Management performance mapping: the value of information for regional prioritization of project interventions

C. E. Buddenhagen[123], J. Andrade Piedra[4], G. A. Forbes[4], P. Kromann[4], I. Navarrete[4], S. Thomas-Sharma[5], Y. Xing[123], K. A. Garrett[123]

[1]Plant Pathology Department, University of Florida, Gainesville, FL, USA

[2]Institute for Sustainable Food Systems, University of Florida, Gainesville, FL, USA

[3]Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

[4]International Potato Center, Lima, Peru, and Quito, Ecuador

[5]Plant Pathology Department, Louisiana State University, LA, USA

karengarrett@ufl.edu

ABSTRACT

Identifying the locations and settings where technologies are most likely to have important effects can make the most of development or extension efforts. In the context of development and applied ecology, decisions must often be made by policy makers and donors about where to implement projects designed to improve management. Implementation in some regions may provide substantially higher payoffs to investment, and higher quality information may help to target the high-payoff locations. The value of information (VOI) in this context is formalized by comparing the benefits from decision making guided by a set of information and the results of acting without taking the information into account. We present a framework for management performance mapping and for evaluating the value of information for decision making about geographic priorities in regional intervention strategies. In our case studies of Andean and Kenyan potato seed systems, we evaluate seed health and yield information from farms, plots, and individual plant observations. We use Bayesian networks and recursive partitioning to efficiently characterize the relationship between these performance measures and the environmental and management predictors used in studies aimed at understanding seed degeneration. These analyses return the expected performance of an intervention for predictor variables mapped across the landscape. We link the scientific process and the learning cycle to the value of information assessments to support a culture of continuous improvement that informs strategic agricultural development. Assessment of the value of information demonstrates the value of science as an integral part of targeted development programs.

Additional keywords: agricultural development, disease, Ecuador, GIS, intervention ecology, Kenya, pest management, potato, seed degeneration, translational science, value of information, virus, yield gap

A central problem in applied spatial ecology is how to partition management efforts across landscapes. Across larger spatial extents, interventions (e.g., by development organizations or governments) may be designed to increase yield across a region by improving management of agroecosystems. International governmental and non-governmental organizations that seek to reduce poverty, enhance food security, and improve natural resources and ecosystem services, need to understand how to prioritize regional interventions. We propose "management performance mapping" as a tool for translating experimental results to support decision making by policy makers and donors, and assessing the value of information (VOI) to support the analysis. Assessing the VOI involves determining the expected benefit of reducing or eliminating uncertainty (Canessa et al. 2015), as described below. In the absence of uncertainty, when the true state of the system is known, optimal actions can more readily be identified. Often data about agricultural management performance exist, or can be collected inside of existing intervention projects, but the data are collected at the scale of fields, farms or individual plant performance measures. Multiple factors influence plant productivity apart from management, creating uncertainty about the pay-off even where data are relatively abundant. Scaling up models based on limited observations is necessary to visualize how specific interventions are likely to perform at a regional scale (Altieri and Nicholls 2008; van Bussel et al. 2015a; van Wart et al. 2013; Grassini et al. 2015). Management performance mapping can be implemented to visualize the impact of proposed interventions, to improve decision-making and policy setting, in the development of innovations, and during project implementation – as a component of adaptive management in development (Fig. 1).

Data available for identifying good geographic prioritization strategies are generally limited, but evidence-based decision making can be designed to make the most of available data. VOI analyses offer a means of both valuing information and assessing the role of uncertainty in making good decisions (Hirshleifer and Riley 1979; Macauley 2006; Canessa et al. 2015). VOI analyses indicate the value of outcomes for decision making with and without particular information. The VOI should ideally be considered in the context of the mean outcome, associated uncertainty, and what is at stake with respect to making good or bad decisions (e.g., the yield difference measured in yield or money) and how much stakeholders are willing to pay to act on their choices. The utility functions in studies of willingness to pay are related to VOI (Breidert, Hahsler, and Reutterer 2006; Asante Bright Owusu et al. 2011; Hanemann 1991). Many examples in the VOI literature focus on agriculture, such as the uncertainty risk distribution for farm yield (Hirshleifer and Riley 1979), the value of weather forecasting for farmers (Lave 1963), and risk assessment for crop futures (Danthine 1978). A related area of application of VOI concepts is in invasion biology more generally and in conservation biology, where decisions must also be made about where to prioritize efforts (Canessa et al. 2015; Johnson et al. 2017; Wilson 2015). Of course, decision-maker willingness to act based on information is an important prerequisite for information valuation to be meaningful. For example, overly confident decision makers may not be influenced by new information, or they may not reflect on the uncertainty that is inherent in the information available. To our knowledge, VOI analyses have not been applied to plant pathology, crop epidemiology, or to seed system development, where they have the potential to improve decision making.

Precision agriculture and species distribution models both address components of spatial prioritization, at different scales. The question of how to optimize information use for decision making is addressed at the within-field scale in precision agriculture (Tittonell and Giller 2013), allowing well-resourced farmers to collect and utilize spatially explicit data sets (in near real time) about crop performance (Devaux et al. 2010). Farmers are able to choose and optimally apply inputs such as fertilizer, pesticides, and irrigation to areas of the field where they are most needed to maximize yields. Species distribution models address the problem of optimal targeting indirectly, by providing information about where invasive (or endangered) species including pathogens are most likely to be found. They generally are designed to draw inference beyond regions where data were collected by assessing species niche parameters based on occurrences throughout a species' native and introduced range (Hijmans and Graham 2006; Condori et al. 2014; Sparks et al. 2014; Aguirre-Gutiérrez et al. 2017; Austin 2007; Sutherst 1985). Management performance mapping can draw on pathogen species distribution models as one of the components determining how important disease management is likely to be, and how likely it is to be successful.

We present a case study that focuses on smallholder management of seed degeneration in agricultural systems. "Seed degeneration" is the reduction in yield or quality caused by an accumulation of pathogens (especially viruses) and pests in planting material over successive cycles of propagation, where vegetatively propagated crops deserve particular attention because of their higher risk of disease transmission (Sharma and Kang 2003; Thomas-Sharma et al. 2016; Thomas-Sharma et al. 2017; Iritani 1968; Kawakami 1962). Establishing improved seed systems is challenging, especially in low-income countries, due in part to the many system components that must be integrated for seed system success (Jaffee and Srivastava 1994; McGuire and Sperling 2016; McQuaid et al. 2016; Sperling 2008; Gildemacher et al. 2009; Bentley and Vasques 1998). Seed system improvement has been a major focus of agricultural development efforts funded by many agencies (e.g., national plant protection agencies, The Bill and Melinda Gates Foundation, USAID, and FAO), but has often proven to be challenging to implement (Jaffee and Srivastava 1994; Almekinders C. J. M., Louwaars N. P., and De Bruijn G. H. 1994; McGuire and Sperling 2016). In informal seed systems in low-income countries farmers typically use seed saved from the previous season for replanting, often leading to reduced yields, e.g., 5-50% reduction (Devaux et al. 2010), especially when farmers are unfamiliar with approaches for selecting seed with reduced pathogen risk.

Optimizing yield by reducing disease impacts, and improving seed quality, is a primary goal of many seed system interventions. Governments and institutions with a strong focus on science for development, such as CGIAR, work on a suite of factors linked to seed system health. Farmer training efforts focus on options for disease management and optimal decision-making. International development efforts to improve seed systems seek to increase farmer access to disease-free, disease-resistant, high-quality seed, and improve farmer practices to implement "integrated seed health strategies" (Thomas-Sharma et al. 2017; Thomas-Sharma et al. 2016). Attempts to formalize seed systems by imposing seed quality standards may fail if thresholds are unrealistic (Choudhury et al. 2017) and despite many interventions systems may revert to largely informal systems, with sub-optimal seed sourced from on farm much of the time, e.g., 98% of potato seed sources in the Andes are informal (Louwaars, de Boef, and Edeme 2013; Devaux et al. 2010). Interventions are more likely to succeed if they are affordable, and help farmers to be profitable (McGuire and Sperling 2013; Sperling, Ortiz, and Thiele 2013). As an example, positive selection is an on-farm management intervention that can provide large yield benefits, e.g. 28-55% increases (mean 32%) (Gildemacher et al. 2012, 2011). Positive selection is often recommended as part of an integrated seed health strategy. Under positive selection, farmers select healthy appearing plants and mark them for later harvesting of seed potato tubers.

Training farmers in the techniques of positive selection is an attractive potential intervention to support integrated seed health strategies.

A risk framework for an integrated seed health strategy (Thomas-Sharma et al. 2017) examines the factors influencing seed system success (e.g., on farm seed management, disease-free seed sources, amount of inoculum) and provides a useful starting point for the type of interventions that could be included in a management performance mapping analysis for seed systems. Management performance maps for seed systems should use reference data sets that focus on these key factors. In practice, the best available data sets, collected via literature reviews, experiments or on-farm observational studies, will fall short of an ideal data set. As such, scientists, policy-makers and funders interested in evidence-based decision-making often must make decisions based on the available data, even if data are sub-optimal. Management performance mapping can provide a useful starting point for assessing the most effective interventions for large scale (large extent, low resolution) interventions. We envisage that management performance maps could be useful for assessing the likely relative value of management interventions across regions, and as a means of reporting on the effectiveness of interventions. Management performance maps can be used to target limited resources to places in the landscape where impacts are likely to be highest.

Our objectives in this study are to (i) introduce and illustrate the concept of management performance mapping and associated methods, (ii) introduce the use of VOI analysis in this context, (ii) compare methods for identifying predictor variables for management outcomes, and (iv) illustrate the application of management performance mapping to potato seed degeneration data from the Ecuadorian Andes and Kenya.

MATERIALS AND METHODS

A primary decision for agricultural development is where to invest limited resources for an intervention. For the case of seed systems, potential interventions include training, seed multiplier subsidies, and improved variety dissemination. We start by describing the steps involved in producing management performance maps, using the example of training in positive selection by farmers to identify plants more likely to produce healthy seed (Fig. 2). Then we illustrate management performance mapping using a detailed seed degeneration data set from a potato seed study in Ecuador. As a step in preparing the management performance maps, we illustrate the application of Bayesian networks and recursive partitioning for assessing the influence of disease, environmental factors, and management on yield, and similarly assess factors that influence virus incidence. We also evaluate the potential value of the information for guiding selection of locations in development interventions, if the estimated mean performance is correct, for potato seed health in Ecuador and Kenya.

Management performance maps: steps. Linking real world farm or field level management data to mapped landscape level data involves the following steps (Fig. 2): (i) gather management performance data, (ii) identify performance measures and predictor variables, (iii) evaluate management performance for the proposed interventions, (iv) identify the best management intervention or policy based on currently available data, (v) assess the value of information, (vi) link predictor variables to GIS data layers, (vii) implement interventions, (viii) facilitate and assess technology adoption by farmers. We illustrate steps i through v, while steps vi and vii would be key to achieving outcomes in the field.

Gathering management performance data

To illustrate the process, we use two data sets, the first being data related to potato production in Ecuador, the elements of which are designed to support parameter estimation for a seed degeneration model (Thomas-Sharma et al. 2017). The main components of this model relate to seed health (virus incidence, seed age since certified seed was produced), variety, environmental factors (weather), management (seed propagation and selection) and yield data (Kromann et al. 2017). A single field represented each of the scenarios (treatment combinations) in this data set, so variability within a scenario can only be evaluated at the individual plant level. We focus on yield data as the response as an example, and the intervention of positive selection versus roguing and random seed management strategies. Second, we used published data about seed heath management, and positive selection training and adoption rates in Kenya to explore how information about the likelihood that farmers in a region will adopt a technology can be integrated (Gildemacher et al. 2012).

Identifying performance measures and predictor variables

A variety of methods are useful for selecting predictor variables for performance indicators. We focus on two types of machine learning algorithms: Bayesian networks and recursive partitioning trees (Therneau, Atkinson, and Ripley 2010). These were selected because of their under-appreciated utility in applications such as disease or pest management, and their utility for our problem. As performance indicators, we focused on yield.

Classification and regression trees have been applied in agricultural systems for the purposes of land and soil classification, climate change impact assessment, risk assessment, toxin levels and disease conduciveness for plants (Langemeier et al. 2016; Novak and LaDue 1999; Etter et al. 2006; Caley and Kuhnert 2006; Paul and Munkvold 2004; Tittonell and Giller 2013). These models are known for their ability to identify important variables in classification problems, identify hierarchically the most important predictor variables, and to support decision making processes. Recursive partitioning is a type of decision tree model. The strength of the recursive partitioning method lies in its ability to deal with non-linearity in the data, and depict and interpret the outputs in decision-tree format. A limitation of this method is that it may perform relatively poorly with continuous variables or large numbers of unordered variables. In R there are several packages that perform recursive partitioning, where randomForest, caret, party and rpart are among the most used (Therneau, Atkinson, and Ripley 2010). We used rpart, designed to perform well even when missing data are common.

Bayesian networks have been applied in natural resource management systems for vegetation classification, optimal decision making, disease management and expert elicitation (Geenen and Van Der Gaag 2005; Aguilera et al. 2011; Kristensen and Rasmussen 1997; Perez-Ariza, Nicholson, and Flores 2012). A Bayesian network is a directed, acyclic graph whose nodes represent predictor variables, and links represent dependencies. The relationships between variables are quantified by conditional probability tables. All of these tables together represent the full joint distribution. These models are designed to incorporate learning from both empirical data and from expert assessment, to build conditional probability tables. Important strengths of the Bayesian network method lie in its ability to infer probabilistic relationships between many variables simultaneously. The network structure can be set manually by the user or learned from the data using a variety of algorithms. In the case of exact estimation algorithms, it is possible to set values for any combination of nodes and produce new posterior probabilities for each variable in the network. A limitation of this method is the cost of some of the most advanced Bayesian network software. In addition, mixed data of continuous and categorical data can be problematic for commonly used Bayesian network algorithms. These algorithms tend to work best for cases where all the variables are continuous, or all are discretized. Tools available for Bayesian network analysis include BI-CAMML, Hugin and Netica (Aguilera et al. 2011). R packages include bnlearn, gRain and pcalg (Nagarajan, Scutari, and Lèbre 2013). We selected Netica for this illustration because it is relatively affordable, the algorithms it uses allow for immediate updating of conditional probabilities based on selected levels for variables, it has a powerful graphical interface, and it is widely used in ecological and environmental analyses (Aguilera et al. 2011).

Evaluating management performance

Training in positive selection methods is often useful for familiarizing farmers with disease symptoms and benefits (Gildemacher et al. 2012). We consider data (Kromann et al. 2017) from potato seed production in Ecuador, asking at what locations the benefits of training in positive selection are likely to be greatest. These data (Kromann et al. 2017) focus on potato yield and virus incidence for six viruses, under different management practices, such as positive selection and the number of plantings since disease-free seed was obtained. Using Bayesian network analysis, as described above, the estimated yield with and without positive selection was evaluated for each of three generations of seed at elevations above and below 2895 m a.s.l., the threshold elevation identified in the Bayesian network analysis. Thus, the decision about where to invest in positive selection training might be phrased in terms of (a) uninformed (random) site selection, where each location is equally likely to be selected for training, and (b) informed site selection, where locations above or below 2895 m a.s.l. may be targeted for training, depending on which provides greater benefits, and (c) misinformed site selection, selecting higher or lower elevations based on an incorrect belief about where positive selection will provide greater benefits. For example, there could be a prior belief that a particular pathogen will be more prevalent at lower elevations, due to a higher abundance of vectors, thus making positive selection more important there.

***Yield evaluated in Bayesian networks* -** The benefit of positive selection (in the third year after disease-free seed purchase) was estimated using a Bayesian network in Netica. Netica's Tree-Augmented Naive Bayes (*TAN*) classifier algorithm was used to estimate the conditional probability tables and the network structure. From the conditional probability tables we obtained estimated yields above (7.7 t/ha) and below (3.2 t/ha) 2895 m.a.s.l.

Interacting factors evaluated with recursive partitioning. We can assess scenarios where we have information about interacting predictors, and these are or are not taken into account in the targeting of training. Here we use experimental data about yield under positive selection (as opposed to roguing or random seed selection), at different altitudes and for certified seed that has been propagated for three seasons or less than three seasons.

Considering variety with recursive partitioning. The two most important ware potato varieties in Ecuador, Superchola and Fripapa, were sold at a ratio of 2:1 by volume by growers in a cooperative in Tungurahau, Ecuador (Buddenhagen et al. 2017), and there was a similar ratio of farmers growing each type. For this illustration, let us ignore the other 14 varieties that were grown in this cooperative, as the volumes reported were less than 10%. Some farmers grow both varieties. We consider variety, altitude, and management by estimating mean per plant yields under different conditions, and by using recursive partitioning in rpart. The observed proportion of Ecuadorian farmers using certified seed was reported at 2% (Devaux et al. 2010), though for some organized groups the proportion can be as high as 46% (Buddenhagen et al. 2017). In summary, the reported ratios for this example are as follows: farm altitude (high to low, 51:49), potato variety (Fripapa to Superchola, 33:66) and potato seed age (old to new, 98:2).

Regional differences in adoption of training recommendations. Another example of evaluating management performance, and where management performance is greatest, is based on regional differences in adoption of positive selection after training (Table 1). In one study, adoption varied for three Kenyan counties: Nakuru 46%, Nyadarua 19%, and Narok 18% (Gildemacher et al. 2012). The average benefit of positive selection was 3.4 tons per ha (~USD 350 per ha). This translated to a per-household benefit of $156 per season for a farm of average size for the region. Meanwhile the cost of training was $38 per farmer. In this case, the average value for each region combined gives the expected benefit of $44 where training occurs in a random region.

Applying models to a map of the relevant region

We selected for analysis and extrapolation a major potato growing region stretching from southern Ecuador to southern Colombia. Using data from mapSPAM (HarvestChoice 2014), we identified the region for evaluation by selecting pixels with >200 ha potato production per pixel (where a pixel represents 5-arc minutes, approximately 10,000 ha). Here 51% of potato production is above 2895 m based on MapSpam estimates (You et al. 2012), the elevation cut-off for predicting yield from the Bayesian network analysis.

The data layer of management performance estimates is one important factor for deciding about where to prioritize management efforts. This analysis effectively treats each location as independent from other locations. However, some locations will have more important roles in epidemics than others, due to factors such as environmental conduciveness to disease and position in spatial epidemic networks. We evaluated the layer of management performance estimates for positive selection with a data layer of the potato cropland connectivity risk index, a measure of the likely importance of locations for spatial movement through potato growing areas (Xing et al. 2017).

Estimating the value of information

We assess the value of information for decisions about where to invest management interventions. For the purposes of this study we consider cases where decision makers either have or do not have information about geographic differences in management performance. In the absence of information, they might select any location for management with equal probability. An estimate of the value of information in this case would be the difference in the benefit of investment for locations selected based on the information ("informed location selection"), and the benefit for locations selected randomly ("uninformed location selection"). In the case where decision makers have a prior belief that is not supported by the data, or a misconception, the value of information would be the difference between investments based on the misconception ("misinformed location selection"), versus informed investments. We look at the uninformed and informed management choices related to spatially distributed differences in yield, disease, variety and rates with which best practices are adopted.

We evaluated uncertainty through the lens of how frequently the better management choice would be made (in terms of managing seed degeneration over the three years) based on the data available, as follows. For each of the 28 potential pairwise comparison of management scenarios (each pair of treatment (positive selection versus random) x year (year 1 or 3) x altitude combinations), the difference in yield randomly drawn from the set of observed yields for the treatment combination was collected 10,000 times. (Note that yields were available at the individual plant level, rather than the individual farm level across multiple farms.)

RESULTS

Identifying performance measures and predictor variables, and evaluating management performance

Positive selection and yield for Andean potato. *The Bayesian network analysis* (Fig. 3) showed that, compared to low yield plants, high yielding plants were found in plots more commonly where first generation seed was used, at higher altitudes, with the Fripapa variety, and with lower minimum temperature and higher rainfall six months after planting, as well as low levels of PVX, PLRV, and PVY. The uncertainty is high compared to the observed values. *In the recursive partitioning analysis* (Fig. 4), higher per plant yields were generally obtained from Fripapa (compared to Superchola) in the first two years after certified seed was bought, the highest yields being obtained for altitudes over 3278 m. Meanwhile the highest yields for Superchola were found above 2895 m altitude. If a farmer can afford to replace seed more frequently, and the farm is over 3200 m, Fripapa might be a better choice than Superchola if only a single variety is grown, based solely on yield and assuming the value in the market is comparable (0.29 versus 0.33 cents per kg respectively).

Year 3 plants yield ~300 g per plant less on average, and the benefits of positive selection allow yields to approach the average yields for recently purchased certified seed (if values from cycle 1 and 2 are combined). Since there are no differences with respect to variety in year/cycle 3 after certified seed it would seem positive selection is equally valuable in both varieties for seed that has gone through more than a two planting cycles.

The price farmers obtain for each variety was estimated to be \$0.33 USD per kg for Superchola and \$0.29 USD per kg for Fripapa (Navarrete, perssonal observation). The variety Superchola has entered into the common parlance and is widely regarded as superior, even though there is some evidence that consumers cannot easily recognize the difference between ware potato of it and other varieties (Kromann, personal observation).

Positive selection and yield for Kenyan potato. This analysis was based on the probability of adoption of positive selection, where higher adoption rates result in a higher payoff for investment. Adoption rates were 46, 19 and 18% in three regions (Table 1) (Gildemacher et al. 2012, 2011). Thus, based on this measure alone, selection of the region with better training uptake would double the benefits obtained of a training intervention focused only in the other areas.

Applying models to a map of the relevant region and integrating data layers

The mapped estimates of the management performance of positive selection for Andean potato yield (Fig. 5A) and the locations where cropland connectivity risk was highest (Fig. 5B) could be combined to identify locations both independently likely to be successful and of importance for regional management.

VOI for targeted implementation of positive selection (yield as response, Bayesian networks to identify predictors) For equivalent farm sizes above and below 2895 m, we can estimate the benefit of training under uninformed (random) site selection by using the weighted mean of the benefit above and below 2895 m.a.s.l., which is 6.5 t/ha (0.51 * 7.7 + 0.49 * 3.2 = 6.5).

The estimated benefit under informed site selection, selecting locations above 2895 m, is 7.7 t/ha – a difference of 1.2 t/ha from random site selection. Under misinformed site selection, if the assumption was that positive selection provides more benefits at low altitude due to greater pathogen load, then the benefit is 3.2 tons per ha, 4.5 tons per ha less than the optimal allocation, and 3.3 tons per ha less than the uninformed (random) site selection option.

VOI for targeted implementation of positive selection (yield as response, rpart to identify predictors) Assuming that positive training targeted farmers randomly with respect to the observed frequency of the variables, the weighted mean benefit of positive selection would be 8.7 tons per ha. Preferentially targeting sites at high altitude (but sampling randomly with respect to seed age and variety) provides higher benefits of 9 tons per ha, otherwise targeting low altitude sites provides lower returns at 8.3 tons per ha. Targeting farmers who plant old seed (3 years since certification) provides little benefit: 8.8 tons per ha compared to random targeting of farmers (under the scenario where use of certified seed is rare at 2%), but targeting the 2% that do use new seed provides a benefit of 3.1 tons per ha. By far the greatest benefit is provided by targeting farmers that grow Fripapa (benefit of 10.9 tons per ha) as opposed to Superchola (benefit of 7.6 tons per ha).

Regional differences in adoption of training recommendations in Kenya.
In the example of Kenyan potato seed technology adoption, targeted selection of high adoption rate areas for training (where the per-farmer benefit was $72) would increase the return by $28 per farmer trained. Realized benefits would vary depending on farm sizes for the targeted farmers.

DISCUSSION

The value of information was assessed by comparing (a) allocating development resources (with the example of positive selection training for farmers) without regard to the observed site characteristics, and (b) targeted allocation of training resources toward sites with known characteristics. In our example, the data (and expert expectation) show that positive selection is often an effective way to reduce seed degeneration. An NGO or government extension agency could implement a rural development intervention where farmers are trained to use positive selection without regard to their specific farm conditions, variety used or the frequency with which they can buy improved seed. However we showed that this option is not optimal as a means of obtaining higher yield, and lower disease incidence. We compared uninformed and informed allocation of resources, to assess the value of the information used for targeting interventions. In the simplest scenario, using results from Bayesian networks, we showed that the benefit of positive selection was highest (4.5 tons per ha) at high altitudes, and uninformed allocation of farmer training would provide a net benefit of 1.2 tons per ha less than targeted training.

Using data about adoption rates from Kenya, we showed that unless adoption rates were higher than 24%, the first-year benefit per household would not exceed the $38 per farmer cost of training (though presumably the benefits would continue to accrue in subsequent years). Also random allocation of training effort would only yield a $44 dollar benefit (over the cost of the training) per household. Gildemacher et al. (2017) also point out that adoption rates were lower in drought years, suggesting that prediction of adoption rates could be difficult if based on regional patterns in a single year. It is easy to imagine scenarios where observed adoption rates (say in a training scheme) would vary in predictable ways based on in season weather conditions, language spoken, literacy, cultural differences between trainer and trainee, wealth or other factors. Further spatial data could be available for key predictor variables, and could form a part of selection criteria for farmer training initiatives (and the approach to the training could be altered to improve adoption rates).

Then in a more complex scenario, we showed that targeting farmers growing the variety Fripapa, or at high altitude was the optimal strategy. Not examined was the size of the farms, which would logically determine the returns on training efforts per farmer. Over and above the availability of data, the value of information in this scenario requires that the data be linked to real world distributions of farm or farmer characteristics and behavior (we used altitude, variety use, rate of seed replacement).

We showed that recursive partitioning and Bayesian networks provide easily interpreted graphics and estimates of the predictor variables related to yield in seed degeneration studies. We anticipate that these tools will be useful in other studies examining disease incidence or yield and predictor variables related to management or environmental conditions.

Combining data layers for evaluating optimal intervention strategies can provide more insight, along with challenges due to uncertainty. Evaluating the risk of disease due to cropland connectivity (Xing et al. 2017) in combination with independent location traits can position the analysis in the larger context of disease management for the region. A broader systems analysis – for example, impact network analysis (Garrett 2018; Garrett et al. 2018), which integrates across management performance, socioeconomic networks, and biophysical networks such as epidemic networks – can aid in identifying intervention locations that prioritize across multiple goals.

Management performance mapping is potentially applicable to any problem in intervention agriculture or ecology with measurable outcomes. Yield gap analyses that incorporate maps can address some of the same goals as management performance mapping (Schulthess et al. 2013; Silva et al. 2017; Lobell, Cassman, and Field 2009; Lobell et al. 2015; van Ittersum et al. 2016; Grassini et al. 2015; van Bussel et al. 2015b). For example, yield gap analysis attempts to identify the most important factors that influence yield, and are controllable. The focus of management performance mapping, however, is on providing spatial information about intervention impact of management options. Management performance maps would ideally incorporate and account for interacting human dimensions (e.g., learning, financial liquidity, capital, institutions) and biophysical aspects of crop production (Arneth et al., 2014)

To make maps, management performance measures are modelled as a function of predictor variables available in geographic data layers, e.g., climate variables, altitude, or soil fertility. Useful outcome variables include yield, disease incidence, measures of crop quality, technology adoption levels, or measures of food security. As in species distribution modeling, data to inform management performance mapping may come from experiments, surveys or monitoring of management carried out at any scale. Species distribution models for crops can, themselves, be a useful predictor variable for evaluating the importance of management of invasive species. Worldwide crop distribution models are valuable for defining the relevant locations for consideration of management options (You et al. 2014; Monfreda, Ramankutty, and Foley 2008). Alternatively, there could be cases where maps are derived from crop distributions are well understood via more direct means such as remote sensing, surveys or farm level measurements (e.g., Lobell et al. 2015). The extent and depth of the reference data, the magnitude of management effect sizes, and the degree to which geographic extrapolation is appropriate will determine the level of confidence decision makers have in management performance maps.

Our example decision, deciding where to implement training for improved disease management, represents a class of decisions where there is confidence that the activity will provide a benefit. Management performance mapping is applied to guide implementation to locations where there is some evidence that the benefit will be greater than in other locations. For this class of decisions, the risk is low that limited data is “worse than no data at all”. In the language of hypothesis testing, there is not a strong motivation to avoid Type I error (rejecting a null hypothesis when the null hypothesis is true), because a Type II error (failing to reject a null hypothesis when the null hypothesis is false) is arguably just as bad. In the management performance mapping context, the null hypothesis is that the benefit of implementation will be the same in all locations. The main risk of “bad data” would be from data with a strong bias that would lead to misinformed decisions. The cost of “bad data” may also go up if the logistical costs (of transport, communications, etc.) of targeting locations incorrectly identified is higher than targeting locations at random or selecting locations based on convenience.

We addressed management performance mapping with performance defined in terms of the mean performance observed. Other potential criteria for selecting regions for investment might emphasize different priorities (Table 2). Going forward with applying management performance mapping, it will be important to consider not only the value of information under a reasonable set of assumptions, but also the role of uncertainty. One of the applications of VOI analysis is to determine whether collecting more or better data about management performance is justified, not just for the sake of more statistical power but because the information improves farmer decision-making under a realistic range of conditions. Scientists, funders and policy makers will need to evaluate whether decreasing uncertainty is likely to lead to shifts in the mean of the performance measures. There may be little value in collecting more evidence about management performance if the mean is little influenced. Estimates of uncertainty were obtained from the Bayesian and recursive partitioning methods we used, but we emphasized the differences in the estimated value of the management rather than the spread of the uncertainty. In our example data, our only estimate of uncertainty within a scenario was based on variability among individual plants, while a person making decisions about regional priorities would strongly prefer to have information about farm-to-farm variability within each scenario. However, emphasizing mean differences in management outcomes could be justified, in general, even if uncertainty is high, particularly if there is reason to believe more data will not lead to major shifts in the ranking of mean management performance. Management performance mapping provides a process to extrapolate from available data to make evidence-based decisions about where to invest in disease and crop management or training initiatives.

ACKNOWLEDGMENTS

We appreciate support by the CGIAR Research Program for Roots, Tubers and Bananas, the CGIAR Research Program on Climate Change and Food Security (CCAFS), USDA NIFA grant 2015-51181-24257, the USAID Feed the Future Haiti Appui à la Recherche et au Développement Agricole (AREA) project, The Ceres Trust, NCR SARE Research and Education Grant LNC13-355, USDA APHIS grant 11–8453–1483-CA, US NSF Grant EF-0525712 as part of the joint NSF-NIH Ecology of Infectious Disease program, US NSF Grant DEB-0516046, and the University of Florida.

LITERATURE CITED

1. Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., and Salmerón, A. 2011. Bayesian networks in environmental modelling. Environmental Modelling & Software. 26:1376–1388.
2. Aguirre-Gutiérrez, J., van Treuren, R., Hoekstra, R., and van Hintum, T. J. L. 2017. Crop wild relatives range shifts and conservation in Europe under climate change. Diversity and Distributions. 23:739–750.
3. Almekinders C. J. M., Louwaars N. P., and De Bruijn G. H. 1994. Local seed systems and their importance for an improved seed supply in developing countries. Euphytica. 78:207–216.
4. Altieri, M. A., and Nicholls, C. I. 2008. Scaling up Agroecological Approaches for Food Sovereignty in Latin America. Development. 51:472–480.
5. Arneth, A., Brown, C., and Rounsevell, M. D. A. 2014. Global models of human decision-making for land-based mitigation and adaptation assessment. Nature Climate Change. 4:550–557.
6. Asante Bright Owusu, Otoo Emmanuel, Wiredu Alexander Nana, Acheampong Patricia, Osei-Adu Jonas, and Nsiah-Frimpong Benedicta. 2011. Willingness to adopt the vine multiplication technique in seed yam production in the forest savanna transition agro-ecological zone, Ghana. Journal of Development and Agricultural Economics. 3 Available at: http://www.academicjournals.org/JDAE/abstracts/abstracts/abstracts2011/Dec/26%20Dec/Asante%20et%20al.htm [Accessed December 9, 2013].

7. Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling. 200:1–19.
8. Bentley, J. W., and Vasques, D. 1998. The seed potato system in Bolivia: organisational growth and missing links. Agricultural Research and Extension Network. :1–12.
9. Bertschinger, L., Bühler, L., Dupuis, B., Duffy, B., Gessler, C., Forbes, G. A., et al. 2017. Incomplete Infection of Secondarily Infected Potato Plants–an Environment Dependent Underestimated Mechanism in Plant Virology. Frontiers in Plant Science. 8:74.
10. Breidert, C., Hahsler, M., and Reutterer, T. 2006. A review of methods for measuring willingness-to-pay. Innovative Marketing. 2:8–32.
11. Buddenhagen, C. E., Hernandez Nopsa, J. F., Andersen, K. F., Andrade-Piedra, J., Forbes, G. A., Kromann, P., et al. 2017. Epidemic Network Analysis for Mitigation of Invasive Pathogens in Seed Systems: Potato in Ecuador. Phytopathology. 107:1209–1218.
12. van Bussel, L. G. J., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H., et al. 2015a. From field to atlas: Upscaling of location-specific yield gap estimates. Field Crops Research. 177:98–108.
13. van Bussel, L. G. J., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H., et al. 2015b. From field to atlas: Upscaling of location-specific yield gap estimates. Field Crops Research. 177:98–108.
14. Caley, P., and Kuhnert, P. M. 2006. Application and evaluation of classification trees for screening unwanted plants. Austral Ecology. 31:647–655.
15. Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J. J., Southwell, D. M., Armstrong, D. P., Chadès, I., et al. 2015. When do we need more data? A primer on calculating the value of information for applied ecologists ed. Olivier Gimenez. Methods in Ecology and Evolution. 6:1219–1228.
16. Choudhury, R. A., Garrett, K. A., Klosterman, S. J., Subbarao, K. V., and McRoberts, N. 2017. A Framework for Optimizing Phytosanitary Thresholds in Seed Systems. Phytopathology. 107:1219–1228.
17. Condori, B., Hijmans, R. J., Ledent, J. F., and Quiroz, R. 2014. Managing Potato Biodiversity to Cope with Frost Risk in the High Andes: A Modeling Perspective ed. Ji-Hong Liu. PLoS ONE. 9:e81510.
18. Danthine, J.-P. 1978. Information, futures prices, and stabilizing speculation. Journal of Economic Theory. 17:79–98.
19. Devaux, A., Kromann, P., and Ortiz, O. 2014. Potatoes for sustainable global food security. Potato Research. 57:185–199.
20. Devaux, A., Ordinola, M. E., Hibon, A., and Flores, F. A. 2010. *El sector papa en la región andina: Diagnóstico y elementos para una visión estratégica (Bolivia, Ecuador y Perú).* International Potato Center.
21. Etter, A., McAlpine, C., Wilson, K., Phinn, S., and Possingham, H. 2006. Regional patterns of agricultural land use and deforestation in Colombia. Agriculture, Ecosystems & Environment. 114:369–386.
22. Garrett, K. A. 2018. Impact network analysis: evaluating the success of interventions. PeerJ Preprints 6:e27037v1.
23. Garrett, K. A., Alcalá-Briseño, R. I., Andersen, K. F., Buddenhagen, C. E., Choudhury, R. A., Fulton, J. C., Hernandez Nopsa, J. F., Poudel, R., and Xing., Y. 2018. Network analysis: A systems framework to address grand challenges in plant pathology. Annual Review of Phytopathology 56:559-580.

24. Geenen, P. L., and Van Der Gaag, L. C. 2005. Developing a Bayesian network for clinical diagnosis in veterinary medicine: from the individual to the herd. In *Proceedings of the Third Bayesian Modelling Applications Workshop; Edinburgh*, Citeseer.
25. Gildemacher, P. R., Demo, P., Barker, I., Kaguongo, W., Woldegiorgis, G., Wagoire, W. W., et al. 2009. A description of seed potato systems in Kenya, Uganda and Ethiopia. American Journal of Potato Research. 86:373–382.
26. Gildemacher, P. R., Leeuwis, C., Demo, P., Borus, D., Schulte-Geldermann, E., Mundia, P., et al. 2012. Positive selection in seed potato production in Kenya as a case of successful research-led innovation. International Journal of Technology Management & Sustainable Development. 11:67–92.
27. Gildemacher, P. R., Schulte-Geldermann, E., Borus, D., Demo, P., Kinyae, P., Mundia, P., et al. 2011. Seed Potato Quality Improvement through Positive Selection by Smallholder Farmers in Kenya. Potato Research. 54:253–266.
28. Grassini, P., van Bussel, L. G. J., Van Wart, J., Wolf, J., Claessens, L., Yang, H., et al. 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research. 177:49–63.
29. Hanemann, W. M. 1991. Willingness to pay and willingness to accept: how much can they differ? The American Economic Review. 81:635–647.
30. HarvestChoice. 2014. *Crop Production: SPAM*. Washington, DC and St. Paul, MN: International Food Policy Research Institute & University of Minnesota. Available at: http://harvestchoice.org/node/9716.
31. Hijmans, R. J., and Graham, C. H. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology. 12:2272–2281.
32. Hirshleifer, J., and Riley, J. G. 1979. The Analytics of Uncertainty and Information-An Expository Survey. Journal of Economic Literature. 17:1375–1421.
33. Iritani, W. M. 1968. Factors affecting physiological aging (degeneration) of potato tubers used as seed. American Journal of Potato Research. 45:111–116.
34. van Ittersum, M. K., van Bussel, L. G. J., Wolf, J., Grassini, P., van Wart, J., Guilpart, N., et al. 2016. Can sub-Saharan Africa feed itself? Proceedings of the National Academy of Sciences. 113:14964–14969.
35. Jaffee, S., and Srivastava, J. 1994. The roles of the private and public sectors in enhancing the performance of seed systems. The World Bank Research Observer. 9:97–117.
36. Johnson, F. A., Smith, B. J., Bonneau, M., Martin, J., Romagosa, C., Mazzotti, F., et al. 2017. Expert Elicitation, Uncertainty, and the Value of Information in Controlling Invasive Species. Ecological Economics. 137:83–90.
37. Kawakami, K. 1962. The physiological degeneration of potato seed tubers and its control. Potato Research. 5:40–49.
38. Kristensen, K., and Rasmussen, I. A. 1997. The use os a Bayesian network to construct a decision support system for mechanical weed control in malting barley. In *Computer Modelling in Weed Science*, ed. J E Jensen. , p. 30–32.
39. Kromann, P., Andrade-Piedra, J. L., Navarrete, I., Taipe, A., and Gómez, J. 2017. Dataset for: Potato seed degeneration in Ecuador. International Potato Center Dataverse, V2. Available at: http://dx.doi.org/10.21223/P3/3CT90C [Accessed December 1, 2017].
40. Kromann, P., Montesdeoca, F., and Jorge Andrade-Piedra. 2016. Integrating formal and informal potato seed systems in Ecuador. In *Case Studies of Roots, Tubers and Bananas Seed Systems.*, eds. J. L. Andrade-Piedra, J. W. Bentley, C. Almekinders, K. Jacobsen, S. Walsh,

and G. Thiele. Lima, Peru: CGIAR Research Program on Roots, Tubers and Bananas (RTB), p. 14–32.

41. Kromann, P., Taipe, A., Perez, W. G., and Forbes, G. A. 2009. Rainfall thresholds as support for timing fungicide applications in the control of potato late blight in Ecuador and Peru. Plant Disease. 93:142–148.
42. Langemeier, C. B., Robertson, A. E., Wang, D., Jackson-Ziems, T. A., and Kruger, G. R. 2016. Factors Affecting the Development and Severity of Goss's Bacterial Wilt and Leaf Blight of Corn, Caused by Clavibacter michiganensis subsp. nebraskensis. Plant Disease. 101:54–61.
43. Lave, L. B. 1963. The Value of Better Weather Information to the Raisin Industry. Econometrica. 31:151–164.
44. Lobell, D. B., Cassman, K. G., and Field, C. B. 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. Annual Review of Environment and Resources. 34:179–204.
45. Lobell, D. B., Thau, D., Seifert, C., Engle, E., and Little, B. 2015. A scalable satellite-based crop yield mapper. Remote Sensing of Environment. 164:324–333.
46. Louwaars, N. P., de Boef, W. S., and Edeme, J. 2013. Integrated Seed Sector Development in Africa: A Basis for Seed Policy and Law. Journal of Crop Improvement. 27:186–214.
47. Macauley, M. K. 2006. The value of information: Measuring the contribution of space-derived earth science data to resource management. Space Policy. 22:274–282.
48. McGuire, S., and Sperling, L. 2013. Making seed systems more resilient to stress. Global Environmental Change. 23:644–653.
49. McGuire, S., and Sperling, L. 2016. Seed systems smallholder farmers use. Food Security. 8:179–195.
50. McQuaid, C. F., Sseruwagi, P., Pariyo, A., and van den Bosch, F. 2016. Cassava brown streak disease and the sustainability of a clean seed system. Plant Pathology. 65:299–309.
51. Monfreda, C., Ramankutty, N., and Foley, J. A. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochemical Cycles. 22:n/a-n/a.
52. Nagarajan, R., Scutari, M., and Lèbre, S. 2013. *Bayesian networks in R.*
53. Navarrete, I., Panchi, N., Kromann, P., Forbes, G., and Andrade-Piedra, J. 2017. Health quality of seed potato and yield losses in Ecuador. bioRxiv. :108712.
54. Novak, M. P., and LaDue, E. 1999. Application of Recursive Partitioning to Agricultural Credit Scoring. Journal of Agricultural and Applied Economics. 31:109–122.
55. Paul, P. A., and Munkvold, G. P. 2004. A model-based approach to preplanting risk assessment for gray leaf spot of maize. Phytopathology. 94:1350–1357.
56. Perez-Ariza, C. B., Nicholson, A. E., and Flores, M. J. 2012. Prediction of coffee rust disease using bayesian networks. In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models*, , p. 259–266.
57. Schulthess, U., Timsina, J., Herrera, J. M., and McDonald, A. 2013. Mapping field-scale yield gaps for maize: An example from Bangladesh. Field Crops Research. 143:151–156.
58. Sharma, K. K., and Kang, G. S. 2003. Appraisal of rate of degeneration in some potato cultivars. Potato Journal. 30:123–124.
59. Silva, J. V., Reidsma, P., Laborte, A. G., and van Ittersum, M. K. 2017. Explaining rice yields and yield gaps in Central Luzon, Philippines: An application of stochastic frontier analysis and crop modelling. European Journal of Agronomy. 82:223–241.

60. Sparks, A. H., Forbes, G. A., Hijmans, R. J., and Garrett, K. A. 2014. Climate change may have limited effect on global risk of potato late blight. Global Change Biology. 20:3621–3631.
61. Sperling, L. 2008. *When Disaster Strikes: a guide to assessing seed system security*. CIAT.
62. Sperling, L., Ortiz, O., and Thiele, G. 2013. *Seed systems. Conceptual frameworks for guiding practical interventions.* ed. CGIAR. CGIAR.
63. Sutherst, R. 1985. A computerised system for matching climates in ecology. Agriculture, Ecosystems & Environment. 13:281–299.
64. Therneau, T. M., Atkinson, B., and Ripley, B. 2010. rpart: Recursive partitioning. R package version. 3:1–46.
65. Thomas-Sharma, S., Abdurahman, A., Ali, S., Andrade-Piedra, J. L., Bao, S., Charkowski, A. O., et al. 2016. Seed degeneration in potato: the need for an integrated seed health strategy to mitigate the problem in developing countries. Plant Pathology. 65:3–16.
66. Thomas-Sharma, S., Andrade-Piedra, J., Carvajal Yepes, M., Hernandez Nopsa, J., Jeger, M. J., Jones, R., et al. 2017. A risk assessment framework for seed degeneration: Informing an integrated seed health strategy for vegetatively-propagated crops. Phytopathology. 107:1123–1135.
67. Tittonell, P., and Giller, K. E. 2013. When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. Field Crops Research. 143:76–90.
68. van Wart, J., van Bussel, L. G. J., Wolf, J., Licker, R., Grassini, P., Nelson, A., et al. 2013. Use of agro-climatic zones to upscale simulated crop yield potential. Field Crops Research. 143:44–55.
69. Wilson, E. C. F. 2015. A Practical Guide to Value of Information Analysis. PharmacoEconomics. 33:105–121.
70. Xing, Y. et al., 2017. Global cropland connectivity: A risk factor for invasion and saturation by emerging pathogens and pests. bioRxiv: https://doi.org/10.1101/106542.
71. You, L., Crespo, S., Guo, Z., Koo, J., Ojo, W., Sebastian, K., et al. 2012. MapSpaM - Home of the Spatial Production Allocation Model. Available at: http://mapspam.info/ [Accessed December 15, 2017].
72. You, L., Wood, S., Wood-Sichra, U., and Wu, W. 2014. Generating global crop distribution maps: From census to grid. Agricultural Systems. 127:53–60.

Table 1. Regional adoption rates after positive selection training in Kenya and the expected realized benefit of training given the adoption rate. The average benefit is that expected under random allocation of training effort to the regions without regard to adoption rates.

Region	Observed adoption rate	Per household benefit $USD	Expected realized benefit of training in year 1
Nakura	0.46	156	72
Nyadarua	0.19	156	30
Narok	0.18	156	28
		Average Benefit	44

Table 2. Potential criteria for identifying priority sites for interventions (such as training farmers to use positive selection for improved seed health).

Criterion	**Rationale**
Regions where expected absolute benefit is greatest	Greatest benefit to regional food production
Regions where expected proportional gain is greatest	Greatest benefit to regional farmers
Regions where outcomes before intervention are lowest	Benefit to regions in greatest need
Regions where outcomes before intervention are highest	Benefit to regions currently best adapted for production
Incorporating measures of uncertainty at the farm level	
Regions where the 5^{th} percentile benefit is greatest	Consistent benefit across farmers
Regions where the trimmed mean is greatest	Greatest benefit for typical farmers

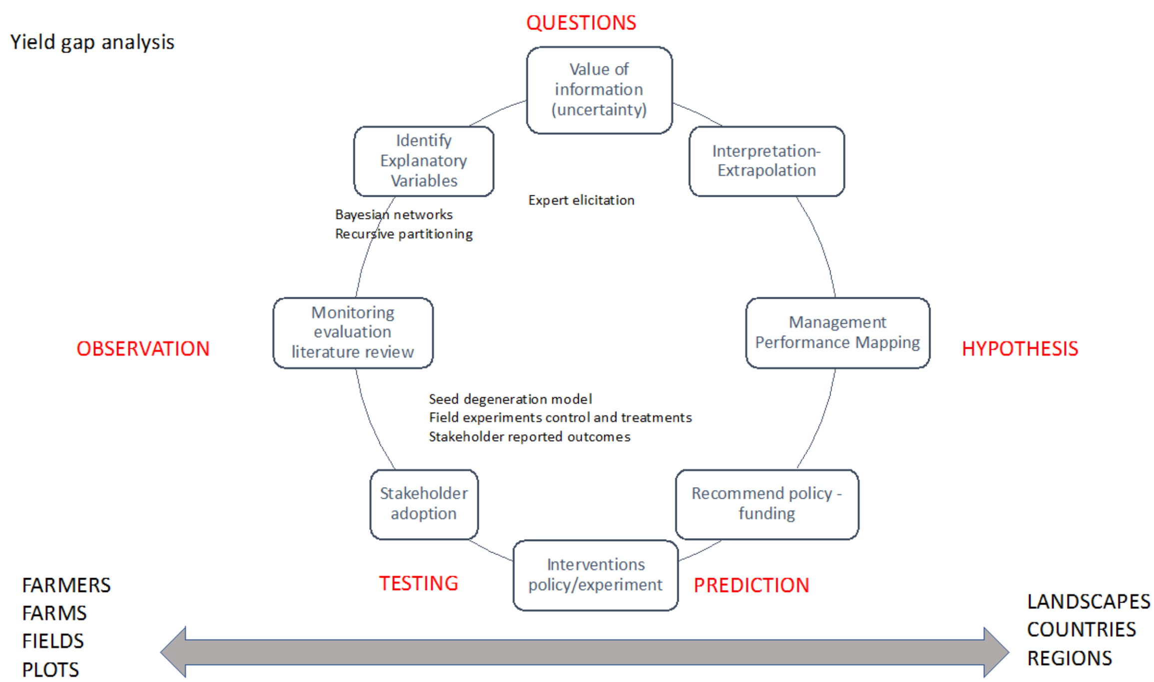

Fig 1. Development efforts ideally take place in a culture of continuous improvement, based on continuous monitoring and evaluation, and incorporating experimentation to facilitate adaptive management. The cycle described here closely resembles the learning cycle and the scientific process, in general. Management performance mapping operates in this context by scaling up field, farm, and plot derived information to larger scale landscapes, regions or countries.

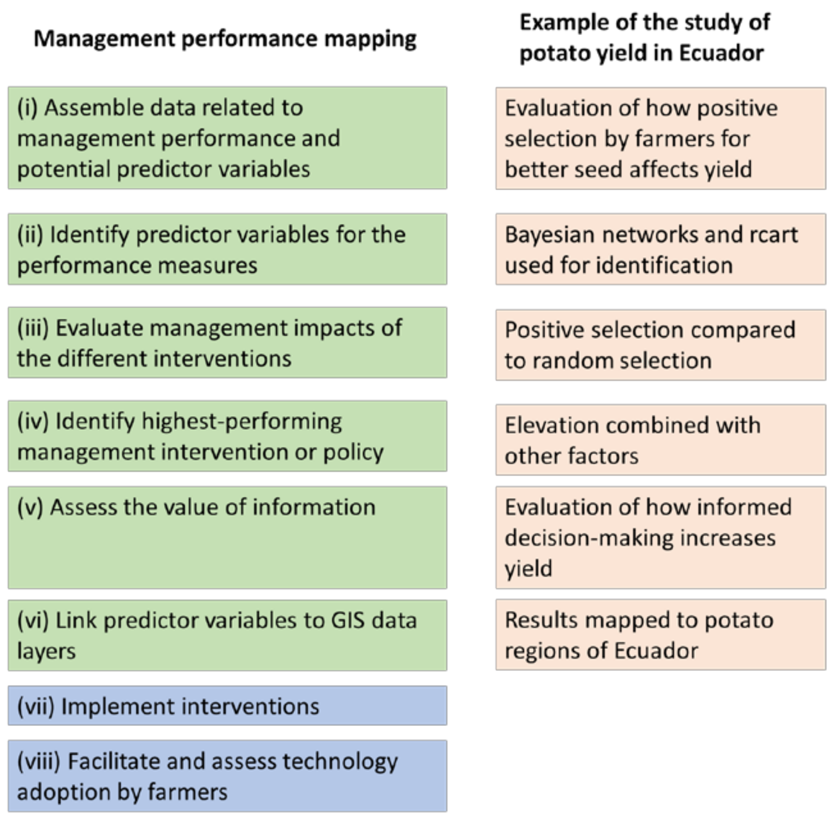

Figure 2. The steps in a management performance mapping "pipeline", illustrated for the use of positive selection for seed potato in Ecuador.

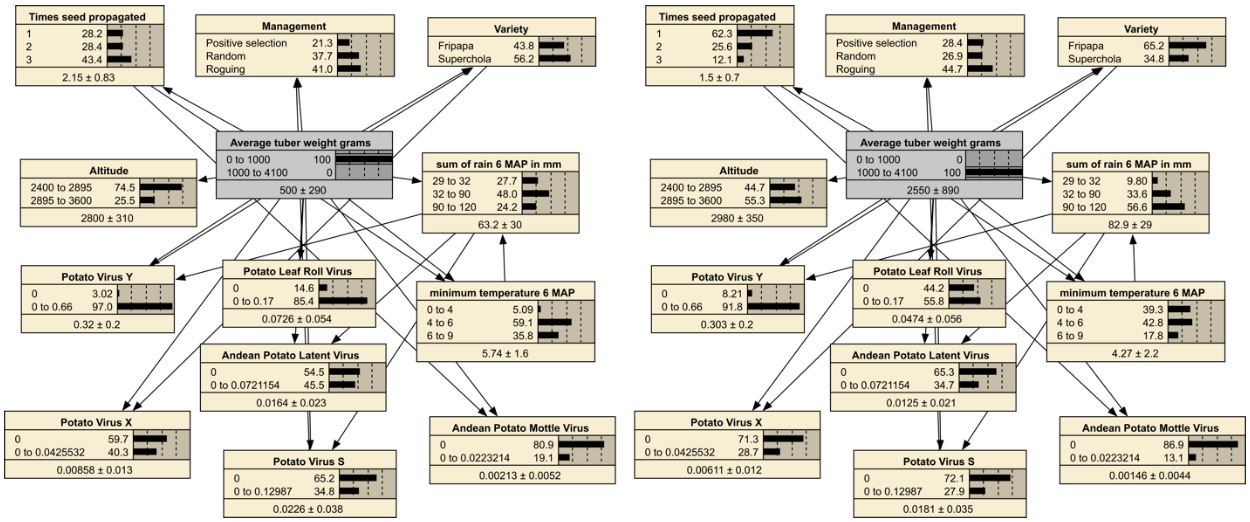

Figure 3. Ecuadorian potato yield and the factors associated with yield from a Bayesian network analysis in Netica, where the impact of selected conditions can be seen for plants with low yield (left) and high yield (right) and the posterior probabilities in other nodes is returned. The lower text for each node gives the estimated mean and uncertainty.

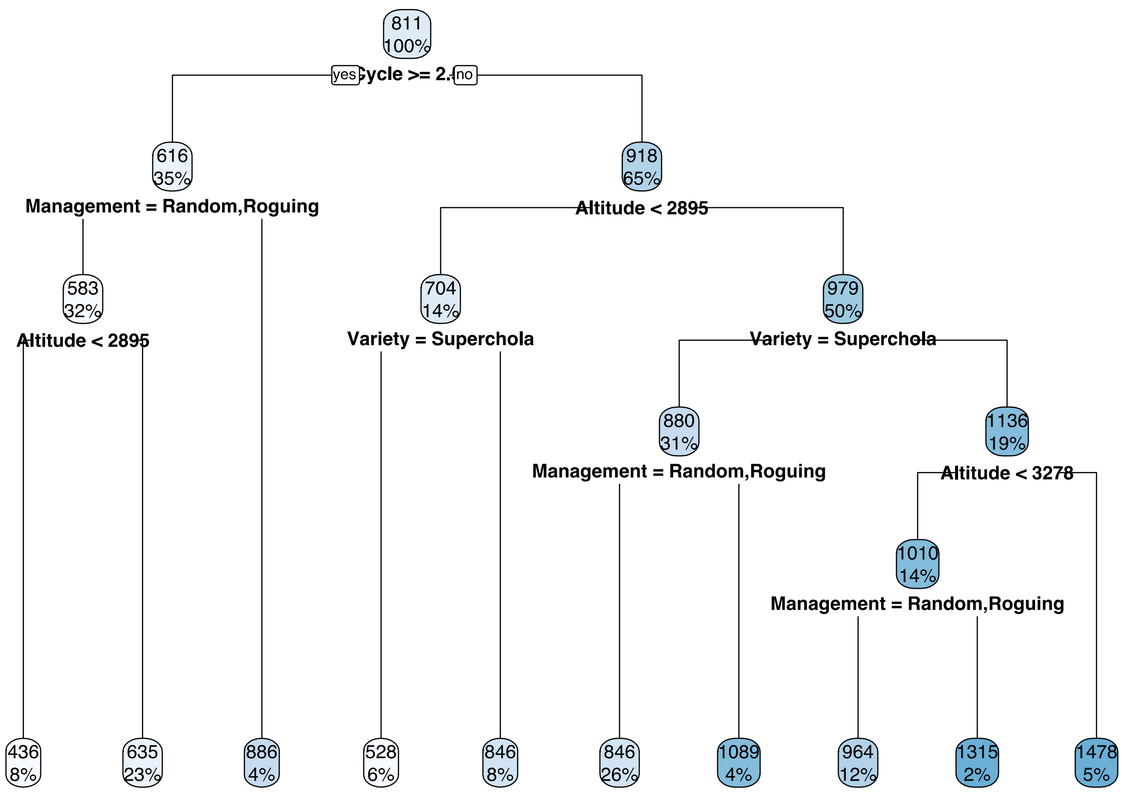

Fig. 4. Recursive partitioning results in decision tree format, with per plant yield (g) as the dependent variable. Branches to the left are results when the logical statements at the nodes are true, and branches to the right are results when the logical statements are false. The upper numbers in the boxes are the mean yields for that condition, and the percent values are the proportion of the data for which the condition applies.

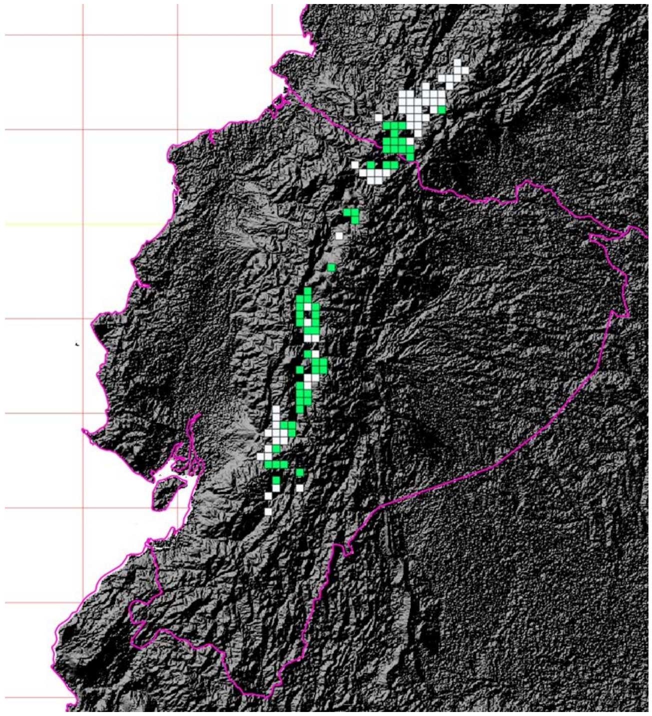

Figure 5A. Ecuador and surroundings, with potato production indicated based on MapSpam estimates. Pixels above and below 2895 m.a.s.l. are indicated in green and white, where pixels are included if the harvested area estimate is greater than 200 ha. We find that 51% of the harvested is above this threshold. The country border is for Ecuador, and the graticules are 1 degree squares.

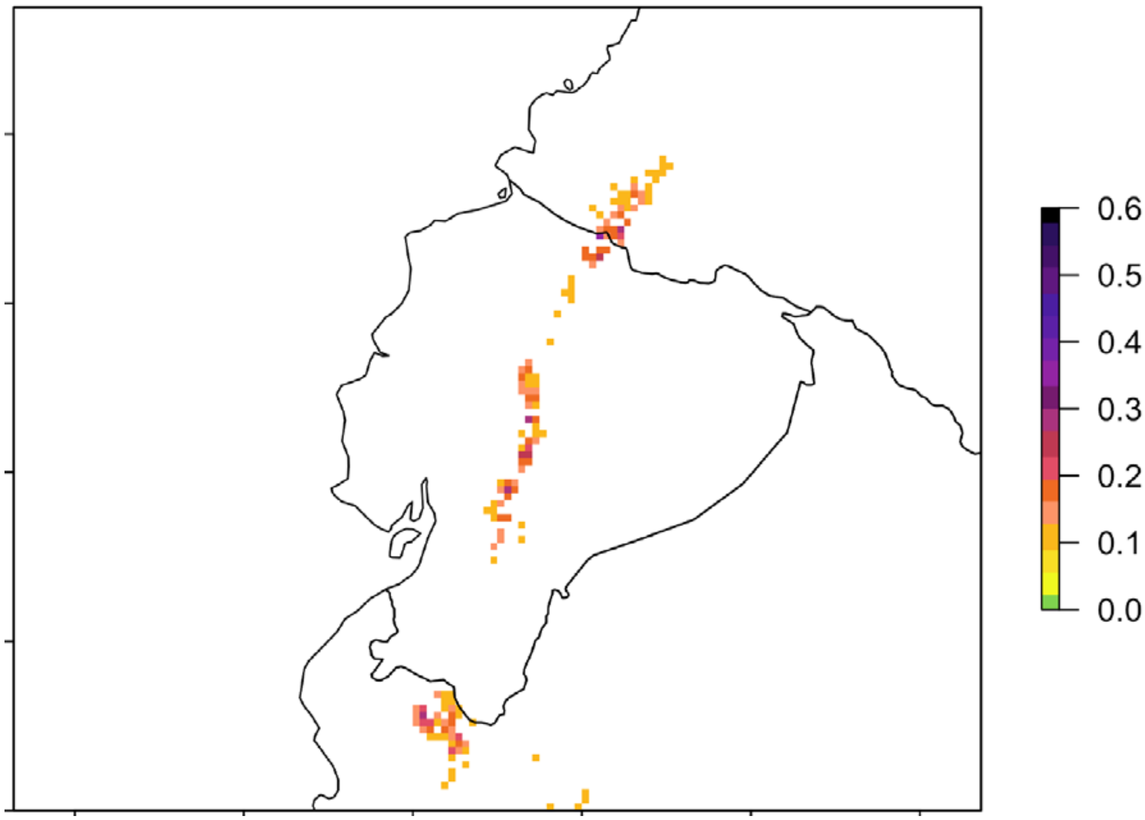

Figure 5B. The potato cropland connectivity risk index estimated for Ecuador and southern Colombia, based on the mean from uncertainty quantification across multiple parameter combinations (Xing et al., 2017).