasets can identify deviance at geographic and temporal scales that were not possible before. But guidance is needed on how this new data can be employed in the positive deviance approach, and how it can be combined with more traditional data to gain deeper, more meaningful, and context-aware insights.

This paper presents such guidance through a data-powered method that combines both traditional and non-traditional data to identify and understand positive deviance in new ways and domains. This method has been developed iteratively through six development projects covering five different domains – sustainable cattle ranching, agricultural productivity, rangeland management, research performance, crime control – with global and local development partners in six countries. The projects combine different types of non-traditional data with official statistics, administrative data and interviews. Here, we describe a structured method for data-powered positive deviance developed from the experience of these projects, and we reflect on lessons learned. We hope to encourage and guide greater use of this new method; enabling development practitioners to make more effective use of the non-traditional digital datasets that are increasingly available.

1. Introduction

Positive deviance (PD) is based on the observation that in every community or organisation, a few individuals or groups develop uncommon practices or behaviours to produce better solutions to problems than their peers who face the same challenges and barriers (Pascale et al. 2010). Those individuals are referred to as positive deviants and adopting their solutions is referred to as the PD approach. This is an approach that, particularly since the turn of the century, has found a growing niche within development research and practice. However, there are challenges that have constrained the spread of the PD approach; some of which are data-related. Recognising this, it has been proposed that recent developments in the increasing availability of non-traditional digital data provides an opportunity to identify and understand positive deviants in new ways; potentially helping address some of these challenges (Albanna and Heeks 2019). We refer to the use of such non-traditional data to replace or complement traditional data as the "data-powered positive deviance" (DPPD) method. 'Non-traditional data' in this context broadly refers to data that is digitally captured (e.g. mobile phone records and financial data), mediated (e.g. social media and online data) or observed (e.g. satellite imagery). 'Traditional data' refers to data captured manually such as official statistics, observation

* Corresponding author.
E-mail address: basma.albanna@manchester.ac.uk (B. Albanna).
[1] Consultant

https://doi.org/10.1016/j.deveng.2021.100090
Received 7 September 2021; Received in revised form 25 November 2021; Accepted 15 December 2021
Available online 24 December 2021

data, surveys and interviews.

This paper provides an exposition of the DPPD method by describing a methodological framework that guides the combined use of traditional data sources and non-traditional digital data sources to identify and characterise positive deviants in development-related challenges. The framework was first outlined by Albanna and Heeks (2019) and then further developed through its application in a global initiative collaboratively conducted by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) Data Lab, United Nations (UN) Global Pulse Lab Jakarta, the United Nations Development Programme (UNDP) Accelerator Labs Network and the University of Manchester (Data-powered Positive Deviance 2020). This initiative refined the DPPD method by applying it to five distinct domains, spanning six developing countries to identify and understand: farmers achieving higher than usual cereal crop productivity in Niger and Indonesia; cattle farmers in Ecuador who are deforesting below average rates; research output outperformance among Egyptian researchers; public spaces in Mexico City where women are safer; and communities in Somalia which are able to preserve their rangelands despite frequent droughts. The framework presented here should provide a tool for development professionals to identify outperformance in different development sectors by mixing analytical insights from traditional and non-traditional data. Such insights should help amplify innovative, locally-sourced and evidence-informed solutions to development challenges.

In what follows, we first present the history and challenges of positive deviance and the potential for non-traditional data and data science to address those challenges. Following that, we explain how we developed the DPPD method. We then present the three core stages of the method: assessing problem-method fit, determining positive deviants, and discovering the underlying factors leading to positive deviance. We summarise preliminary results from applying these stages in the pilot projects, then discuss lessons learned from applying DPPD, and end with conclusions, including thoughts on future application of the method.

2. Background

Positive deviance was used for the first time in 1976 to inform the design of food supplementation programmes in Central America by identifying dietary practices developed by mothers in low-income families having well-nourished children (Wishik and Van Der Vynckt 1976). The full method and results of this study were not published, limiting uptake. However, in the 1990s the PD approach became more widely recognised as a credible strategy for operational and academic research in nutrition, based on extensive observations and a strong emphasis on impact (Zeitlin 1991; Sternin et al. 1997, 1998). Its first large-scale adoption was by Save the Children Foundation, which used PD as a strategy to reduce malnutrition in Vietnam, rehabilitating an estimated 50,000 malnourished children in 250 communities (Sternin 2002). But it was not until the 2000s that PD started to attract wider attention, when Sternin and colleagues introduced it as a development approach for social change and demonstrated how it can be operationalised as a domain-agnostic approach (Sternin and Choo 2000; Sternin 2002). Since then, PD has been applied across multiple development domains, with public health being the most prominent.

Originally, the PD approach was designed to study the characteristics and practices of individuals who are able to achieve better results in response to a specific development challenge. A more recent set of PD studies has been interested in how certain individuals or groups respond to a development intervention programme significantly better than their peers who are targeted by the same intervention (Post and Geldmann 2018). This is similar to randomised control trials in the sense that it compares post-intervention performance with pre-intervention performance. But in PD studies, the interest is not in the difference in performance between the control and intervention groups as much as in the variation in the performance of units within the intervention group, and potential factors that led to this variation. Identifying the reasons behind the exceptional response of the positive deviants can be used to inform intervention strategies and to increase overall adoption by "bad responders".

Notwithstanding the growth in prevalence of positive deviance as an approach to international development, its adoption has been constrained by a number of challenges (Albanna and Heeks 2019). Given these challenges, there are obvious opportunities for innovation in PD and our particular interest here is in the innovative opportunities offered by non-traditional, digital data sources like big data following the increasing "datafication" of development and growing availability of big datasets in a variety of development sectors (Hilbert 2016). The opportunities have been identified via a systematic literature review of positive deviance and big data in development (Albanna and Heeks 2019):

- **Time and cost of data collection:** traditional PD studies rely mainly on primary data collection to identify positive deviants and to understand their underlying practices and strategies; something that involves significant time, cost and risk. These could be ameliorated by use instead of existing big datasets if they contain indicators of relevance to positively-deviant performance.
- **Positive deviant identification:** because of the costs of data collection, and notwithstanding the substantial impacts of some PD projects, the overall population sample in traditional PD studies tends to be relatively small.[2] Given they are the exceptions in any population, this makes the number of positive deviants in these samples very small, constraining generalisability of conclusions about their particular features and practices. Big data, by contrast, may cover large populations, making it possible to identify a larger number of positive deviants and thus to improve the generalisability of conclusions for practice. Additionally, data sources in traditional PD studies provide a static, cross-sectional reflection of performance whereas some big data could provide a dynamic picture due to its longitudinal coverage. Finally, traditional PD studies have tended to focus on individuals or individual households as positive deviants because they are most amenable to field survey methods. Big data might offer opportunities via direct coverage or aggregation to identify positively-deviant communities or even regions.
- **Monitoring and evaluation:** because of time, cost, logistical and other challenges, traditional PD studies rarely evaluate the impact of any interventions developed as the result of positive deviant identification and analysis. If a big dataset longitudinally captures relevant performance indicators, then it could relatively easily be used for monitoring and evaluation of the effects of scaling positively-deviant practices into an intervention population.
- **Expanding the scope of PD:** despite the spread of positive deviance noted above, there has been a domain and geographic skew in its application. According to Albanna and Heeks (2019), 89% of the sample of PD studies they reviewed were in public health (a form of path dependency due to the success of its first application in nutrition), with 83% targeting rural communities. Big data could help break PD from its current narrow focus, due to the existence of big datasets dealing with a variety of development domains and locations.

However, the role of big data in development has itself been criticised, given that big datasets may often be decontextualised (Taylor and Broeders, 2015). A number of studies have suggested integrating "thick data"[3] with big data to extract meaning and value from it and to rescue it from the potential context loss (Bornakke and Due, 2018; Smets and

[2] The sample size is typically tens or hundreds at most, with the exception of a few studies in health care (Bradley et al., 2009; Wallace and Harville 2012).

[3] Data collected through qualitative and ethnographic methods to uncover individual behaviours and attitudes (Bornakke and Due 2018).

Lievens, 2018; Ang, 2019). Such a combination could be seen as particularly relevant for positive deviance. In order to identify "true" positive deviants[4] that are performing unexpectedly well due to uncommon behaviours and strategies, it is crucial to control for the contextual variables that could influence this performance. Given its decontextualised nature, big data rarely contains such variables and they must therefore be sought in other, traditional data sources. Therefore, combining traditional data with big data is an integral part of the DPPD method presented in this paper.

3. Methodology

The potential value of non-traditional data to the positive deviance approach can only be realised if action researchers and practitioners are provided with a clear method through which to make use of these data. The aim of this paper is therefore to present a systematic method for data-powered positive deviance (DPPD) by testing and validating the use of big data and other types of non-traditional data in positive deviance. In order to do this, we built on the preliminary framework proposed by Albanna and Heeks (2019) which sought to integrate the use of big data into the five main stages of the PD approach (Positive Deviance Initiative 2010):

1) "Define the problem, current perceived causes, challenges and constraints, common practices, and desired outcomes.
2) Determine the presence of positive deviant individuals or groups in the community.
3) Discover uncommon but successful practices and strategies through inquiry and observation.
4) Design and implement interventions to disseminate PD practices and strategies.
5) Monitor and evaluate the resulting project or initiative".

The first version of the DPPD method was applied in a case study of Egyptian researchers who outperformed their peers in terms of research outputs. Following that, we iteratively developed the method through a collaborative initiative between the GIZ Data Lab, UN Global Pulse Lab Jakarta, the UNDP Accelerator Labs Network and the University of Manchester. Action research was chosen as the research strategy because it bridges the gap between research and practice by integrating, rather than chronologically separating, the two processes of research and action (Somekh 1995). It would therefore allow the application of the DPPD method to be fed back into its conceptualisation; that re-conceptualisation then refining practice in an iterative cycle.

In addition to the Egypt case study, the action research cycles were applied in five other PD projects (see Table 1). These were chosen following a call for proposals, to which GIZ field offices and the UNDP Accelerator Lab Network responded. Proposals were selected based on judgement of their viability and the diversity of development domains and countries to which the DPPD method could be applied. As shown in Table 1, the projects also offered diversity in terms of non-traditional data types – citation data, remote sensing data, mapping and cadastral geographic data – and both proprietary and open data sources. This was complemented by a variety of traditional data sources: official statistics, administrative data, surveys and interviews. The units of analysis covered different aggregation levels starting with individuals, farms and communities up to geographical units representing urban areas and villages. This diversity of domains, countries, data and scales was seen as important in helping to broaden the testing base for the DPPD method and to strengthen its likely generalisability.

Within the overall collaborative initiative, a central group was responsible for revising the DPPD method. Its application was led by country-level practitioner teams drawn from domain specialists in GIZ field offices and UNDP Accelerator Labs working in continuous contact with the central group.

[4] Positive deviants who are not false positives that were mistakenly identified as positive deviants because of a contextual advantage that was not accounted/controlled for.

Table 1
Summary of DPPD method projects.

Project	Unit of Analysis	Definition of Positive Deviants	Data Used
Research publication outperformance in Egypt (Albanna et al. 2021)	Individual researcher	Information systems researchers in public universities who achieved significantly higher-than-average scores in one or more of six citation metrics	Citation data from Google Scholar, research publications on Scopus, university websites, interviews and surveys
Rice-farming outperformance in Indonesia (Albanna et al. 2020)	Village	Rice-farming villages that have higher than expected rice productivity as measured by Enhanced Vegetation Index (EVI) scores while controlling for their climatic, socio-economic and demographic conditions	Remote sensing data, official statistics, administrative boundary data and crop masks
Rangeland preservation by pastoral communities in Somalia (Abdullahi et al. 2021)	Community	Communities in the same land capability class that were able to sustain or enhance their rangelands' health (since the 2016 drought) as measured by the Soil-Adjusted Vegetation Index (SAVI)	Remote sensing data, settlement location data, observation data and semi-structured interviews
Cereal-farming outperformance in Niger (Gluecker et al. 2021)	Community	Communities which – despite drought and conflict – achieve high cereal yields as calculated by higher than expected SAVI while controlling for soil, evapotranspiration, precipitation and land use	Remote sensing data, administrative boundary data, land use data, observation data and semi-structured interviews
Public spaces in Mexico City where women are safer (Cervantes et al. 2021)	AGEB[a]	AGEBs where gender-based crimes and crimes with female victims are lower than expected given their population density, demographics, socio-economic status and urban infrastructure	Mexico City open data portal, 911 calls, administrative boundary data, official statistics, observation data and semi-structured interviews
Low deforestation cattle farming in the Ecuadorian Amazon (Grijalva et al., 2021)	Farm	Cattle-raising farms operating in areas of potential forest clearance with deforestation rates that are significantly lower than expected for three consecutive years, while controlling for the size of the farm, the land use, soil adaptability, socio-economic status and cattle density	Remote sensing data, vaccination data, cadastral data, official statistics, land use data, observation data and semi-structured interviews

[a] Área Geoestadística Básica (AGEB), which is the basic geo-statistical area in Mexico City.

The DPPD method that emerged from this process follows the same five stages as the PD approach outlined above, but uses pre-existing non-traditional data sources instead of – or in conjunction with – traditional data sources across the five stages. As detailed in the following section, this requires a series of new and specific methods and practices that are not required in the conventional PD approach. The first stage is also somewhat different because it not only defines the problems but also checks if it is suitable and feasible to use the DPPD method for the proposed project.

4. The data-powered positive deviance method

This section presents the three core stages of the DPPD method. We focus on these three for two reasons: first, because these are the stages so far achieved by the action research projects and, second, because these are the stages that differ most significantly from the conventional PD approach and which therefore most require new guidance. Stage 1 defines the problem and validates if it is suitable and viable to use the DPPD method, hence we refer to it as 'Assessing problem-method fit' instead of 'Defining the problem' (the original name of this stage in the PD approach). Stage 2 seeks to identify positive deviants within the available datasets, and Stage 3 seeks to uncover the factors underlying positive deviant outperformance. Fig. 1 provides a summary of these three core stages of the DPPD method and the different steps conducted in each stage.

4.1. Stage 1: Assessing problem-method fit

In a similar way to the positive deviance approach, the first step of the data-powered positive deviance method is to define the problem and the desired outcome. However, in order to move from the problem to the desired outcome, one has to make sure that using a PD approach is suitable for the problem at hand, to check access to the various data sources and capabilities needed to identify and characterise positive deviants, while ensuring no harm is likely to affect the observed units. So, before applying the DPPD method to the identified problem, it is important to first answer three main questions:

- **Suitability**: Is the positive deviance approach suitable to address this type of development problem?
- **Feasibility**: Is there access to data sources and capabilities that would make it feasible to reach the desired outcome using the DPPD method?
- **Desirability**: Who is likely to benefit from or be harmed by the project, including any potential unintended negative consequences from data analysis?

4.1.1. Defining the problem

When defining the problem, it is important to specify the study population and the unit of analysis. The 'study population' is the group of individuals, communities or geographic units who are suffering from or causing the problem and will be included in the analysis. The 'unit of analysis' is the level at which one can find solutions to the addressed problem. For example, in the Mexico safe public spaces project, the problem we are trying to tackle is the high rates of violence against women in public spaces. The study population is public spaces in Mexico City, our units of analysis are AGEBs and the desired outcome is to reduce violence against women and girls in public spaces (Cervantes et al. 2021). In this step, it is also important to identify the different stakeholders who should be involved (community members and leaders, development professionals, government officials, etc.) in discussions around the current perceived causes of the problem, and to better understand the community's context including challenges and constraints, existing human and natural resources, common practices and normative behaviours. Having the buy-in of the different stakeholders at the very beginning guarantees, to some extent, the adoption and amplification of findings from the PD inquiry later on.

4.1.2. Assessing suitability

There are two key criteria to determine whether a PD approach is suitable: 1) The nature of the development problem being addressed, and 2) The likelihood that positive deviants exist. Neither the conventional PD approach nor DPPD will be suitable if the addressed problem requires mainly a technical solution, e.g. building a road or constructing a dam – in such circumstances, the positive outcome is likely not related to individual practices and strategies. A PD/DPPD approach is much more likely to be applicable if the problem has social components and requires some form of behavioural change or a shift in mindsets, as seen in the project examples discussed here.

Even in this situation, before starting a PD intervention, it is important to check if positive deviants exist. While it may be hard to do this before diving into the data, there are ways to assess whether positive deviants exist or not by engaging with relevant stakeholders that are concerned with the issue at hand. Meeting with key development actors including community leaders and government officials familiar with the targeted sector will help give a sense if outperformers exist. It is also useful to review academic and grey literature to check if there are direct examples of positive deviants or indirect evidence of local solutions sourced from communities living in comparable contexts and facing similar challenges. For example, before starting the Ecuador cattle-farming project, we knew through conversations with a key development actor that certain farmers adopt more sustainable cattle-ranching practices and deforested less than others. Similarly, in Somalia, through interviews with an officer from the Ministry of Environment and Rural Development, we learned about a positively-deviant community that was protecting its trees from cutting and burning for charcoal production. Fig. 2 shows four PD suitability quadrants that can be used to judge whether a PD approach is suitable or not. The projects for which the PD approach is best suited generally lie in the top right quadrant, where positive deviants are likely to exist, and scaling their practices should contribute to solving the problem. Projects in the bottom left quadrant, where positive deviants are unlikely to exist, and scaling practices will likely have a limited impact on the problem, are not suitable for the PD approach.

4.1.3. Assessing Feasibility

The DPPD method relies heavily on existing non-traditional digital datasets that complement more traditional secondary data to identify positive deviants. Given the dependency on existing datasets for the DPPD method to work, a number of conditions regarding data availability, accessibility and adequacy need to be met. In terms of **availability**, it is important to ensure that there are outcome and contextual data, which are crucial for the identification of positive deviants. Outcome data is used to directly or indirectly measure the performance of the target group. It should be capable of identifying individuals, groups or, more generally, units within the target group that outperform their peers. DPPD leverages the potential of readily available non-traditional, digital data, like earth observation data, online or mobile data, citizen-generated data, or sensor data, to capture outcomes.[5] Contextual data is then needed to control for factors that are likely to impact performance but are not related to practices or behaviours. A large portion of those factors should emerge from the perceived causes of the problem identified in the previous step. Contextual data helps put the outcome measure in perspective and guarantees that positive deviants are identified based on performance relative to comparable peers rather than absolute performance. This data can be extracted from traditional data sources (e.g. census) and non-traditional data sources (e.g. remotely sensed climate data). Both the outcome and contextual data

[5] More detail on this is provided in section 4.2.1.

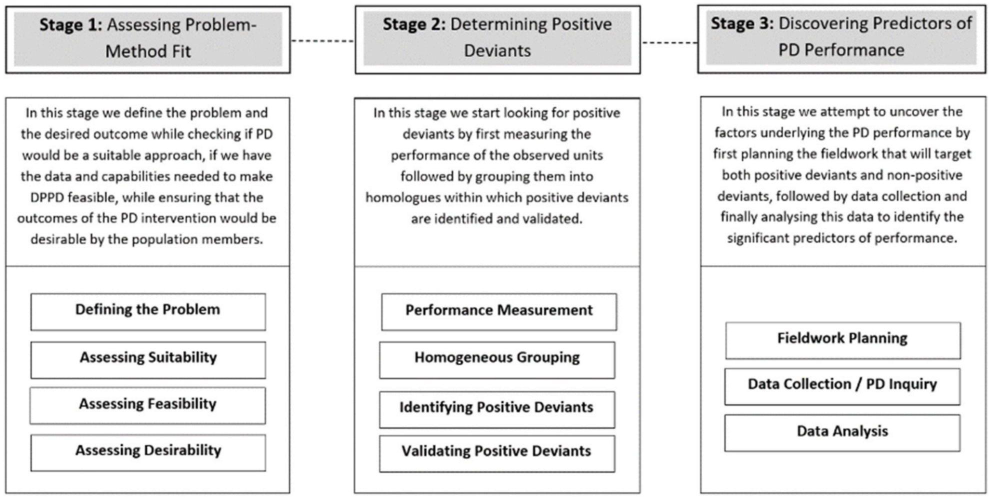

Fig. 1. The first three stages of the DPPD method.

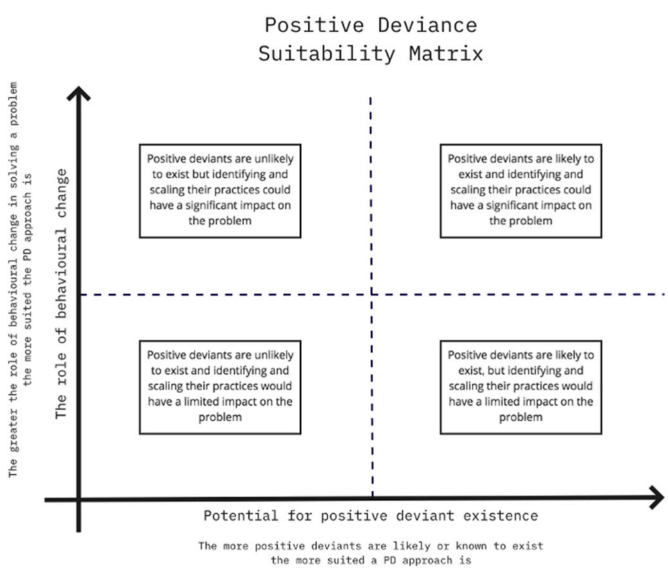

Fig. 2. Positive deviance suitability matrix.

should be spatially and temporally relevant to the problem at hand, i.e. data should be at an aggregation level that is sufficient to capture the outcomes of the unit of interest and it should be recent enough to ensure the relevance of the field investigation.

For instance, in the Niger and Indonesia projects, when measuring the agricultural outcomes of cereal farming villages, we needed remote sensing-based measures of vegetation health. To extract those village-delimited vegetation indices, administrative data about their boundaries was required. And to identify positive deviants within groups sharing similar resources and context, we needed digitally available climate and soil data, administrative data about land cover and agro-ecological zones, and official statistics in order to capture their socio-economic and demographic conditions (Albanna et al., 2020; Gluecker et al., 2021).

After identifying available relevant data, comes the question of data **accessibility**. When there is a need to use non-public data, having data access agreements in place with data providers is a real asset. It can take significant time to negotiate the conditions over access to such data. The Ecuador cattle-farming project grew out from a current project called ProAmazonia, run by UNDP and the Ministries of Environment and Agriculture. Through this partnership it was possible to access cattle vaccination data from the Ministry of Agriculture, training datasets for land cover analysis through the Ministry of Environment, and cadastral data as well as farm boundary data from municipalities through ProAmazonia.

Finally, and most importantly, the available and accessible data should fit the scope of the project, the know-how needed for the analysis should be attainable, and the choices of both data and skills should account for the project's time and budget limitations. This data **adequacy** is usually achieved after several iterations between problem framing and data mapping until a suitable match is found. This process should not compromise the initial purpose of the project. It might however call for starting with a somewhat flexible problem definition (or lens) in well-defined domains with clear development challenges. This flexibility allows navigation through different proxy options to capture the outcomes of the target group. In the Somalia rangelands project, we moved from identifying drought resilient pastoral communities by measuring their livestock numbers remotely, to identifying them based on their ability to sustain the health of their rangelands, which is necessary to maintain livestock (Abdullahi et al. 2021). The data for the former was costly to obtain (very high resolution imagery) and required extensive and complex analytical skills that are rare to find, whereas the data for the latter was readily available and could be more easily analysed because the team already has a remote sensing analyst at hand.

The process of "Assessing Feasibility" presented in this section may appear linear. In practice, though, an accessible data source may turn out to be not available, or an available source may transpire to be not adequate. Hence, the actual process will be iterative until data is found that enables reliable capture of the outcomes and context of the target group.

4.1.4. *Assessing desirability*

This step is about closely looking at all those individuals and communities that stand to benefit or lose from the identification and amplification of the practices, strategies and other factors associated with the outperformance of positive deviants. Should this assessment yield a potential outcome that would harm those three groups – the positive deviants, the non-positive deviants, or the wider community – it may be advisable to adjust the overall design of the project or to abandon the idea. Having said this, some evaluation of net public benefit needs to be undertaken. Some reordering of incentives and benefits may be appropriate, even if it disadvantages one group, if it is to the greater public good.

The PD approach assumes that positive deviants are not aware of their innovativeness and/or impact of their uncommon practices and strategies. But what if they are aware and have deliberately chosen not to share their strategies with other members of the community? For example, they might fear losing their competitive advantage over others, or depleting a resource they alone are aware of, if it were to be shared with other community members, which makes this solution unsustainable. Hence, it is important to assess if it is desirable for a positive deviant to share their practices and behaviours with others. This is likely to be less of an issue where cultural norms dictate against competitive strategies, such as child malnutrition or health. However, it might be more problematic in areas where people more overtly compete with one another, as in the Egypt research performance case study. The following questions can be used to assess desirability: Is it generally safe to assume that it will be desirable to scale the behavioural practice in question? Are we endangering the competitive advantage of positive deviants by sharing their practices and strategies with others? Might we risk harming a positive deviant or a non-positive deviant by revealing their identity? Will inviting others to adopt a PD strategy trigger this strategy's obsolescence? As an example, there was concern about what might happen if we promoted a particular transhumance destination in Somalia as a strategy to help community rangelands recover, given this might lead to overgrazing of rangelands at the promoted destination.

If the DPPD method is seen to be feasible from a data and capabilities point of view, it is important to ensure protection of the privacy of individuals and communities involved: "The availability or perceived publicness of data does not guarantee lack of harm, nor does it mean that data creators consent to researchers using this data" (Zook et al., 2017). Questions to be asked here are: Is the data to be used of sensitive nature, e.g. personally-identifiable information? Are safeguards in place for safe and secure data access and processing? Has consent been given (directly or indirectly) by the data subjects to use this data? To whom can the identity of positive deviants and non-positive deviants be revealed? Given the next stage is the identification of positive deviants, such questions must be thought through at this point.

4.2. *Stage 2: Determining positive Deviants*

After defining the problem and ensuring the applicability of the DPPD method comes the stage of looking for the positive deviants. This section outlines the different steps of this stage, starting with performance measurement, followed by homogeneous grouping and positive deviant identification, and finally the preliminary validation of the potential positive deviants. This is undertaken via data analysis and through reference to existing literature and frameworks that identify relevant measures and variables.

4.2.1. *Performance measurement*

This step attempts to identify the core performance measure for positive deviance; a measure that captures a desirable development outcome as defined by the different stakeholders of the investigated problem. The DPPD method advocates deriving this measure from non-traditional, digital data sources (e.g. big data), and using it either alone or in combination with some other measure. For instance, in the Niger and Indonesia agricultural projects, we used remotely-sensed vegetation indices (e.g. SAVI and EVI) to measure vegetation health – and, hence, agricultural productivity – as the core performance measure of agricultural communities (Albanna et al., 2020; Gluecker et al., 2021).

The data sources that are available are often collected for a different purpose than that of a positive deviance project. In such cases, it is possible that data provides only indirect insights into the subject of interest, rather than direct measures. Hence, the data source measures should be considered as proxies of the actual phenomena that need to be measured, and the validity of using these proxies must be ascertained. This validation could be as simple as checking prior literature showing a strong correlation between the proxy and the desired outcome in a context similar to the one being investigated. If prior studies are lacking, then the proxy relationship should be ground-truthed using direct measures of performance, and their suitability should be validated with local domain experts; triangulating between multiple experts to reduce the dangers of bias within any individual source. In the Niger agricultural project, for example, use of SAVI – rather than other measures – as an indicator of vegetation health and crop productivity was based on local expert advice that SAVI was suitable for semi-arid areas like Niger given it incorporated a soil brightness correction factor (Gluecker et al. 2021).

Depending on how the desired development outcome is defined, the study can have one or multiple performance measures. In the Egypt research publication case study, six research citation metrics were used to evaluate performance because these enabled a balanced consideration of both scientific productivity and impact while controlling for factors like article and author age (Albanna et al. 2021). From each metric, positive deviants were identified and the final set of outperformers included positive deviants from all six metrics. There are also techniques that can be used to summarise multiple performance measures into a single index or at least into fewer measures. A basic approach here assumes all measures to have equal weight. If the performance measures are seen to be of varying importance, a weighted average can be used but this of course requires some collective determination of the weight (importance) of each measure. Summarisation techniques (assuming that measures have equal weight) include principal component analysis, which replaces the original set of measures with a smaller number of uncorrelated measures that account for most of the information in the original set (Abdi and Williams 2010).

4.2.2. *Homogeneous grouping*

Having identified the measure that determines positive deviance, the next step is to divide the study population into homogeneous or peer groups having similar contextual factors, to then identify individuals who deviate positively from their peers. This way positive deviants are identified relative to their context and not in an absolute sense. This grouping also increases the likelihood of identifying positive deviance that can be attributed to particular attributes, practices and strategies that can be transferred; not deviance due to structural factors – contextual variance that impacts the studied outcome but is beyond the control of the unit of analysis – that cannot be transferred. Fig. 3 provides an illustrative example of grouping from the Mexico safe public spaces project where areas were divided into three peer groups based on socio-economic level, daily incoming trips and population density.

The grouping procedure can be done manually based on professional experience and intuition or can be done through unsupervised machine learning techniques such as clustering. The aim of the grouping is to minimise the variance of those structural factors within the groups and maximise it between the groups. There are three main drivers for this grouping:

1. The essence of the PD approach is to uncover context-aware solutions that are associated with the performance of positive deviants. Since it is difficult to capture all the contextual factors driving performance, one aim is to group observations based on aggregated structural variables (e.g. a district poverty index) that would correspondingly reduce variance from underlying disaggregated contextual factors that are not accessible (e.g. household income).
2. A number of studies (Nathan and McMahon 1990; Trivedi et al. 2011, 2015) demonstrate how clustering a population into homogeneous groups can make the per-cluster-prediction better. This is because, rather than seeking to build models that explain the natural variation between clusters, the focus is on within-cluster variation, which increases the model's performance.
3. When a study population is divided into homogeneous groups, findings can be extrapolated with more confidence (Nathan and

[6] This map was provided by Codeando Mexico, the consultants who carried out the data analytics for the Mexico project.

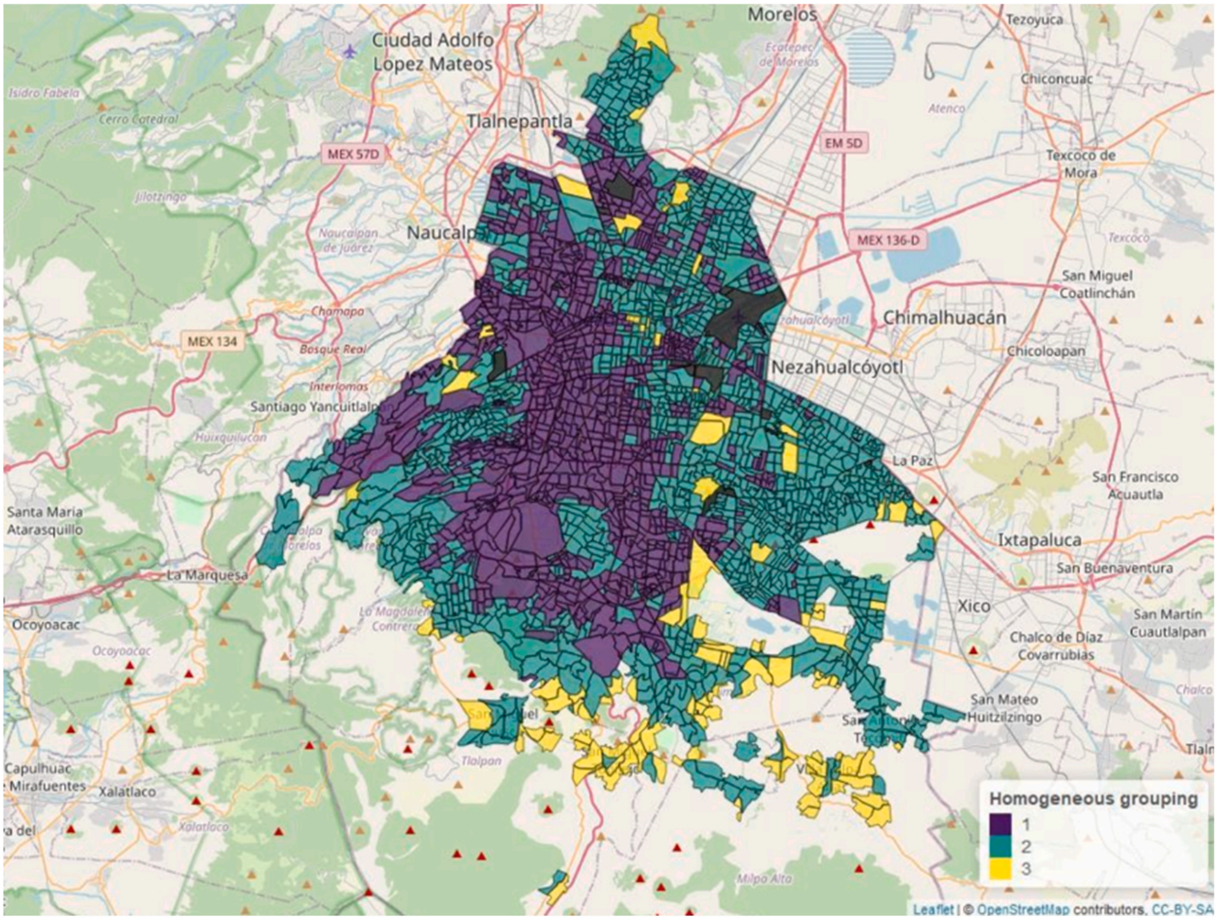

Fig. 3. Homogeneous grouping of areas in Mexico City.[6]

McMahon 1990). This is because more detailed and localised information can be extracted from homogeneous groups having similar conditions (Kovács et al., 2014). This is particularly useful in the following DPPD stages that seek to uncover positive deviants' practices and strategies and to design interventions that disseminate them.

One of the main challenges associated with homogeneous grouping or clustering is the selection of the variables that will be used to assess the degree of similarity between the different observations. Clustering techniques are capable of generating clusters with literally any set of variables, so it is crucial to select variables based on their relevance to the problem. In the DPPD projects, our clustering variables were selected based on theoretical and empirical research indicating that they have a significant impact on the outcome measure. For example, there is a well-established relationship between socio-economic conditions and crime rates (Vilalta and Muggah 2016). Therefore, it was crucial, in the Mexico safe public spaces project, to cluster urban areas into groups with similar socio-economic levels, as shown in Fig. 3. It is also possible to find existing groupings of homologues or clusters that were created for a different purpose but which can be reused for PD identification. As an example, in the Niger agricultural project, we found a recently-released map developed by the Adapt'Action Facility that divided Niger into agroecological zones with similar biophysical, ecological and climatic conditions (Hauswirth et al., 2020). This zoning also included access to data on natural resources, land tenure, farming systems, socio-demographic and economic status. We decided to use these zones instead of creating our own homologues, especially as they took into account valuable local and contextual knowledge which would have been difficult for us to incorporate in our grouping.

4.2.3. *Positive deviant identification*

After dividing the study population into homogeneous groups comes the stage of identifying *outliers* or positive deviants within each group separately. Positive deviants are identified within homologous groups because, as mentioned earlier, it is their relative performance when compared with peers that have similar structural constraints that is important, rather than their absolute performance. This identification requires defining the techniques and cut-off points – the limits beyond which observations are considered positive deviants – which distinguish positive deviants from non-positive deviants. Depending on how performance is measured, there are several ways to identify positive deviants:

- **Univariate Analysis:** this is used in cases where only one variable (i.e. the performance measure) is used to identify outliers. This variable can be categorical i.e. pass/fail, win/lose or healthy/sick. In such cases, positive deviants are those that succeed when most fail. Alternatively, the variable can be continuous and, depending on the underlying distribution, a suitable outlier detection method can be used. For instance, if the data can be assumed to follow a normal distribution, then positive deviants can be defined as observations at the extreme end of the distribution, where the cut-off point might be two standard deviations from the mean. If no assumption of normality can be made about the underlying distribution, extreme value analysis can be used, which deals with the extreme deviations from the median of distributions (De Haan, Ferreira and Ferreira 2006). Proximity-based models can also be used (e.g. clustering and density-based methods), where outliers are points isolated from the remaining data on the basis of similarity or distance functions (Aggarwal 2013). In the Egypt research publication case study

(Albanna et al. 2021), positive deviants in each citation metric were identified using a density-based method called the interquartile range (IQR). IQR segments an ordered dataset into quartiles and the values that separate them are denoted by Q1, Q2 and Q3 (Hampel 1974). Positive deviants were defined as observations that fall above Q3 + 1.5*(Q3-Q1).

- **Multivariate Analysis:** this is used in cases where there are contextual variations among the observed units belonging to the same homogeneous group that need to be controlled for (i.e. to reduce their effect). Those contextual/structural variables are used to predict performance for each observed unit using regression analysis, and the positive deviants are identified based on how far the observed performance is from the predicted performance. This increases the likelihood that the identified positive deviants are over-performing due to individual practices and strategies and not due to structural and contextual factors that can be accounted for in the regression. When the performance measure is categorical, probabilistic models such as logistic regression can be used. Positive deviants in this case are the false negatives i.e. observations that based on the independent variables are expected to fail but in fact succeeded. When dealing with continuous performance measures a least-squares fit is typically used (Aggarwal 2013). In the Ecuador cattle-farming project, we used a model to predict farm deforestation rates as a function of farm cattle density, size, soil adaptability, socio-economic variables and the different land uses (Grijalva et al., 2021). Positively-deviant farms were then identified based on the residual values i.e. the difference between predicted and observed deforestation rates.
- **Posteriori Expectation:** a phenomenon based on historical observation is the basis on which the cut-off point is determined. For example, according to the International Union for Conservation of Nature, threatened species are defined as species that suffer a decline in population for three generations, or over 30 years. A positive deviant could be a population of a species whose size is increasing, or is stable, for three generations or more, when the size of other populations of the species is decreasing rapidly.
- **Exceptional Responders:** exceptional responders are units that perform better than expected in response to a certain intervention. An example would be an intervention to protect forests. Forest cover could be measured both inside and outside a protected area. The difference between the inside and outside can be used to generate an average expected effect of protection and positive deviants would be the protected areas significantly exceeding the expected effect. This can be done using the difference-in-differences method (Abadie et al. 2010), where positive deviants would be the units having the largest difference in differences.

4.2.4. *Positive deviant validation*

The previous step aims to identify outliers. However, one risk of using non-traditional data is that it is possible to find "spurious correlations" or, more generally, to misinterpret statistical outliers as positive deviants. This spuriousness is usually caused by confounding factors that the data could not capture, or due to making wrong comparisons in the homogeneous or peer grouping i.e. not comparing like with like (Blastland and Dilnot 2008). Hence, potential positive deviants identified in the previous stage should be approached as a starting point for asking questions rather than as the basis for drawing conclusions. Field research will be needed to ascertain if these are indeed positive deviants. However, there are ways to validate these potential positive deviants before going to the field. We generally recommend reaching out to community leaders, government officials, local domain experts and development professionals who are engaged in activities, projects or services related to the targeted areas before doing the field research. Sharing with them the initial list of potential positive deviants could lead to an early, better understanding of performance, and insight into factors that might have been overlooked or that could have biased results. For instance, there might be development interventions just for positive deviants, such as external support, which can explain their outperformance but which cannot be known from the digital dataset. Additionally, checking if significant contextual predictors of positive deviant performance (e.g. type of irrigation, month of rainfall, age demographics) are in accordance with existing literature and local domain knowledge, could count as a means of validation in itself.

There are also more quantitative ways to validate whether what is identified is simply random noise or false positives, or whether it is a sign of actual positively-deviant performance. One way is to look longitudinally (if data is available) and see if the identified positive deviants outperform over time, or whether their outperformance is a one-off event. In the Indonesia agricultural project, a time series analysis was conducted to see if the performance of rice farming villages was independent of climatic patterns over time compared to non-positively-deviant villages. This was done by developing a model to predict village average enhanced vegetation index (EVI) as a function of precipitation and temperature in 2013 by training it using historic climate and EVI data from 2000 until 2012. The observed performance of positively-deviant villages was significantly higher than the observed performance of non-positively-deviant villages. This implies that outlier villages have likely adopted specific approaches and practices that others have not, and have established production systems that delink climatic patterns and productivity. This provided an initial validation of their positive deviance.

Other validation methods include trying out different sources of data and different techniques to identify positive deviants. Continuing on the Indonesia agricultural project, we used both univariate and multivariate outlier detection techniques to identify potential positive deviants (Albanna et al. 2020), and in the Ecuador cattle-farming project, we modelled deforestation rates using both yearly predictors and interannual variations in predictors. In both case studies, there was greater confidence in the validity of positive deviants that were identified across multiple approaches.[7] An alternative approach could be to use a different dataset. For example, in the Somalia rangeland project, the use of the remote sensing datasets was complemented by the use of open-source high-resolution imagery available from Google Earth for pre-fieldwork visual inspection. The latter was used to rule out false-positive deviants in the former analysis whose vegetation scores were inflated by interventions (e.g. government reserves), and to look for early signs of pastoral and agro-pastoral activities, visible soil and conservation techniques, and other rangeland management practices. It was also used to check if the area was actually occupied by permanent or semi-permanent settlements or not at all. Through this remote inspection, as illustrated in Fig. 4, we identified patterns indicating the existence of soil and water conservation techniques at a number of potential positively-deviant communities (Abdullahi et al. 2021).

4.3. *Stage 3: Discovering predictors of positive Deviant performance*

This section outlines the different steps needed to discover the factors underlying positive deviance. It follows the "Determining Positive Deviants" stage which results in a list of potential positively-deviant units that will be included in the fieldwork sample for further inquiry. The inquiry in this stage refers to the process of finding positive deviants' uncommon but successful strategies and practices that can be shared and acted upon by the population of interest. It starts with fieldwork planning, followed by data collection and ends with data analysis.

[7] Consistency across time, analysis techniques and data sources can all indicate that the identified positive deviants are not random errors, hence, increasing the likelihood of them being true positive deviants. However, this does not totally exclude the possibility that their outperformance could be due to externalities that were not controlled for in the identification and were not accounted for in the validation.

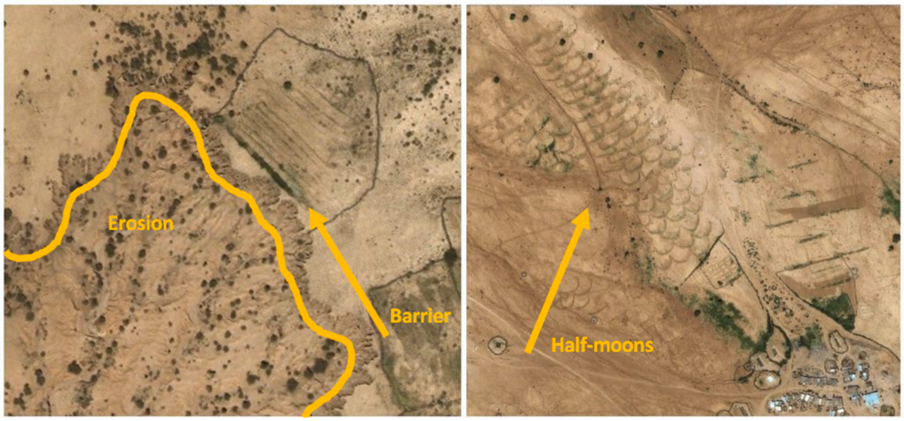

Fig. 4. Examples of soil and water conservation techniques. (On the left, there is a shrub barrier placed to limit the expansion of soil erosion. On the right can be seen half-moon techniques to reduce water run-off (Source: Abdullahi et al., 2021).).

4.3.1. Fieldwork planning

The goal of the fieldwork is twofold: 1) to confirm the validation of positive deviants identified in the previous stage, and 2) to uncover the underlying factors responsible for their deviance. The latter should include other stakeholders who have an indirect or direct relationship with the unit of analysis, and could influence its performance. For example, in the Ecuador cattle-farming project, our unit of analysis was cattle-raising farms and this stage therefore targeted both farm owners and farm workers as direct stakeholders, with government officials identified as indirect stakeholders. Fieldwork planning should therefore start with a scoping activity: identifying the different stakeholders, and developing further familiarity with the social and cultural environment of the targeted population.

Conceptual Framework: Before developing the data collection tools, it is necessary to identify relevant variables for the field study, and to understand how they might relate to each other, and how they will be measured. One way in which this may happen is through use of existing conceptual or theoretical frameworks from the literature that have been used to explain the investigated phenomenon. Where they exist, such frameworks will likely already have been identified in Stage 2 as part of mapping relevant PD outcome and contextual variables. They may otherwise need to be created at this point. Having linked variables through a conceptual framework, this can then be discussed with key informants in the project domain and with the actors involved in the previous stage to make sure that all relevant variables are included. This particularly helps ensure that any contextual variables used in positive deviant identification that might require field validation, will be included in the data collection tools. Fig. 5 presents the example of a framework that was used as the basis to develop the questionnaire tool in the Ecuador cattle-farming project.

Study Design: After developing the conceptual framework and mapping out the different stakeholders, the strategy for collecting data from those stakeholders must be determined. This can use a qualitative approach (e.g. interviews), a quantitative approach (e.g. surveys) or a mix of both. In the following 'Data Collection' step we will present the different methods that can be used in each of those approaches. However, due to the nature of the DPPD method – which covers populations that are relatively larger than in the conventional PD approach – a mixed-methods approach is likely most appropriate. This is because it supports the combined analysis of a small information-rich sample of positive deviants to qualitatively generate hypotheses about individual, cultural, social and structural predictors of positive deviant performance via inductive reasoning, while also leveraging large samples to validate the generated hypotheses quantitatively via deductive reasoning. Fig. 6 presents the proposed mixed methods study design for DPPD projects. This was used, for example, in the Egypt research publication case study, where positively-deviant researchers were interviewed first to generate hypotheses about the basis for their performance, and then quantitative data were collected from both positive-deviant researchers and nonpositive-deviant researchers to validate those hypotheses and identify significant differences between both groups (Albanna et al., 2021).

The first step of Fig. 6 could also include a few non-positive deviants to establish an understanding of the population's normative and common behaviour before interviewing positive deviants; hence, making it easier to identify uncommon practices and strategies. However, it is advised to build this normative framework at an earlier stage of the study ('Assessing Problem-Method Fit') because it might uncover key variables that are needed in determining positive deviants. Furthermore, in cases when there are limited resources constraining the ability to conduct large scale surveys, following a qualitative approach targeting both positive deviants and non-positive deviants rather than a mixed-methods approach could be more practical. Conversely, when doing retrospective studies using secondary data sources or when it is difficult to have face-to-face engagement with the study participants, a quantitative approach could be more suitable.

4.3.2. Data collection

This step involves collection of the data needed to identify factors that enable positive deviants to achieve better outcomes than their peers. While the conventional PD approach focuses mainly on individual-level factors, the DPPD method – due to its breadth of coverage – can employ a 'systemic' lens that takes into account factors beyond individuals (e.g. infrastructure, policies, social system dynamics, etc.). This enables DPPD to capture a more comprehensive understanding of the complex forces at play behind a 'solution' and can lead to both community-level and policy-level interventions; though at both these and individual levels, the focus should always be on identifying factors that are transferable and controllable. Hence, it is important to design the data collection instruments in such a way as to capture this mix of factors. There are several methods that can be used for this purpose, and the choice of participants included in each method depends on the study design.

While time-consuming, *qualitative methods* provide deeper insight into the factors underlying positively-deviant performance and, as shown in Fig. 6 for a mixed-methods approach, are particularly associated with generating hypotheses about positive deviants. Interviews, focus groups and observation are all relevant techniques. In the projects, we used semi-structured interviews targeting a sample of positive deviants and non-positive deviants (for example, 18 farmers – 9 positive deviants and 9 non-positive deviants – were interviewed at their farms in the Ecuador cattle-farming project). The interview schedules contained closed-ended questions and observational checklists to capture contextual, demographic and socio-economic variables in addition to variables developed from the conceptual framework. Open-ended questions were

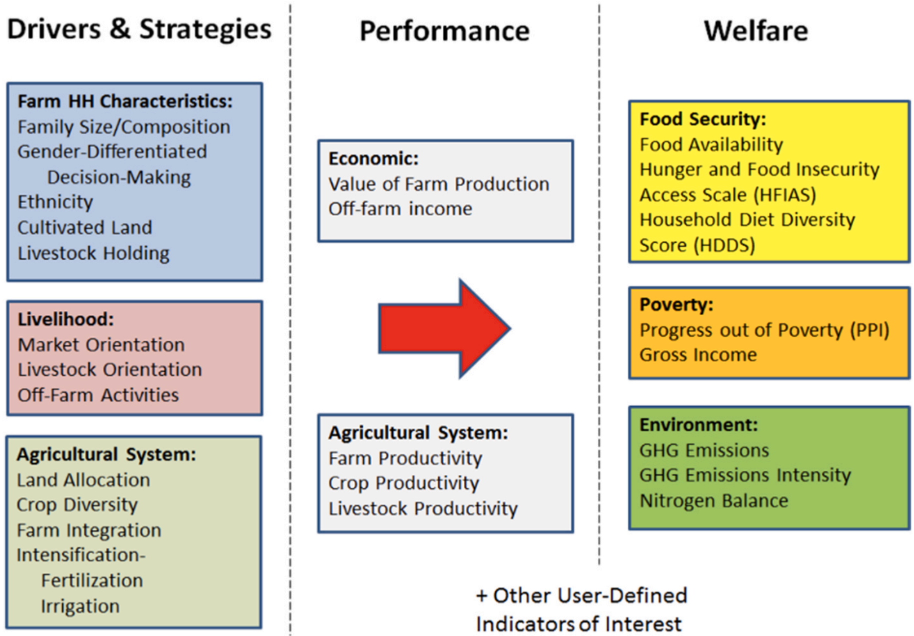

Fig. 5. Conceptual framework of key farm livelihood indicators (Wijk et al., 2016).

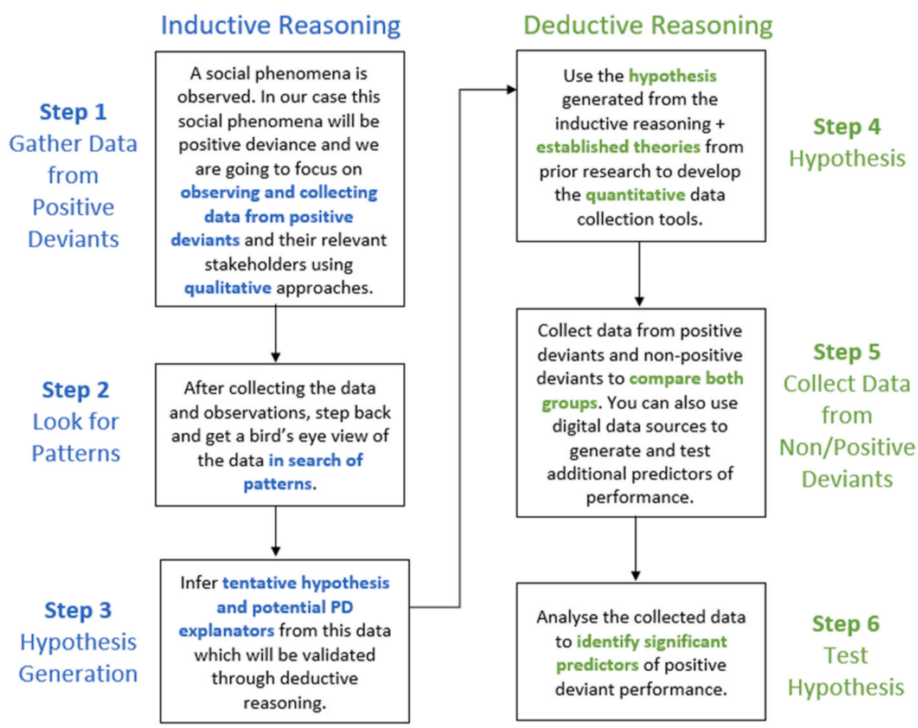

Fig. 6. DPPD study design.

also used to uncover the uncommon strategies, attitudes and practices of positive deviants by comparing them with those of non-positive deviants.

Community-based participatory methods have been quite widely used in the conventional PD approach. They have been shown to mobilise populations, create buy-in, increase knowledge and change attitudes. In such methods the researcher acts as a facilitator who creates a space to integrate the expertise of both insiders and outsiders who contribute equally to the PD inquiry, community capacity building and action (Teufel-Shone et al., 2019).

Examples of qualitative participatory methods, the first of which has been used in the projects to date, include:

- **Community Mapping:** the process and product of a community getting together to map their own assets, values, beliefs, spatial units of interest or any other self-selected variable. In the Somalia rangelands projects, participatory mapping (see Fig. 7) has been used to: 1) identify community resources and infrastructure; 2) understand the mobility patterns of livestock and pastoralists (i.e. transhumance); 3) understand the conditions of the rangeland and map both problem

Fig. 7. Community mapping at the village of Shilmaale

areas and bright spots of rangeland conservation; and 4) understand the different land uses and user groups in the community and areas where there is potential conflict. Similarly, feminist participatory cartography was used in the Mexico safe public spaces project, to visually represent the experiences and knowledge of women who live, work, study, visit or transit through the urban spaces that were targeted for fieldwork. 89 women across the 16 selected spaces (10 positive deviants and 6 non-positive deviants) were invited to highlight on a map where they hang out, rest, shop, among other activities, as well as places where they feel safe and places where they feel unsafe and why they feel this way about those places. Those individual area maps were then amalgamated into a collective one as shown in Fig. 8.

- **Participatory Sketching**: a method of collective drawing employed to obtain enriched narratives from participants (Greiner et al. 2010). Participants jointly draw a sketch describing what they envision as good practice or an ideal model in a physical space, and then share and discuss. This has been used in PD studies when visual aids are required to identify positively-deviant practices (Nieto-Sanchez et al., 2015).
- **Discovery and Action Dialogues (DADs):** a key technique used in PD, the aim of DADs is to ensure that in the presence of a facilitator, people in the group, unit, or community discover by themselves the positively-deviant practices (Escobar et al., 2017). DADs are argued to create favourable conditions for stimulating participants' creativity in spaces where they can feel safe to invent new and more effective practices; to reduce resistance to change as participants are given the freedom to choose the practices they will adopt and the problems they will tackle; and to increase the likelihood that solutions will be adopted by creating local ownership. DADs are thus seen as a basis for both discovering PD practices and mobilising communities to take action.
- **Photo Elicitation:** a method used in visual anthropology that introduces pictures to elicit comments (Lindlof and Taylor 2017). For example, pictures taken during interviews with positive deviants can be presented to focus group participants. Using these pictures as reference, the participants are asked to reflect on the captured practices and solutions.
- **Data-Driven Participatory Approaches:** include methods that engage community members in interpreting the data that were collected about them in order to catalyse dialogue and debate around the challenges they are facing and means to address those challenges (Cañares 2020). Alongside transforming community members from passive producers of data into active users, this can be used to elicit PD-related evidence from the community.

Quantitative methods, as noted above and in Fig. 6, can be used to test hypotheses about positive deviants using statistical analysis. For example, quantitative surveys can collect structured data from both positive deviants and non-positive deviants to identify statistically significant differences between both groups. Quantitative observation checklists can also be applied, containing a list of things that the observer will look at when observing positive deviants and non-positive deviants. Usually, it incorporates contextual variables that are used to identify positive deviants and require ground validation. For instance, in the Ecuadorian cattle-farming project, we used vaccination data as a proxy of cattle numbers in the farm (Grijalva et al., 2021). The field team had to validate this proxy by counting the real number of cattle. Knowing we found a good correlation between the vaccination data and actual cattle headcount, we were then able to propose that such data can likely be used for the same purpose in other studies.

The DPPD method provides an opportunity to use quantitative digital datasets not just for identification of positive deviants but also for understanding their underlying behaviours and practices. While this is not possible in most cases, it is still important to ask the question "Are there digital traces that can shed light on positive deviant behaviours and practices?". In the Egypt research publication case study, we applied machine learning and content analysis techniques to the researchers' publications to identify paper-extrinsic factors (e.g. number of pages), paper-intrinsic factors (e.g. topics covered) and publication outlets (e.g. where do they publish their research) that could shed light on publication strategies and tactics of those positively-deviant researchers whose research was highly cited (Albanna et al. 2021).

4.3.3. Data analysis

The main aim of this step is to identify significant predictors of positive deviants that distinguish them from non-positive deviants. Data analysis techniques will largely depend on the selected study design: qualitative, quantitative or mixed-methods. At the heart of the qualitative data analysis is thematic analysis of verbatim interview and focus group transcripts to extract the attributes, attitudes, practices and strategies of positive deviants. Such analysis can also quantify the frequency of occurrence of these variables, offering some measure of difference between positive deviants and non-positive deviants. In a mixed-methods approach as per Fig. 6, the themes inductively identified can be used to develop a survey instrument that seeks to quantitatively validate the qualitative findings (uncommon PD predictors) using a large representative sample of the population. For example, in the Egypt research publication case study, the qualitative analysis of the interviews with PDs led to the discovery of predictors that proved to be significant in the following quantitative analysis of the surveys targeting both positive deviants and non-positive deviants. Examples of those predictors include, but are not limited to: publishing with foreign reputable authors, and taking scientific and formal writing courses (Albanna et al. 2021).

PD studies use three main types of quantitative analysis: descriptive statistics, inferential statistical tests and regression analysis. Descriptive statistics are used as the first step of statistical analysis for either two-group comparison (positive deviants vs. non-positive deviants) or three-group comparison (positive deviants vs. two other groups: average performers and negative deviants who significantly underperform). Statistical tests provide basic comparative information for these groups (differences between group means, minima and maxima, etc.) and also establish whether differences are statistically significant when comparing either the two groups (e.g. via student t-test, Mann Whitney and Fisher exact test) or three groups (ANOVA, Kruskal–Wallis and chi-square). For example, in the Mexico safe public spaces project, this

[8] This map was provided by the Community Cohesion and Social Innovation organisation, which is the fieldwork consultancy for the Mexico project.

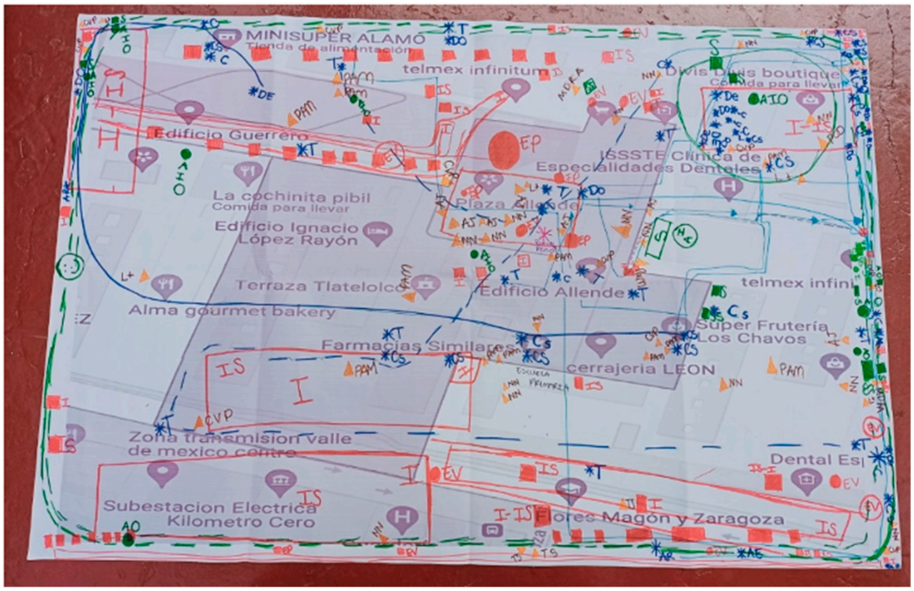

Fig. 8. Collective map of safe (green) and unsafe (red) areas for women in one part of Mexico City.[8] (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

analysis was used to identify differences between positively-deviant AGEBs and non-positively-deviant AGEBs. Early findings revealed that positively-deviant AGEBs, in some homogeneous groups, had a higher percentage of streets with informal commerce and public lighting, more poles with cameras and more "intersecciones seguras" interventions[9] compared to non-positively-deviant AGEBs. Regression analysis is used to examine the relationship between the identified positive deviant performance measure (as dependent variable) and the independent variables. It can include both the structural variables and controls that were used in the positive deviant identification step and the socio-demographic and behavioural variables captured in the data collection step. For each homogeneous group, we recommend having a separate model to identify significant predictors of performance that are relevant to the context of the respective groups.

5. Results

This section summarises the preliminary results from applying the first three stages of the DPPD method across the five pilot projects that conducted fieldwork.[10].Table 2 presents for each project the study population, the number of potential positive deviants identified from Stage 2, and potentially-transferable explanators of outperformance that were generated from the fieldwork investigation and/or the analysis of secondary data.

The predictors of PD performance could be divided into those (listed in the table) that could potentially be transferred in some way, and those that could not. For example, in the Egypt research performance case study, we were able to identify potentially-transferrable practices that could enhance the publication outcomes of Southern researchers, such as participation in multi-country teams, taking specific training courses, or focusing on publishing journal articles rather than conference papers. But there were also factors contributing to better outcomes that are not transferable such as seniority, publication age or being male. Similarly, in the Ecuador cattle farming project, although we did find some attitudes and practices in favour of forest conservation that might be scaled out to other farmers, in most cases, low deforestation was a result of contextual and structural factors. For instance, in some farms there were topographic limitations where trees on ravines, swamps and mountainous zones were untouched simply due to their geographic inaccessibility and the difficulty of land use.

The potentially-transferable PD explanators emerging from the field could feed into both community and policy interventions. Explanators relevant to community interventions are mainly practices, strategies and know-how of positively-deviant individuals that could be transferred and replicated by other individuals in the group. Examples include Zai holes (holes for seeds that contain organic fertiliser and trap rain) and the knowledge of what would be "useful rain" in Niger; and rotational reseeding and strip grazing in Somalia. Explanators feeding into higher-level policy design include existing successful policy interventions, government schemes or physical features of positively-deviant geographic units that could be adjusted to achieve the desired outcomes. For example, in the Mexico safe public spaces project, better lighting and the presence of informal commerce, especially if the merchants are known to women or are women, were associated with lower reported crime rates. These elements could inform policies on the design of urban spaces, making them safer. In the Somalia rangelands project, pastoralists stressed the importance of establishing government reserve areas. These reserves emerged as crucial for saving livestock during the dry season or droughts and, hence, their creation and effective management would be a policy recommendation.

When designing community-level interventions from these projects, we will focus on creating activities that enable people to share, learn and practice the behaviours of positive deviants. Activities will be designed to warrant the active participation of those who developed the solution i.e. positive deviants, those who stand to benefit from adopting a positively-deviant practice i.e. non-positive deviants, as well as the different stakeholders who might have an influence on the overall adoption of the solution. Examples for communities could be community gatherings where positively-deviant pastoralists demonstrate how they build the stone lines and bunds to limit soil erosion, or where they explain how village leaders limit illegal land enclosures and manage conflict in the communal areas using effective customary laws.

6. Discussion: lessons learned

Having provided details on the first three stages of the DPPD method

[9] Safe intersections programme: https://www.eluniversal.com.mx/metropoli/cdmx/con-intersecciones-seguras-se-redujo-un-30-los-accidentes-viales.

[10] The Indonesia agricultural project covered only the first two stages of the DPPD method.

Table 2
Summary of results from DPPD projects.

Project	Study Population	Positive Deviants	Potentially-Transferable Positive Deviance Explanators
Research publication outperformance in Egypt (Albanna et al. 2021)	203 information systems researchers who are affiliated to public universities	26 potential positive deviants were identified	- Taking scientific writing and English language courses - Supervising a large number of postgraduate students - Publishing with foreign authors - Establishing research teams overseas - Obtaining PhD degrees from global North universities - Securing research grants and travel funds - Publishing more journal articles and fewer conference papers - Working on established research areas rather than on radical research topics - Having multiple authors and affiliations in their publications
Rangeland preservation by pastoral communities in Somalia (Abdullahi et al. 2021)	314 communities in the West Gollis area of Somaliland, which is a zone with a majority pastoral community	13 communities were identified as potential positive deviants	- Women are part of the village development committee and are involved in decision making - Trained community-based animal health workers - Dedicated village committees for conflict resolution - Dedicated natural resource management committees - Stone lines and stone bunds to slow down water runoff - Rotational reseeding and strip grazing of pasture mixes such as Rhodes grass and Lablab legume - Income diversification (beekeeping, power food preservation, growing fodder) - Community land enclosure policies - Customary laws for tenure management - Drip irrigation - Moving towards agro-pastoralism - Tree planting
Cereal-farming outperformance in Niger (Gluecker et al. 2021)	12,093 communities in the Sahelian region, where there is a predominance of rain-fed agriculture	180 communities were identified as potential positive deviants	- Leaving millet stalks and stems in the field to protect the soil from wind erosion and help restore the organic matter - Existence of *Faidherbia albida* ("Gao tree") which helps fertilise the soil - Existence of Zai holes and stone bunds to reduce surface water run-off - Sowing only after useful rains[a] - The rational use of mineral fertiliser with organic manure
Public spaces in Mexico City where women are safer (Cervantes et al. 2021)	2431 AGEBs in Mexico City	32 AGEBs were identified as potential positive deviants	- Presence of informal commerce - Better lighting (intensity, scheduling, location and distribution) - More poles with cameras - The existence of safe intersection programmes - Lower percentage of green areas - Presence of activities and facilities for specific demographic groups (women, children, families and elderly)
Low deforestation cattle-farming in the Ecuadorian Amazon (Grijalva et al. 2021)	5332 farms in Joya de los Sachas canton and 5701 farms in Sucúa canton	53 potential positive deviants were identified across both cantons	- Income diversification - Motivation to plant native trees - Rotational grazing - Finding other sources of animal feed which reduces pressure on pasture - Realising the value of trees in their grazing systems

[a] Useful rain (14 mm in fallen water height) is the point at which producers can sow usefully. Sowing only after a useful rain helps to reduce seed loss considerably.

and the results produced to date, we now draw out some of the key lessons we learnt while applying the method, as developed from reflections of both the global and country-level teams during learning calls and online surveys undertaken as part of the six projects. These reflections highlight both the limitations and opportunities of applying DPPD, and its future potential.

6.1. DPPD is not universally applicable

DPPD is not a method that could be applied to every PD-amenable problem. This is because non-traditional digital data that could be utilised in DPPD must be capable of capturing the performance of the observed units without compromising their privacy. This is difficult to achieve in culturally-sensitive domains, such as limiting HIV transmission or fighting against female genital mutilation, where the conventional PD method has been applied successfully. Additionally, open digital data is rarely available at the level of individuals, mainly due to the prerequisite to de-identify and aggregate digital observations to make them open. This makes DPPD better suited to development problems with communities or geographical areas as the unit of analysis, with the exception of a few domains where the digital outcomes of individuals can be traced and quantified without compromising their privacy (e.g. scientific research outputs). Finally, the DPPD method relies heavily on the existence of reliable and accessible digital and secondary data that captures outcomes directly related to the addressed development problem. In domains and countries with poor data landscapes, applying the DPPD method may not yet be feasible.

6.2. The right know-how must be available

Finding potential positive deviants from non-traditional data without sufficiently understanding their contextual realities will likely lead to false positives. Hence, a unique combination of local, domain and data knowledge is needed before conducting any data analysis. Country-specific domain knowledge is crucial in understanding the normative behaviours of the investigated population, if positive deviants exist, and the contextual and structural factors that have an effect on their outcomes. Domain-specific data knowledge is required to identify relevant performance indicators from the available data, in addition to mapping out suitable data sources that could be used. However, such expertise is usually missing within international organisations. Therefore, an initial mapping of existing and missing relevant know-how for the project can help uncover necessities for bringing in additional know-how.

6.3. Control for contextual variables

The conventional PD method generally covers small sample sizes, e.g. a few dozen families, in a homogeneous context, e.g. a single village or district. This makes it very accurate in singling out a particular behaviour that explains a successful practice since non-behavioural factors can be largely neglected as they are more or less the same for the entire (small) population being investigated. Digital performance measures used in DPPD can cover large geographic areas enabling the inclusion of larger populations in the analysis. This increases the heterogeneity of the sample and the likelihood of potential confounding factors when identifying positive deviants. For example, structural factors such as access to roads and levels of rainfall, and socio-economic factors such as population density, differ across large populations and could contribute to differences in performance among units of analysis. Failure to control for those structural factors when identifying positive deviants leads to an inability to single out the particular attributes, practices and strategies that need to be disseminated. The biggest challenge here is identifying additional data sources, both traditional and non-traditional, that can link the context to the digital performance measure. Additionally, such contextual data should have an overlapping time frame and spatial resolution with the performance measure to be useful. In Ecuador, we used satellite imagery to calculate deforestation rates for a large sample of cattle-raising farms. However, cattle density is an important confounding factor, as higher density makes the recovery of pasture harder, and requires more grazing space which creates pressure to deforest. We used cattle vaccination data as a proxy of cattle numbers on the farm to identify positive deviants with low deforestation rates relative to their cattle density and not in absolute terms. This increased the chances of attributing low deforestation rates to sustainable cattle-ranching practices and not to lower cattle density.

6.4. If possible, measure performance over time

An advantage of using digital measures of performance is that they often have a longitudinal coverage and are collected at regular intervals. This allows performance to be evaluated over time, and to observe moves towards or away from positive deviance. Furthermore, it enables identification of persistent positive deviants: those who appear as positive deviants in the data for several consecutive years are more likely to be “true” positive deviants. In the Ecuador cattle-farming project, we were able to measure deforestation rates over a five year period. We were able to develop a more nuanced understanding of positive deviants: those who became positive deviants over time (from low performing to high performing) or those who stopped being positive deviants (from high performing to low performing). Such diversity in positive deviant categories can help uncover interesting factors that trigger moving from one state to another and can inform the design of interventions. Moreover, the same digital datasets that are used to capture performance longitudinally can readily be used to monitor and evaluate the mid-to-long-term effects of scaling the practices and strategies of positive deviants across intervention populations.

6.5. Adopt a holistic approach in understanding PD

The potentially-wide spatial coverage of DPPD, when compared with the conventional PD method, provides an opportunity to observe units of analysis that are beyond individuals e.g. villages or regions. Discovering determinants of outperformance within such units requires a new type of inquiry that looks at factors beyond individuals that could be modified and transferred. Such factors include, but are not limited to, governance mechanisms, development interventions, policies, systemic changes, etc. Early findings from our pilots suggest that the DPPD method might be a promising way to better understand the interactions between individual and supra-individual factors. This can inform the design of nuanced interventions that take into account such interactions, hence, increasing their effectiveness and contextual fit. This is different from the conventional PD method which is placed in a more ‘controlled’ environment where variation in performance might indeed be attributed only to individual-level factors. As a case in point, in the Somalia rangelands project, we realised through conversations with local experts that rangelands health is influenced by individual and community behaviours e.g. soil and water conservation techniques, alternative livelihoods, along with land tenure policies and campaigns against private enclosure. Hence, when planning for our field investigation of positively-deviant communities we decided to embrace the complex dimensions of the rangeland problem and explore positive deviance as a system behaviour instead of looking into positive deviants as individuals in isolation from the larger system. Findings from this investigation could thus inform the design of both community-level and policy-level interventions.

6.6. Earth observation data is a low hanging fruit for DPPD

After applying the DPPD method to multiple projects and domains, it is clear that earth observation (EO) data can play an instrumental role in the viability and scalability of the DPPD method. EO gathers data about the physical, chemical and biological systems of the planet using remote sensing technologies (Rast and Painter 2019). It is considered the most cost-effective technology able to provide data at a global scale. It can be acquired at low cost, over long periods of time, and thanks to the recent advances in remote sensing technologies, it is witnessing a growing availability at a high resolution including coverage of lowest-income countries where other datasets are lacking. Such attributes of EO data make it possible to overcome a number of data accessibility limitations, while being able to capture the potential gains of using big data in PD such as reducing the cost, time and risk of measuring performance at large scale. Of course, limitations must be acknowledged given that EO data is applicable only for problems where the impact of human behaviours and practices on natural and built environments can be observed and measured remotely. For instance, EO data has proved useful in our projects to identify positive deviance in vegetation health and forest cover (assuming that this observed deviance can be linked to individual practices and strategies, or successful policies and governance mechanisms on the ground). Additionally EO data was useful in the homogeneous grouping step, where remote sensing-derived covariates (e.g. temperature) were extracted to create peer groups having similar conditions. High resolution satellite imagery also proved useful in validating potential positive deviants by ensuring the accuracy of the land covers used for PD identification and in identifying hints of positively-deviant practices in the rare case that they are observable.

7. Conclusion

This paper has presented the three core stages of the data-powered positive deviance (DPPD) method; a new way of applying the positive deviance approach by combining non-traditional, digital data (e.g. online and remote sensing data) with traditional data (e.g. interviews, official statistics). These core stages are: assessing problem-method fit, determining positive deviants, and discovering positive deviant practices and strategies. The remaining two stages covering the design of interventions and monitoring and evaluating the effects of those interventions were not included in the presented method for two reasons: the majority of the projects reported here did not yet reach these stages, and these stages should not differ much from the conventional PD approach. However, investigation of the potential value-added benefits that could be incorporated into those two stages from the use of non-traditional data is a future direction of this work.

More generally, the DPPD method makes it possible to identify and characterise positive deviants at temporal and geographical scales that are not possible using the conventional approach. While the use of existing datasets may reduce initial time/financial costs of PD identification compared with traditional PD methods, DPPD overall is not yet demonstrably cheaper and quicker because there can be additional costs associated with the access and analysis of datasets, because DPPD itself will typically involve fieldwork, and because none of the pilot projects is yet in a position to allow total and comparative lifecycle costs to be calculated. The presented method was developed iteratively through its application by the DPPD initiative partners in six projects across five different development domains. Through readily available digital data we were able to observe and capture outcomes of large populations in relation to the addressed development problems; however, this came with the challenge of controlling for numerous contextual factors, parts of which were feasible while others not. The large temporal coverage of digital data enabled not only the identification of sustained positively-deviant behaviour (e.g. in consecutive years) but also changes in behaviours (i.e. becoming positive deviants or no longer being positive deviants). Additionally, accessing relevant data turned out to be much harder than expected. This highlights the necessity to forge the right partnerships and involve the various data-controlling stakeholders at an early stage of the project.

The DPPD method relies heavily on a digitally recorded or observed performance measure that is directly related to the desired outcomes of the observed units. However, the selection of this measure highly depends on the data availability in a given country and domain. Flexibility and creativity in dealing with a potential lack of data, while adhering to the original focus of the development challenge, requires constant iteration, reflection and discussion with domain experts. It is also evident that DPPD requires an interdisciplinary team which combines the right local, domain and data analysis know-how to conduct plausible data analysis for PD identification that uses relevant performance measures while controlling for potential confounding factors. That specialist expertise can be in short supply and one future aim would be to 'democratise' the method, enabling it to be accessible by a wider range of organisations including, potentially, community-based organisations and others involved with citizen science. Finally, while the conventional PD approach focuses mainly on individual-level factors, the DPPD method – due to its large spatial coverage – could employ a more holistic lens that takes into account both individual and supra-individual factors.

We hope that the details of the DPPD method provided here enable its uptake by development and data science professionals, and we encourage its application to a wider range of development challenges and in a wider set of development domains, with further refinement of both the method and the lessons learned.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Catherine Vogel from the GIZ Data Lab for leading the effort to establish the DPPD initiative. We thank the in-country United Nations Development Programme Accelerator Lab teams, the Deutsche Gesellschaft für Internationale Zusammenarbeit projects and all their local partners for implementing the Ecuador, Mexico, Niger and Somalia projects. We would specifically like to thank Ana Grijalva, Paulina Jimenez and Carolina Michell from the Ecuador project; Gabriela Rios, Alejandra Cervantes and Daniela Torres from the Mexico project; Assane Boukar, Moustapha Sahiro, Moctar Seydou, Rachida Mani and Damien Hauswirth from the Niger project; and Hodan Abdullahi, Erik Fritzche and Florian Fritzche from the Somalia project. We would also like to thank Gunnar Hesch and Esther Barvels from the GIZ, and Erik Lehmann from the GIZ Data Lab for their support in the geographic information system and remote sensing analysis conducted in the Somalia and Niger projects. We are also grateful to the United Nations Global Pulse Lab Jakarta, particularly Dharani Dhar Burra and Michael Dyer, for their support in implementing the Indonesia project. This work was financially supported by the GIZ Data Lab and the in-country UNDP Accelerator Labs and GIZ projects.

References

Abadie, A., Diamond, A., Hainmueller, A.J., 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J. Am. Stat. Assoc. 105 (490), 493–505.

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Comput. Stat. 2 (4), 433–459.

Abdullahi, H., Albanna, B., Barvels, E., 2021. Rangelands Defying the Odds: A Data-powered Positive Deviance Inquiry in Somalia. Data-powered Positive Deviance. https://dppd.medium.com/rangelands-defying-the-odds-a-data-powered-positive-devianceinquiry-in-somalia-90772de392dd. (Accessed 4 August 2021).

Aggarwal, C.C., 2013. Outlier Analysis. Springer, New York, NY.

Albanna, B., Dhar Burra, D., Dyer, M., 2020. Identifying Potential Positive Deviants (PDs) across Rice Producing Areas in Indonesia: an Application of Big Data Analytics and Approaches. Pulse Lab Jakarta, Jakarta.

Albanna, B., Handl, J., Heeks, R., 2021. Publication outperformance among global South researchers: an analysis of individual-level and publication-level predictors of positive deviance. Scientometrics 126 (10), 8375–8431.

Albanna, B., Heeks, R., 2019. Positive deviance, big data, and development: a systematic literature review. Electron. J. Inf. Syst. Dev. Ctries. 85 (1), e12063.

Ang, Y.Y., 2019. Integrating Big Data and Thick Data to Transform Public Services Delivery. IBM Center for The Business of Government, Washington, DC.

Blastland, M., Dilnot, A.W., 2008. The Tiger that Isn't: Seeing through a World of Numbers. Profile Books, London, UK.

Bornakke, T., Due, B.L., 2018. Big–thick blending: a method for mixing analytical insights from big and thick data sources. Big Data Soc. 5 (1), 2053951718765026.

Bradley, E.H., Curry, L.A., Ramanadhan, S., Rowe, L., Nembhard, I.M., Krumholz, H.M., 2009. Research in action: using positive deviance to improve quality of health care. Implement. Sci. 4 (1), 1–11.

Cañares, M., 2020. Three examples of data empowerment. Data Empower. https://medium.com/data-empowerment/three-examples-of-data-empowerment5f3e964ffbdc. (Accessed 21 August 2021).

Cervantes, A., Rios, G., Soto, I., 2021. Identifying Safe(r) Public Spaces for Women in Mexico City. Data-powered Positive Deviance. https://dppd.medium.com/identifyingsafe-r-public-spaces-for-women-in-mexico-city-4f3d49d269d6. (Accessed 8 April 2021).

Data-powered Positive Deviance, 2020. Launching the Data-powered Positive Deviance Initiative. Data-powered Positive Deviance. https://dppd.medium.com/. (Accessed 8 April 2021).

De Haan, L., Ferreira, A., Ferreira, A., 2006. Extreme Value Theory: an Introduction. Springer, New York, NY.

Escobar, N.M.O., Márquez, I.A.V., Quiroga, J.A., Trujillo, T.G., González, F., Aguilar, M.I. G., EscobarPérez, J., 2017. Using positive deviance in the prevention and control of MRSA infections in a Colombian hospital: a time-series analysis. Epidemiol. Infect. 145, 981–989. https://doi.org/10.1017/S095026881600306X.

Gluecker, A., Lehman, E., Barvels, E., 2021. Searching for Positive Deviants Among Cultivators of Rainfed Crops in Niger. Data-powered Positive Deviance. https://dppd.medium.com/searching-for-positive-deviants-among-cultivators-of-rainfed-cropsin-niger-8dbbcceaf4ec. (Accessed 4 August 2021).

Greiner, K., Singhal, A., Hurlburt, S., 2010. "With an antenna we can stop the practice of female genital cutting": a participatory assessment of ASHREAT AL AMAL, an entertainment-education radio soap opera in Sudan. Investigación & Desarrollo. 15 (2), 226–259.

Grijalva, A., Jiménez, P., Albanna, B., Boy, J., 2021. Deforestation, Cows, and Data: Data-powered Positive Deviance Pilot in Ecuador's Amazon. Data-powered Positive Deviance. https://dppd.medium.com/deforestation-cows-and-data-data-powered-positive-deviancepilot-in-ecuador-s-amazon-648aa0de121c. (Accessed 3 August 2021).

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Am. Stat. Assoc. 69 (346), 383–393.

Hauswirth, D., Yaye, H., Soumalia, A.S., Djariri, B., Lona, I., Abba, M.B., 2020. Support for the Concerted Formulation of SPN2A for the Republic of Niger: Identification and Assessment of Climate-Smart Agriculture Options Priority for Adaptation to Changes Climate in Niger (Volume 1). Niamey, Niger. Baastel - BRL - ONFI, Brussels, Belgium.

Hilbert, M., 2016. Big data for development: a review of promises and challenges. Dev. Pol. Rev. 34 (1), 135–174.

Kovács, J., Kovács, S., Magyar, N., Tanos, P., Hatvani, I.G., Anda, A., 2014. Classification into homogeneous groups using combined cluster and discriminant analysis. Environ. Model. Software 57, 52–59.

Lindlof, T.R., Taylor, B.C., 2017. Qualitative Communication Research Methods. Sage Publications, Thousand Oaks, CA.

Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalisation, 121 (1–4), 217–238.

Nieto-Sanchez, C., Baus, E.G., Guerrero, D., Grijalva, M.J., 2015. Positive deviance study to inform a chagas disease control program in southern Ecuador. Memórias do Inst. Oswaldo Cruz 110, 299–309.

Pascale, R., Sternin, J., Sternin, M., 2010. The Power of Positive Deviance: How Unlikely Innovators Solve the World's Toughest Problems. Harvard Business Press, Boston, MA.

Positive Deviance Initiative, 2010. Basic Field Guide to the Positive Deviance Approach. Tufts University, Medford, MA.

Post, G., Geldmann, J., 2018. Exceptional responders in conservation. Conserv. Biol. 32 (3), 576–583.

Rast, M., Painter, T.H., 2019. Earth observation imaging spectroscopy for terrestrial systems: an overview of its history, techniques, and applications of its missions. Surv. Geophys. 40 (3), 303–331.

Smets, A., Lievens, B., 2018. Human sensemaking in the smart city: a research approach merging big and thick data. In: Ethnographic Praxis in Industry Conference Proceedings, vol. 2018, pp. 179–194, 1.

Somekh, B., 1995. The contribution of action research to development in social endeavours: a position paper on action research methodology. Br. Educ. Res. J. 21 (3), 339–355.

Sternin, J., 2002. Positive deviance: a new paradigm for addressing today's problems today. J. Corp. Citizen. (5), 57–62.

Sternin, J., Choo, R., 2000. The power of positive deviancy. Harv. Bus. Rev. 78 (1), 14–15.

Sternin, M., Sternin, J., Marsh, D., 1998. Designing a Community-Based Nutrition Program Using the Hearth Model and the Positive Deviance Approach: A Field Guide. Save the Children Federation, Westport, Conn., USA.

Sternin, M., Sternin, J., Marsh, D.L., 1997. Rapid sustained childhood malnutrition alleviation through a positive-deviance approach in rural Vietnam: preliminary findings. In: Wollinka, O., Keeley, E., Burkhalter, R.B., Bashir, N. (Eds.), The Hearth Nutrition Model: Applications in Haiti, Vietnam, and Bangladesh. World Relief Corporation Headquarters, Baltimore, MD, pp. 49–61.

Taylor, L., Broeders, D., 2015. In the name of development: power, profit and the datafication of the global South. Geoforum 64, 229–237.

Teufel-Shone, N.I., Schwartz, A.L., Hardy, L.J., de Heer, H.D., Williamson, H.J., Dunn, D. J., Polingyumptewa, K., Chief, C., 2019. Supporting new community-based participatory research partnerships. Int. J. Environ. Res. Publ. Health 16 (1), 44.

Trivedi, S., Pardos, Z.A., Heffernan, N.T., 2015. The utility of clustering in prediction tasks arXiv Prepr. arXiv1509.06163.

Trivedi, S., Pardos, Z.A., Heffernan, N.T., 2011. Clustering students to generate an ensemble to improve standard test score predictions. In: International Conference on Artificial Intelligence in Education. Springer, Berlin, Heidelberg, pp. 377–384.

Vilalta, C., Muggah, R., 2016. What explains criminal violence in Mexico City? A test of two theories of crime. Stabil. Int. J. Secur. Dev. 5 (1).

Wallace, M.E., Harville, E.W., 2012. Predictors of healthy birth outcome in adolescents: a positive deviance approach. J. Pediatr. Adolesc. Gynecol. 25 (5), 314–321.

Wijk, M.T.V., Hammond, J., Etten, J.V., Pagella, T., Ritzema, R.S., Teufel, N., Rosenstock, T.S., 2016. The Rural Household Multi-Indicator Survey (RHoMIS): A Rapid, Cost-Effective and Flexible Tool for Farm Household Characterisation, Targeting Interventions and Monitoring Progress towards Climate-Smart Agriculture. CCAFS Info Note.

Wishik, S.M., Van Der Vynckt, S., 1976. The use of nutritional 'positive deviants' to identify approaches for modification of dietary practices. Am. J. Publ. Health 66 (1), 38–42.

Zeitlin, M., 1991. Nutritional resilience in a hostile environment: positive deviance in child nutrition. Nutr. Rev. 49 (9), 259–268.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S.P., Goodman, A., Hollander, R., Koenig, B.A., Metcalf, J., Narayanan, A., Nelson, A., Pasquale, F., 2017. Ten simple rules for responsible big data research. PLoS Comput. Biol. 13, e1005399.