

Resource frontiers and agglomeration economies: The varied logics of land-based investing in Southern and Eastern Africa

Dilini Abeygunawardane ^{a, b, *}, Angela Kronenburg García ^{a, c, d}, Zhanli Sun ^b, Daniel Müller ^{b, e, f}, Almeida Siteo ^c, Patrick Meyfroidt ^{a, g}

^a Earth and Life Institute, Georges Lemaître Centre for Earth and Climate Research (TECLIM), UCLouvain, 1348 Louvain-la-Neuve, Belgium

^b Leibniz Institute of Agricultural Development in Transition Economies (IAMO), 06120 Halle (Saale), Germany

^c Department of Forest Engineering, Universidade Eduardo Mondlane, 1102 Maputo, Mozambique

^d Department of Historical and Geographic Sciences and the Ancient World, University of Padua, 35141 Padova, Italy

^e Geography Department, Humboldt-Universität Zu Berlin, 10099 Berlin, Germany

^f Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität Zu Berlin, 10099 Berlin, Germany

^g FNRS, Fonds de la Recherche Scientifique, 1000 Brussels, Belgium

* abeygunawardane@iamo.de

Abstract

Investment and market conditions across much of Africa are unfavorable. With scarce empirical evidence, the peculiar choice of investing in land in the African frontiers has been the subject of many conjectures. We address this gap by reconstructing the underlying logics of investment location choices in a Bayesian network with firm and actor level interview and spatial data from 37 transnational agriculture and forestry investments across 121 farm and plantation sites in Mozambique, Zambia, Tanzania, and Ethiopia. Drawing from rent theory, we characterized the appeal of investment locations in terms of resource frontiers (i.e., where land resources are abundant relative to other inputs) and agglomeration economies (i.e., where scale economies external to the individual firm exist). Our sample captured a diversity of investments, made up of investors with varying farming or forestry and regional skill sets and product market control, resulting in varied investment priorities and location choices. Experienced investors focusing on high-value crops invested preferentially in remote subsistence frontiers, seeking land with specific agroecological conditions. In contrast, newcomers – such as the typical speculators - focused more on proximity to infrastructures and markets. We identified four comparative advantages of the African frontiers: large tracts of agroecologically suitable land; agroecologically unique land, potential market access, and location advantages conditional on investor track records. Development efforts founded upon sustainable land-based investments would benefit from targeting investors who can survive unfavorable investment contexts to induce scale economies that extend beyond individual firms.

Introduction

African frontiers in the dry woodland and savannah biomes have not received as much scholarly attention as their counterparts have. The region, where smallholder agriculture has been the key driver of deforestation (1, 2), was not known for those historical generalizations of frontier development drawn from Latin America and Southeast Asia. High deforestation rates, state-sponsored land settlement schemes, and industrialized agriculture have been sparse (3). However, the global land-based investments wave that started in the early 2000s and boomed later in the decade around the food and financial crisis marked a renewed interest in African frontiers. Despite the decline in land-based investments in recent years (4, 5), the effects of the investments on social orders and property systems in Africa has been profound (6, 7). The scale of land acquisitions, some involving thousands of hectares, their speculative nature (8-10), and numerous failures in consulting and compensating people (11, 12) have caused suspicion over the potential benefits (9, 13, 14). Consequently, region's large-scale commercial agriculture is at normative crossroads. While there are economic and environmental gains to be had through the positive spillovers of increased productivity (15) and the potential conservation benefits of land-efficient farming (16-18), balancing the trade-offs has remained a challenge. Hampered by incomplete and mostly inconsistent data (19) and preconceptions thinly informed by science (20), the understanding of land-based investments in Sub-Saharan Africa, their rationales, poor implementation, and high failure rates have remained blurry. Addressing this gap, we contribute toward understanding where investments are likely to occur, the reasoning and logics behind these investments, and how these vary across heterogeneous investors in understudied frontiers in Southern and Eastern Africa. We combine firm and actor level interview and spatial data to analyze investors' take on where they invest, in what, and why, using an inductive grounded theory approach (21).

Informed opinions on what attracts investors to Africa, whether it is cheap land, cheap labor, or weak regulations, abound. Empirical evidence is however, scarce. Much of the evidence relies heavily on country-level studies investigating the determinants of foreign direct investments (FDI) in the industry and services sectors, and points to the usual suspects including market proximity, human capital (22), bilateral ties, governance (23-28), and corporate taxes (29). Studies that specifically focus on land-based investments have found agroecological conditions, tenure security, land availability, investor protection, investor-host distance, colonial ties, and common languages to have an effect on location choices (30, 31). However, these findings are mainly based on firm-level models that use country-level investment and covariate data (32). Qualitative analyses and interpretations have mostly relied on deductive normative frameworks such as small versus large farms or family-owned versus capitalist operations (19). These limitations highlight the importance of a data-driven, inductive approach to better understand the logics of land-based investments. Here, we used firm and actor level interview and spatial data and a Bayesian network model (BN) to (i) reconstruct the logic of investment decision making by transnational investors investing in commercial agriculture and forestry in Mozambique, Zambia, Tanzania, and Ethiopia and (ii) identify the determinants of investment locations, given their heterogeneity across gradients of resource frontiers and agglomeration economies.

The Model

Our framework of analysis combines rent and FDI location theories. Classic rent theories of von Thünen and Ricardo model economic rent of the land as a function of the comparative advantage of

market accessibility and bio-physical suitability, respectively. As economies integrate, such comparative advantages lead to local and regional specialization (33, 34). A recent body of literature builds on the notion of external scale economies in explaining regional specialization (35) and specifically frontier expansion (36, 37). These works, unlike early rent-theory models, recognize the endogeneity of firm interactions, producer prices, and the geographical distribution of demand in location choices (36-41). Beyond such comparative advantages, firm-specific characteristics also make a specific location attractive to certain types of investors. The *eclectic paradigm* of FDI theory refers to these firm-specific, spatially-transferable, intangible assets that guide investment decisions as *ownership advantages* - which is disparate from ownership of land (42). Such ownership or proprietary advantages, as distinct from *location advantages*, may consist of crop-specific expertise, technologies, managerial skills, and established brands.

Building on this theoretical framework and our interviews (see Methods and SI text for details), we hypothesized that a given investment location can be explained by a set of selection criteria enumerated by the investors, their production choices, and track record (Fig. 1A). We defined investor track record as the aggregate measure of related skill sets, and existing product market control of all the investors involved in a single investment (Fig.1B). Skill set was measured in terms of previous farming or forestry and regional experiences, and product market control in terms of existing local and export market control. We operationalized these variables using interview and additional firm data (see Methods and Table S2). To characterize farm and plantation locations, we used two key dynamic dimensions: *resource frontiers* (40, 43) and *agglomeration economies* (36, 44). These dimensions, long-established for characterizing commercial agricultural frontiers, encapsulate the compromise an investor makes in choosing cheap land with the potential to expand and achieve scale economies internal to the firm, as opposed to locating closer to an investment cluster to benefit from the existing scale economies that are external to the firm.

To operationalize the two dimensions, i.e., resource frontiers and agglomeration economies, we drew upon the following definitions. A resource frontier is delineated as a transformative moment in-between the discovery or the reinvention of an abundant resource and its consolidation. During this time, the resource, the means of its exploitation, such as labor and capital, and the institutional orders that govern them undergo reconfiguration (45). Some key defining features of this reconfiguration amenable to modelling are the dwindling resource, in this case the land base, and the accruing labor force or capital inputs (40, 46). Here, to quantify the resource frontier we used the two spatial variables; (i) population density and (ii) the proportion of potentially suitable land yet to be converted to agriculture (see Table S3 for data sources). We combined these two measures into a single index, using elicited conditional probabilities (see Table S4 for thresholds), which gauged each investment location along a resource frontier gradient. For agglomeration economies, we adopted the definition that they are scale economies external to the individual firm, but internal to the sector (47, 48). These external scale economies offer localized clusters of specialized knowledge, inputs, and industry-specific infrastructure and institutions (36, 44). To quantify agglomeration economies, we used the two spatial variables; (i) economic activity and market accessibility and (ii) average field size, to proxy the presence of large-scale, capitalized agriculture. Using elicited conditional probabilities, we calibrated a second index to combine these two measures (Table S5). Although resource frontiers and agglomeration economies could evolve independently of each other, their intensities are likely to be negatively associated, presenting a potential trade-off for investors on location choice. Building on this premise, we combined the two indices into a third to construct a typology of four distinct investment locations (Table S6): (i) *populated smallholder land* which is low

on both resource frontier and agglomeration economies indices, (ii) *subsistence frontier*, which ranges medium to high on the resource frontier index and is low on the agglomeration economies index, (iii) *emerging commercial frontier*, which ranges medium to high on both resource frontier and agglomeration indices, and (iv) established markets, which is low on the resource frontier index and high on the agglomeration economies index (Fig.2).

The investors, investments, and key selection criteria

We surveyed 37 investments operated by 29 investors across 121 farm and plantation locations, covering 11% of the total transnational agricultural and forestry investments made between 2000 and 2016 in Mozambique, Zambia, Tanzania, and Ethiopia. We carried out 62 semi-structured interviews across the management chain of these investments and 32 additional interviews with other actors. Our sample contained three main types of investors – *agribusinesses* (55 %), *fund managers* (28 %), and *development finance institutions* (DFIs; 17 %). These investors originated from the UK, South Africa, India, Singapore, USA, Saudi Arabia, Norway, and the Netherlands. We identified *agribusinesses* as those investors who create value through production and *fund managers* as those who generate value through rentier relations, i.e., extracting value through financial transactions involving commodities or property rights rather than through production (49). We identified those who invested with an underlying goal to create jobs and induce scale economies, in places where these would otherwise be unlikely as *DFIs*. Individual investors in our sample, largely corresponded to these three predominant investor types, which have been identified also by previous works (50-54). However, the actual investments were often made up of two or more investors forming new configurations that constantly adapted to the prevailing constraints. Such configurations included corporate agribusinesses who managed supply chain logistics and sold in partnership with small farmers; fund managers who leased out land to contract farmers; and medium- to large-scale agriculture ‘projects’ in which all three types of investors had invested. These investments did not represent the profile of a single partner anymore, but rather an amalgam of the individual investors’ skill sets, experiences, market reach, and motives. Therefore, unless otherwise specified, we report our findings by the investment and not by the investor.

In terms of production, we separated the operations (n=121) into three main types, i.e., *high-value food crops* (11 %), including coffee, macadamia nuts, deciduous fruits such as apples, avocados, citrus, litchi, and pears, *other agriculture* (30 %), including banana, corn, cotton, maize, peas, potato, soya, tobacco, vegetables, horticulture, and livestock, and *forestry* (60 %), including eucalyptus and pine species.

In terms of investor *track records* (Fig. 3), i.e., the aggregate effect of skill sets and product market control, 38 % of the investments (n=37) were made up of investors that had an *extensive track record*. Almost half of the investments (49 %) recorded previous farming or forestry experience similar to the production types targeted by the current investment. On the contrary, 19 % lacked any farming or forestry experience. Around 62 % of the investments had previous commercial experience in the region. Around 14 % had no related skill sets. In terms of product market control, 41 % of the investments had an already established export market for the target crop and 19 % had an already established local market. Around 41 % of the investments were new to the product markets.

The investors identified over 54 criteria that had guided and informed their choice of investment locations and production types (see Table S2 for a description of the criteria). We aggregated these into key groups of selection criteria (Fig. 4). The top six most frequently cited groups of criteria in order of descending importance were land accessibility, agroecology, market proximity, market drivers, infrastructure and logistics, and policy environment. Contrary to widespread narratives, governance conditions did not feature a high priority in guiding investment decisions within Southern and Eastern Africa. This is consistent with global-scale studies using secondary data that found general institutional quality including corruption to have an insignificant effect on land-based investments (30, 31).

Investments focus on pre-commercial agriculture frontier areas

Based on the BN output, around three quarters of the investments were located in regions of low agglomeration economies, while over a fifth of the investments were located in high resource frontier conditions (Fig. 5). Close to 49 % of the investments were located in regions recording medium to high levels of resource frontier conditions and low levels of agglomeration economies, which we identified as *subsistence frontiers*. Around 30 % of the investments were located in regions we defined as *smallholder populated lands* with low levels of resource frontier conditions and agglomeration economies. Only 13 % of the investments were located in *emerging agricultural frontiers*, which recorded medium to high levels of resource frontier conditions and agglomeration economies. With almost half of the investments occurring in places, where land resources are in relative abundance, but scale economies external to individual firms are lacking, we find agricultural frontiers in Southern and Eastern Africa to be predominantly in a “precommercial” context. Mutual information shared between variables (also known as the entropy reduction), i.e., how much knowing one variable reduces the uncertainty of the other (55), revealed that in general, an investment location was most sensitive to the following selection criteria: land accessibility, socio-cultural conditions, market drivers, infrastructure and logistics, and other inputs supply (Fig. 6A). Furthermore, an investment location was more sensitive to investment selection criteria and investor track records than to production types.

Investment locations are associated with investment selection criteria, investor track records, and types of production

Different types of production and investor track records had distinct relations with investment selection criteria and locations (Figs. 6B-D, 7, S1 and S2). Subsistence frontiers remained the most likely investment destination across all production types and investor track records. However, when conditioned on the type of production, forestry investments, which require large tracts of land, favored resource frontier conditions (Fig. S1A), while agricultural investments favored agglomeration economies (Fig S2A). Investments into high-value food crops favored emerging commercial frontiers while investments into other agriculture favored established markets (Fig. 7A). In terms of track records, those with an extensive track record showed little deviation from the general trend (Fig. 7B, S1-2B). Those with limited track records favored low agglomeration economies, in contrast to those without a track record who favored high agglomeration economies (Fig. S2B).

Those with an extensive track record who invested in high-value food crops strongly favored subsistence frontiers (Fig. 7C, S1-2C). These investors weighed heavily on market demand and

strategic reasons compared to other investors (Fig. 8). The key selection criteria most sensitive to the investment locations of these investors were labor supply, policy environment, agroecology, strategic reasons, and operations-related reasons (Fig. 6B). In contrast, forestry investors with an extensive track record favored smallholder populated land. They tended to weigh more on market proximity and infrastructure and logistics (Fig. 8). The most sensitive selection criteria to their investment location choices were limited to agroecology, socio-cultural conditions, land accessibility, market drivers, and infrastructure and logistics (Fig. 6C). Investors who invested in other agriculture with a limited track record favored populated smallholder lands and to a lesser extent subsistence frontiers. Deviations in the investment selection criteria prioritized by these investors included labor supply and policy environment (Fig. 8). The most sensitive selection criteria to their investment choices consisted of market drivers, economic and financial conditions, other inputs supply, strategic reasons, and socio-cultural conditions (Fig. 6D). Notably, investors with extensive track records relied on fewer selection criteria, in contrast to those with limited track records who relied on the entire range of selection criteria in making their investment choices.

The comparative advantages of Southern and Eastern African frontiers

Across much of Southern and Eastern Africa, infrastructure and logistics are basic, input and output markets are imperfect, capital markets are underdeveloped, agglomeration economies are absent, and corruption is prevalent (56). Where states fail to provide basic infrastructure and services, these costs must be borne fully by the investor. The costly endeavor of investing in these “pre-commercial” frontiers may seem a rather peculiar choice. The predilection for narratives built around rent-seeking and corruption in explaining the land-based investment wave in Africa is therefore, unsurprising. In contrast to such conjectures, our empirical findings point to a broader set of comparative advantages. Furthermore, we found the investor’s response to these comparative advantages to be heterogeneous. Here, we identify the main determinants of investment locations into four key comparative advantages and discuss them in the light of investor narratives. The four key comparative advantages are: (i) large tracts of agroecologically suitable land, (ii) agroecologically unique land, (iii) potential market access, and (iv) location advantages conditional on investor track records.

i. Large tracts of agroecologically suitable land

Land, its availability and access, and agroecological suitability were the most common selection criteria guiding investor’s location choices (Fig. 4). Within these broad categories, four aspects stood out; extent, level of occupancy, availability of water resources, and brownfields, i.e., already established farms or plantations. Large tracts of land with low population densities, which at the time of the investments were perceived as unused lands, were a major attraction. Adequate year-round irrigation potential was the single most sought-after agroecological condition. These results are consistent with previous global-scale studies using country-level data that also found land availability (31, 57), agroecological potential (30), and access to irrigation water (31) to be the key determinants of transnational land-based investments. Investors sought brownfield investments for two reasons. First, brownfield investments in many instances offered existing farm infrastructures and in some cases farm or plantation plots with already established tree crops. Second, brownfield investments offered land with existing title deeds, reducing the hurdles of obtaining tenure rights and the risk of conflicts associated with tenure transfers. Much of the current work on land-based investments has specifically focused on the new set of asset-seeking actors who were drawn to land

by poorly performing financial markets, high commodity prices, and attractive historical returns on farmlands (8, 52, 58). However, when the post-2007 land-based investment spike unfolded, farmlands in the global north were already a scarce resource that had sustained high property values (6, 59). Against this looming land scarcity (60), non-substitutability and the immobility of land resources make investments into land attractive, both for rent accruing as well as for production-oriented activities.

ii. Agroecologically unique land

Here we make the distinction between agroecologically suitable land and land with unique agroecological characteristics. Seeking land with agroecological conditions apt for a specific crop entails a different set of constraints than seeking for generic arable conditions. Our results show that investors who focus on high-value crops with specific agroecological requirements are more likely to move deep into scarcely-populated frontiers in pursuit of such land. For example, an investment into macadamia nuts in our sample prioritized land that offered the unique growing conditions but had limited market access, over land with generic agroecological conditions with excellent market access. An investor who supplied avocados to international markets and required guaranteeing year-round supply sought land with a climate that provided a suitable crop calendar to fill gaps in existing supply (61). An investor growing seed potato looked for land that was spatially isolated from other potato farms to minimize the risk of pest infections. While much of the recent focus has been on asset-seeking institutional investors whose investments were swayed by land and crop-price speculation, production-oriented agribusinesses (known as the resource-seeking investors in FDI literature) have been investing in land and occupying the African frontier space for many decades in search of specific agroecological niches.

iii. Potential market access

Markets was another key advantage that the African frontiers potentially offered. These market advantages were largely anticipatory and attracted investors who were new to land-based investments. Among those investors, whose investment decisions were guided by product markets and the availability of infrastructure and logistics, some responded to anticipatory local markets that were expected to benefit from a booming African population. Some responded to anticipatory export markets that were expected to benefit from the growing affluence in Asian markets. Others responded in anticipation of infrastructure developments that potentially created new market connections. Investments that took place through state or donor-sponsored investment schemes in particular (e.g., Nacala and Beira corridors in Mozambique, Southern Agricultural Growth Corridor of Tanzania, and the Malonda Foundation in Mozambique) relied on infrastructure development and economic incentives that were promised but hitherto had rarely been realized. This criticism, often stated in the interviews, echoed the unsuccessful attempts of state-led regional development schemes offering broad incentives to induce firms to locate in less-developed regions (47, 62). In contrast, investors with extensive track records, especially those who had achieved internal scale economies, either owned existing processing centers or had the financial resources large and patient enough to build such facilities. Such investors did not find the basic infrastructures or the lack of logistic services an impediment for investing. Forestry investors were the one exception to this, who despite extensive track records relied on the prospect of product markets, especially in Asia.

iv. Location advantages conditional on investor track records

Location advantages such as large tracts of agroecologically suitable land and access to markets are independent of the investor. Despite heterogeneous responses, these location advantages present equal opportunities for cheaper production, rent accruing, or easy market access to all investors. Investors who specialize in a particular crop for which they possess an extensive track record, command location advantages that stem from firm-specific ownership advantages, such as crop-specific knowledge, technologies, managerial skills, and established brands. Investment logics of such an investor, who operates with a goal to further strengthen core competencies and increase its competitive edge, are conditioned by these ownership advantages. These location advantages are therefore, unique to the investor. For example, an investor in our sample, who had established supply chains, gained managerial proficiency, and amassed information networks in Mozambique, continued to invest in the country, regardless of superior location advantages in a different country. Investors with extensive track records are also likely to pursue unique location advantages tied to the crops they specialize in. For instance, growing coffee in the Harar region in Ethiopia, offered a reputation for quality that only certain established brands could adequately exploit. Growing tobacco in the Tete province in Mozambique benefited from a large pool of customary labor specialized in the crop, with whom an investor had a long-term trust relation. Guided by ownership advantages, investors with extensive track records prioritize specific growing or market conditions, which give them an edge over their competitors. In contrast, investors with limited track records prioritize a range of generic location advantages from infrastructure and logistics to land availability.

Beyond the binary logics

Far from being the homogeneous group of speculators described in much of the land grab discourses, investors who develop medium to large-scale commercial agriculture and forestry in Southern and Eastern Africa show a wide portfolio of skill sets and market reach, which has an effect on where they invest, in what, and why. Investors with extensive track records focusing on high-value crops invest preferentially in remote subsistence frontiers, where they seek land with specific agroecological conditions ideal for crops for which they have strategic reasons to invest in and achieve scale economies by horizontal integration. In doing so, they push the frontier, which likely results in expansion into uncleared natural vegetation. In contrast, newcomers to a sector, including asset-seeking rentiers, focus more on populated smallholder and emerging commercial frontier regions, where they seek the general advantages of proximity to infrastructure and logistics, input markets, and product markets. Similarly, forestry investments tend to be located in smallholder populated regions, which have sufficient basic infrastructures such as roads and electricity, but where the large tracts of land that they need are more easily accessible than closer to centers of commercial development. These investments potentially result in conflicts with more densely settled local populations.

Here, we emphasize the importance of avoiding overly simplistic narratives. While investors with an extensive track record continue to push the frontier boundary, these same investors are more likely to survive the high transaction costs of operating in the pre-commercial frontiers. Compared to other investors, their chances of stimulating employment and economic development, are therefore higher. State-led regional development schemes across the world, where supporting infrastructure and local expertise to leverage cluster formation and agglomeration were lacking, have been largely unsuccessful (47, 62). To realize any lasting developmental impacts on local economies, the positive spillover effects of increased agricultural productivity need to channel through backward (e.g., supplier contracts and input service collaborations) and forward linkages (e.g., processing services)

across the various sectors of the economy. However, such positive spillovers to the external economy have notably been rare in the African frontiers (63). This has been the case even with the much applauded models of contract farming and out-grower schemes that achieve internal scale economies, but not beyond.

It is worth pointing out that the dynamics of land-based investments in Southern and Eastern Africa are changing. High rates of investment failures (4) and the resounding opposition toward large-scale commercial agriculture have not gone unnoticed. The number and scale of new investments have declined (Fig. S3), while attempts to improve investment governance are underway. These attempts range from voluntary investment codes to obligatory funding stipulations to the promotion of private property rights, which include land registration and consolidation (64, 65). There has also been self-learning from the investors part. The more recent investments continue to coalesce different investors to integrate sector-specific skillsets, regional experiences, and patient-enough funding windows that were previously missing. However, what these efforts to date are lacking is a dialogue around the type of investors and investments who can create the catalytic conditions for agglomeration economies with minimal environmental and social trade-offs.

Methods

Data. We compiled a list of transnational agricultural and forestry investments in Mozambique, Zambia, Tanzania, and Ethiopia based on Land Matrix data (4). According to the Land Matrix, between 2000 and 2016, these countries signed contracts for a total of 342 agriculture and forestry investments (see SI text on Study sites and Table S1 for key development indicators of the four countries). Together the investments stretched over an area of 7.7 million hectares and amounted to 41% of the total number and 18% of the total area of land-based investments in Sub-Saharan Africa (4). In choosing the sample countries we focused on investor hotspots within Southern and Eastern Africa that were the common targets of individual investors, with adjustments to capture the heterogeneity of the investment contexts. We reached out to the management using contact details made available on the company websites or through personal contacts. We did an initial round of interviews with the investors who responded to our request and followed on with snowballing by asking those investors to introduce us to others. The final sample includes 37 investments operated by 29 investors across 121 farm and plantation locations.

To generate actor- and firm-level investor data, we carried out 62 semi-structured interviews across the management chain. The top-level management, including chief-investment officers (CIOs), chief executive officers (CEOs), managing directors, and investors, provided information on land-use decisions such as the production choices (e.g., which crops, livestock, horticulture, or forestry species) and locations. The farm and country managers filled in with details on day-to-day management and operations. We conducted 32 additional interviews with local investors and other actors from international organizations, civil society, international and national NGOs, state departments, and academia to better understand the prevailing narratives and the larger context. Building on the previous works and information gathered during a pilot field visit, we developed a semi-structured questionnaire. All the interviews were carried out on site, except for four virtual interviews. The anonymity of the investors was guaranteed. Each interview lasted approximately 1.5 to 2.5 hours and served to collect data on investment locations, the selection criteria adopted in choosing the investment locations, cultivated crops, and the company profiles including previous farming experience, regional experience, and product market control. The interview data was

complemented with additional information on the investments, extracted from company annual reports, company websites, online news reports, and literature. Our sample comprised 11% of the total transnational agricultural and forestry investments made between 2000 and 2016 in the four countries.

To generate the spatial data, we initially geolocalized 268 farm and plantation sites managed by the investors we sampled, using field records, Google Earth, and secondary spatial administrative data (e.g., land use title data for plantations and forestry concessions in Mozambique). We applied a 30 km buffer, to eliminate the neighboring farms that were a part of the same investment, which brought the number of spatially distinct farms and plantation sites down to a total of 121. We used the distance between farms managed under a single investment and the farm sizes as guides to set the buffer size. Farms managed by a single investment were often within a 7 – 8 km radius, but the furthest were located at around a 30 km radius. The large farms in the region ranged between 2000-3000 ha (i.e., 20-30 km²). Due to their vast expanses, forest plantations managed by a single investment spread over much larger ranges. We therefore, chose the upper limit of the distance between farms (i.e., 30 km) as the buffer size to extract variables characterizing investment locations, as it allowed minimizing the spatial dependence in the data on the one hand and capturing the context beyond the farm and plantation boundaries, on the other. Then, we compiled a database of spatial variables capturing land cover, population density, field size, and market influence (see Table S3 for data sources). For each farm or plantation location, we calculated the mean value of population density, field size, and market influence and the proportion of land cover data, within a 30 km buffer using ArcGIS™.

Bayesian network. We converted the conceptual framework (Fig. 1A) into a directed acyclic graph (DAG), as shown in Fig. 5 and operationalized it in a Bayesian network (BN) using firm and actor level interview and spatial data (see SI section on Variables for details on data and processing). A BN is a probabilistic graphical model, which encodes conditional dependencies between variables (66, 67). BNs can combine qualitative and quantitative data and capture the uncertainty and variation inherent in observation and parameter estimation (see SI section on The Bayesian network (BN), for details). This made BNs ideal for our purpose of apprehending the complexity of the underlying decision making (68). Our BN comprised 14 predictor variables including two investor profile and 12 investment selection criteria variables. The two outcome indices, i.e., *resource frontier* (see Table S4 for thresholds) and *agglomeration economies* (see Table S5 for thresholds), were each calibrated using two spatial variables. A final output typology, i.e., *investment location* was derived by combining the two outcome indices (see Table S6 for thresholds). Table S2 details the individual variables (BN nodes), their states and thresholds, and the observed variables used in calculating the conditional probability tables (CPTs). The BN was parameterized using these CPTs. We used Netica™ Version 5.24, to implement the BN and the outputs and sensitivity scores were imported to R (69) for further analysis and plotting (70). We validated the model by conducting a validation exercise with a select group of investors and researchers (See SI section on Model validation).

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No 677140 MIDLAND).

We thank all our study participants, for their time and expertise. Without their support this work would not have been possible. We also wish to thank fellow MIDLAND colleagues Ms. Adia Bey, Ms. Virginia Rodriguez García, and Dr. Eduardo Henrique Da Silva Oliveira for the helpful discussions that shaped this work. Our work contributes to the Global Land Programme (<https://glp.earth>).

References

1. P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, M. C. Hansen, Classifying drivers of global forest loss. *Science* **361**, 1108 (2018).
2. A. B. Brink, H. D. Eva, Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. *Applied Geography* **29**, 501-512 (2009).
3. T. K. Rudel, The national determinants of deforestation in sub-Saharan Africa. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120405 (2013).
4. Land Matrix (2019) (The Land Matrix Initiative).
5. UNCTAD (2019) (United Nations Conference on Trade and Development (UNCTAD), Geneva, Switzerland).
6. P. McMichael, The land grab and corporate food regime restructuring. *The Journal of Peasant Studies* **39**, 681-701 (2012).
7. S. M. Borras Jr., R. Hall, I. Scoones, B. White, W. Wolford, Towards a better understanding of global land grabbing: an editorial introduction. *The Journal of Peasant Studies* **38**, 209-216 (2011).
8. L. Cotula, S. Vermeulen, R. Leonard, J. Keeley (2009) Land grab or development opportunity? Agricultural investment and international land deals in Africa. in *Land, investment and rights*, p 145.
9. K. Deininger *et al.*, *Rising global interest in farmland : can it yield sustainable and equitable benefits?*, Agriculture and rural development (World Bank Group, Washington, D.C., 2011).
10. J. Fairhead, M. Leach, I. Scoones, Green Grabbing: a new appropriation of nature? *The Journal of Peasant Studies* **39**, 237-261 (2012).
11. S. Vermeulen, L. Cotula, Over the heads of local people: consultation, consent, and recompense in large-scale land deals for biofuels projects in Africa. *The Journal of Peasant Studies* **37**, 899-916 (2010).
12. L. German, G. Schoneveld, E. Mwangi, Contemporary Processes of Large-Scale Land Acquisition in Sub-Saharan Africa: Legal Deficiency or Elite Capture of the Rule of Law? *World Development* **48**, 1-18 (2013).
13. K. F. Davis *et al.*, Tropical forest loss enhanced by large-scale land acquisitions. *Nature Geoscience* **13**, 482-488 (2020).
14. G. C. Schoneveld, L. A. German, E. Nutakor, Land-based Investments for Rural Development? A Grounded Analysis of the Local Impacts of Biofuel Feedstock Plantations in Ghana. *Ecology and Society* **16** (2011).
15. P. Collier, S. Dercon, African Agriculture in 50Years: Smallholders in a Rapidly Changing World? *World Development* **63**, 92-101 (2014).
16. B. Phalan, M. Onial, A. Balmford, R. E. Green, Reconciling Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Compared. *Science* **333**, 1289 (2011).
17. A. Balmford *et al.*, The environmental costs and benefits of high-yield farming. *Nature Sustainability* **1**, 477-485 (2018).
18. R. E. Green, S. J. Cornell, J. P. W. Scharlemann, A. Balmford, Farming and the Fate of Wild Nature. *Science* **307**, 550 (2005).
19. C. Oya, Methodological reflections on 'land grab' databases and the 'land grab' literature 'rush'. *The Journal of Peasant Studies* **40**, 503-520 (2013).
20. K. Deininger, Challenges posed by the new wave of farmland investment. *The Journal of Peasant Studies* **38**, 217-247 (2011).
21. B. G. Glaser, A. L. Strauss, *The discovery of grounded theory : strategies for qualitative research* (Aldine Publishing, Chicago, 1967).
22. F. Noorbakhsh, A. Paloni, A. Youssef, Human Capital and FDI Inflows to Developing Countries: New Empirical Evidence. *World Development* **29**, 1593-1610 (2001).

23. M. Busse, C. Hefeker, Political risk, institutions and foreign direct investment. *European Journal of Political Economy* **23**, 397-415 (2007).
24. N. Jensen, Political Risk, Democratic Institutions, and Foreign Direct Investment. *The Journal of Politics* **70**, 1040-1052 (2008).
25. A. Cuervo-Cazurra, Better the devil you don't know: Types of corruption and FDI in transition economies. *Journal of International Management* **14**, 12-27 (2008).
26. A. Cuervo-Cazurra, Who cares about corruption? *Journal of International Business Studies* **37**, 807-822 (2006).
27. S.-Y. Wu, Corruption and cross-border investment by multinational firms. *Journal of Comparative Economics* **34**, 839-856 (2006).
28. P. Egger, H. Winner, Evidence on corruption as an incentive for foreign direct investment. *European Journal of Political Economy* **21**, 932-952 (2005).
29. S.-J. Wei, How Taxing is Corruption on International Investors? *The Review of Economics and Statistics* **82**, 1-11 (2000).
30. R. Arezki, K. Deininger, H. Selod, What Drives the Global "Land Rush"? *The World Bank Economic Review* **29**, 207-233 (2013).
31. J. Lay, K. Nolte, Determinants of foreign land acquisitions in low- and middle-income countries. *Journal of Economic Geography* **18**, 59-86 (2017).
32. B. A. Blonigen, A Review of the Empirical Literature on FDI Determinants. *Atlantic Economic Journal* **33**, 383-403 (2005).
33. G. I. P. Ottaviano, D. Puga, Agglomeration in the Global Economy: A Survey of the 'New Economic Geography'. *The World Economy* **21**, 707-731 (1998).
34. P. Krugman, Increasing Returns and Economic Geography. *Journal of Political Economy* **99**, 483-499 (1991).
35. S. Scotchmer, J.-F. Thisse, Space and competition. *The Annals of Regional Science* **26**, 269-286 (1992).
36. R. D. Garrett, E. F. Lambin, R. L. Naylor, The new economic geography of land use change: Supply chain configurations and land use in the Brazilian Amazon. *Land Use Policy* **34**, 265-275 (2013).
37. P. D. Richards, R. T. Walker, E. Y. Arima, Spatially complex land change: The Indirect effect of Brazil's agricultural sector on land use in Amazonia. *Global Environmental Change* **29**, 1-9 (2014).
38. J. A. Robalino, A. Pfaff, Contagious development: Neighbor interactions in deforestation. *Journal of Development Economics* **97**, 427-436 (2012).
39. Y. le Polain de Waroux *et al.*, Rents, Actors, and the Expansion of Commodity Frontiers in the Gran Chaco AU - le Polain de Waroux, Yann. *Annals of the American Association of Geographers* **108**, 204-225 (2018).
40. P. Meyfroidt *et al.*, Middle-range theories of land system change. *Global Environmental Change* **53**, 52-67 (2018).
41. P. Krugman, What's new about the new economic geography? *Oxford Review of Economic Policy* **14**, 7-17 (1998).
42. J. H. Dunning, The Eclectic (OLI) Paradigm of International Production: Past, Present and Future. *International Journal of the Economics of Business* **8**, 173-190 (2001).
43. E. F. Lambin *et al.*, Estimating the world's potentially available cropland using a bottom-up approach. *Global Environmental Change* **23**, 892-901 (2013).
44. P. Richards, It's not just where you farm; it's whether your neighbor does too. How agglomeration economies are shaping new agricultural landscapes. *Journal of Economic Geography* **18**, 87-110 (2018).
45. M. B. Rasmussen, C. Lund, Reconfiguring Frontier Spaces: The territorialization of resource control. *World Development* **101**, 388-399 (2018).
46. E. B. Barbier, Scarcity, frontiers and development. *The Geographical Journal* **178**, 110-122 (2012).
47. M. E. Porter, Competitive Advantage, Agglomeration Economies, and Regional Policy. *International Regional Science Review* **19**, 85-90 (1996).
48. A. Scott, M. Storper, Regions, Globalization, Development. *Regional Studies* **37**, 579-593 (2003).
49. P. Meyfroidt, D. Abeygunawardane, N. Ramankutty, A. Thomson, G. Zeleke, Interactions between land systems and food systems. *Current Opinion in Environmental Sustainability* **38**, 60-67 (2019).
50. M. Boche, W. Anseeuw (2013) Unraveling "land grabbing": Different models of large-scale land acquisition in Southern Africa. in *LDPI Working Paper* (Land Deal Politics Initiative (LDPI), Cape Town), p 26.
51. A. Ducastel, W. Anseeuw, Agriculture as an asset class: reshaping the South African farming sector. *Agriculture and Human Values* **34**, 199-209 (2017).
52. S. Daniel, Situating private equity capital in the land grab debate. *The Journal of Peasant Studies* **39**, 703-729 (2012).

53. S. Ouma, Getting in between M and M' or: How farmland further debunks financialization. *Dialogues in Human Geography* **5**, 225-228 (2015).
54. Z. Kish, M. Fairbairn, Investing for profit, investing for impact: Moral performances in agricultural investment projects. *Environment and Planning A: Economy and Space* **50**, 569-588 (2017).
55. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers Inc., 1988).
56. A. Dorward, J. Kydd, C. Poulton, Beyond Liberalisation: "Developmental Coordination" Policies for African Smallholder Agriculture. *IDS Bulletin* **36**, 80-85 (2005).
57. M. J. Burger, B. van der Knaap, R. S. Wall, Revealed competition for greenfield investments between European regions. *Journal of Economic Geography* **13**, 619-648 (2012).
58. G. Lawrence, S. Sippel, D. Burch, "The financialisation of food and farming" in Handbook on the Globalisation of Agriculture, G. M. Robinson, D. A. Carson, Eds. (Edward Elgar Publishing, Cheltenham, UK 2015), 10.4337/9780857939838.00023, pp. 309-327.
59. J. Koeninger (2017) History of institutional farmland investment. (HighQuest Partners).
60. E. F. Lambin, P. Meyfroidt, Global land use change, economic globalization, and the looming land scarcity. *Proc Natl Acad Sci U S A* **108**, 3465-3472 (2011).
61. J. H. Dunning, Location and the Multinational Enterprise: A Neglected Factor? *Journal of International Business Studies* **29**, 45-66 (1998).
62. J. R. Markusen, A. J. Venables, Multinational firms and the new trade theory. *Journal of International Economics* **46**, 183-203 (1998).
63. K. Deininger, F. Xia, Quantifying Spillover Effects from Large Land-based Investment: The Case of Mozambique. *World Development* **87**, 227-241 (2016).
64. S. M. Borras Jr., J. Franco, From Threat to Opportunity? Problems with the Idea of a "Code of Conduct" for Land- Grabbing. *Yale Human Rights and Development Law Journal* **13**, 507-523 (2010).
65. J. Quan, A. Seigneret (2019) Investing Responsibly in Agricultural Land: Lessons from responsible land investment pilots in sub-Saharan Africa. in *A LEGEND report*. (Natural Resources Institute, University of Greenwich, Chatham, UK).
66. F. V. Jensen, T. D. Nielsen, *Bayesian Networks and Decision Graphs* (Springer New York, ed. Second, 2007).
67. A. B. Gelman, J. Carlin, H. Stern, D. Rubin, *Bayesian Data Analysis* (2003), vol. 45.
68. D. Hackerman, E. H. A. Mamdani, M. P. Wellman, Real-World Applications of Bayesian Networks. *Communications of the ACM* **38**, 24-26 (1995).
69. R Core Team (2020) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria).
70. H. Wickham (2016) ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag New York).
71. FAOSTAT (2019) (Food and Agriculture Organization of the United Nations (FAO)).
72. UNECA (2003) Land tenure systems and sustainable development in southern Africa. (Southern Africa SubRegional Development Centre, Addis Ababa).
73. W. Crewett, A. Bogale, B. Korf, Land tenure in Ethiopia: Continuity and change, shifting rulers, and the quest for state control. *International Food Policy Research Institute (IFPRI), CAPRI working papers* 10.2499/CAPRIWP91 (2008).
74. J. Pearl, Causal inference in statistics: An overview. *Statist. Surv.* **3**, 96-146 (2009).
75. P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, A. Salmerón, Bayesian networks in environmental modelling. *Environmental Modelling & Software* **26**, 1376-1388 (2011).
76. D. Heckerman, D. Geiger, D. M. Chickering, Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning* **20**, 197-243 (1995).
77. U. B. Kjærulff, A. L. Madsen, *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. U. B. Kjærulff, A. L. Madsen, Eds. (Springer New York, New York, NY, 2013), <https://doi.org/10.1007/978-1-4614-5104-4>, pp. 382.
78. B. G. Marcot, Metrics for evaluating performance and uncertainty of Bayesian network models. *Ecological Modelling* **230**, 50-62 (2012).
79. Z. Sun, D. Müller, A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. *Environmental Modelling & Software* **45**, 15-28 (2013).
80. J. Cain, *Planning Improvements in Natural Resources Management: Guidelines for using Bayesian networks to support the planning and management of development programmes in the water sector and beyond* (Centre for Ecology and Hydrology, Wallingford, UK, 2001), vol. 124.
81. S. L. Lauritzen, The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis* **19**, 191-201 (1995).

82. M. C. Hansen *et al.*, High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013).
83. T. S. Jayne, D. Mather, E. Mghenyi, Principal Challenges Confronting Smallholder Agriculture in Sub-Saharan Africa. *World Development* **38**, 1384–1398 (2010).
84. World Bank, Enquête Agricole de Conjoncture Intégrée 2014 (EAC-I 2014). <https://microdata.worldbank.org/index.php/catalog/2583>.
85. World Bank, World Development Indicators. <https://databank.worldbank.org/source/world-development-indicators>.
86. W. Li *et al.*, Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). *Earth Syst. Sci. Data* **10**, 219–234 (2018).
87. C. T. Lloyd *et al.*, Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data* **3**, 108–139 (2019).
88. WorldPop, Center for International Earth Science Information Network (CIESIN) at Columbia University, Global High Resolution Population Denominators Project www.worldpop.org
89. M. Lesiv *et al.*, Estimating the global distribution of field size using crowdsourcing. *Global Change Biology* **25**, 174–186 (2019).
90. P. H. Verburg, E. C. Ellis, A. Letourneau, A global assessment of market accessibility and market influence for global environmental change studies. *Environmental Research Letters* **6**, 034019 (2011).

Figures

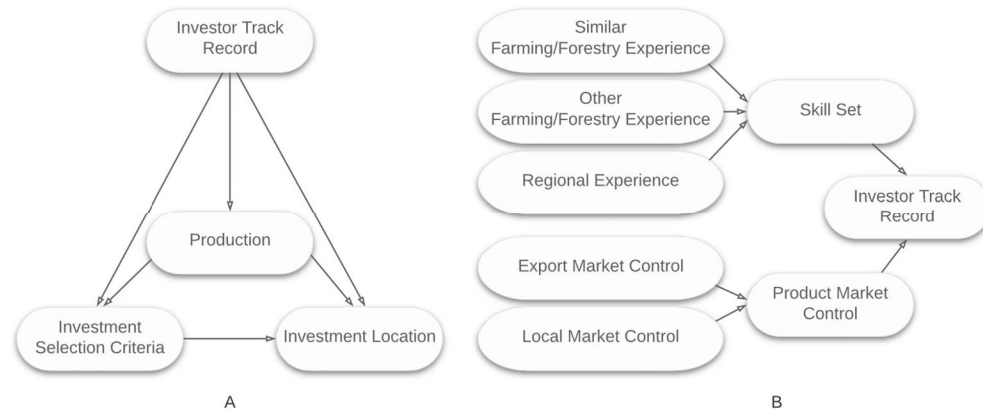


Figure 1. The conceptual model A) interdependencies between the determinants of investment locations and B) variables measuring investor track record. Investor track record is an aggregate measure of related skill sets and existing product market control. Skill set is an aggregate measure of previous farming or forestry and regional experiences. Product market control is an aggregate measure of existing export and local market control.

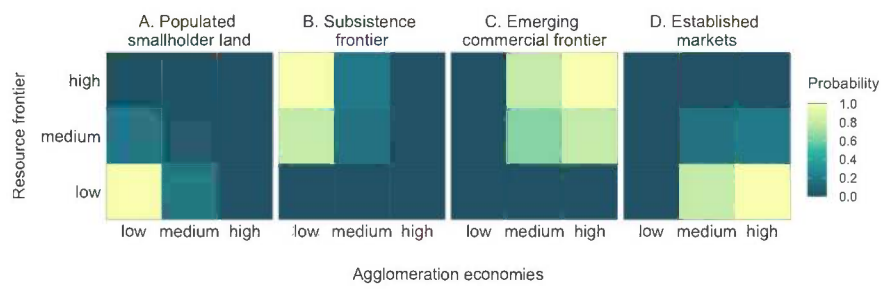


Figure 2. Typology of the four distinct investment locations. i.e., A) populated smallholder land, B) subsistence frontier, C) emerging commercial frontier, and D) established markets, along the low-medium-high gradients of resource frontier and agglomeration economies.

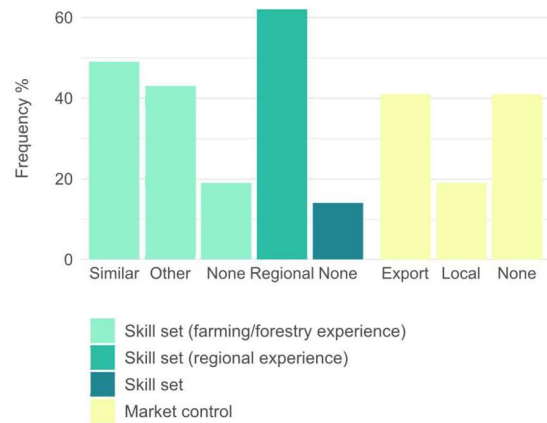


Figure 3. Investor track records including skill sets and market control (n=37). The columns, in the order from left to right, show the frequency of investors: with previous farming or forestry experience in the exact same or similar production as the target production of the investment under study (Similar); with farming or forestry experience in other production (Other); without farming or forestry experience (None); with regional experience (Regional); without sector-related (i.e., farming or forestry and regional) skill sets (None); with existing export market control for the exact same or similar production as the target production (export); with existing local market control for the exact same or similar production (Local); and without export or local market control for the exact same or similar production.

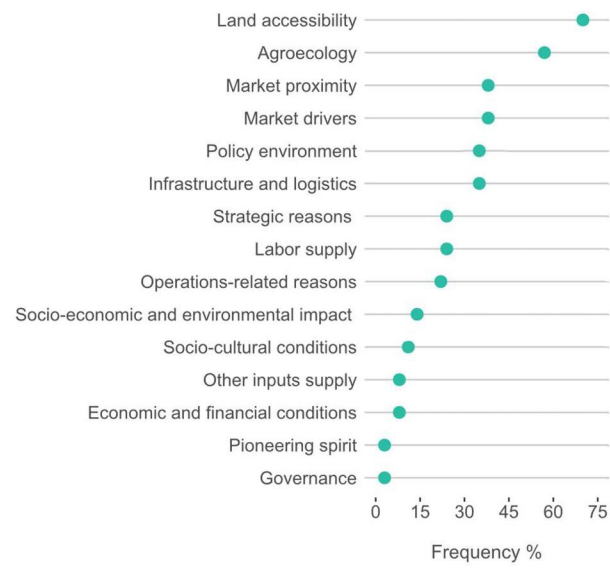


Figure 4. Key groups of investment selection criteria cited by the investors ($n = 37$).

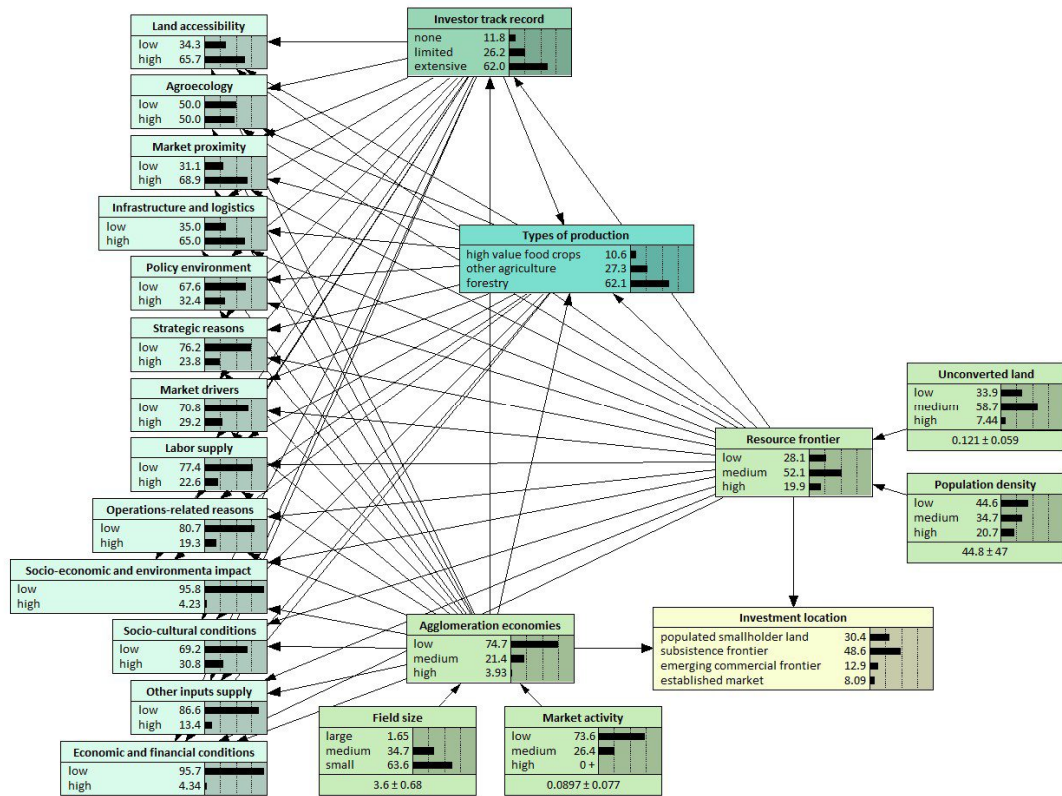


Figure 5. The BN presenting the determinants of investments locations in the African frontiers.

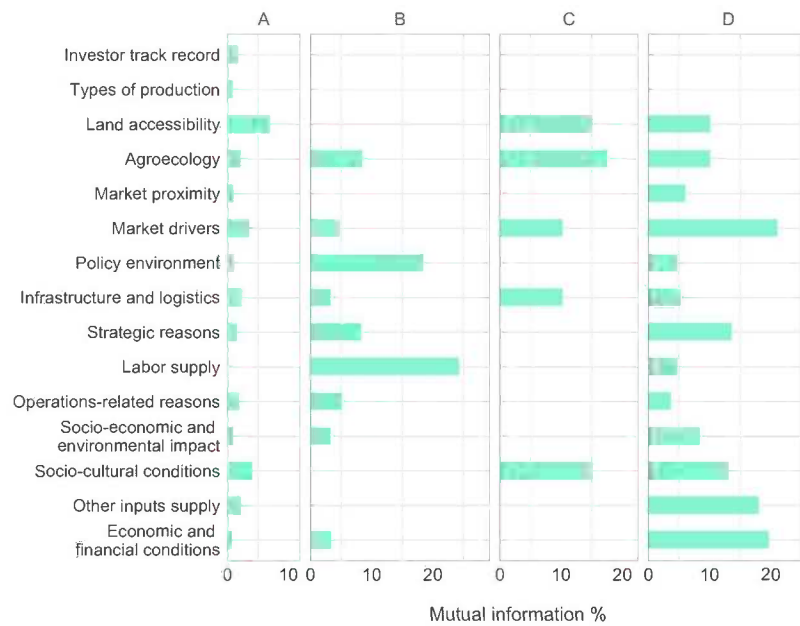


Figure 6. The sensitivity of each variable to the investment locations corresponding to A) general trend, and when controlled for investor track records and types of production as in B) high-value food crop investors with an extensive track record, C) forestry investors with an extensive track record, and D) other agriculture investors with a limited track record.

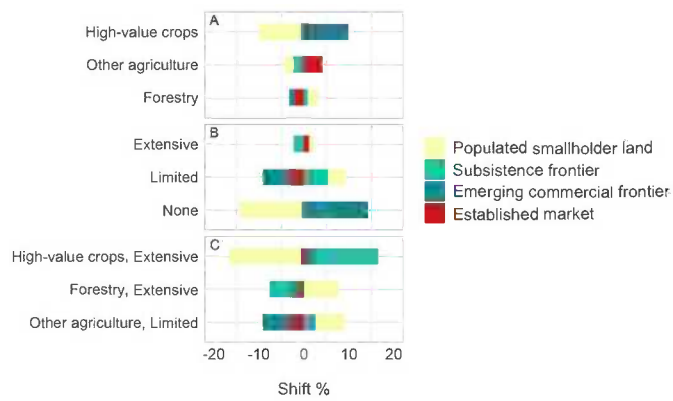


Figure 7. Shifts in the likelihood of investment locations conditional on different A) production types, B) investor track records, and C) the aggregate of production types and track records.

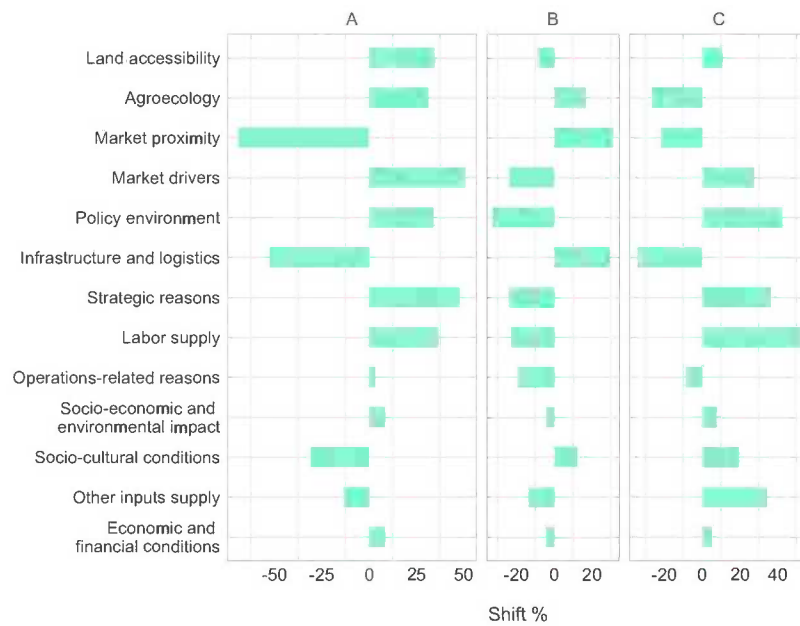


Figure 8. Shifts in investment selection criteria, when conditioned on investor track records and types of production as in A) high-value food crop investors with an extensive track record, B) forestry investors with an extensive track record, and C) other agriculture investors with a limited track record.

Supplementary Information Text

Study sites: The four countries in our sample, i.e., Mozambique, Zambia, Tanzania, and Ethiopia are characterized by rural agrarian economies. A large fraction of the population is poor and derives its livelihood and incomes primarily from farming and related activities (Table S1). The main crops grown in the region include cassava, maize, and sugar cane (71). The four countries have pluralistic land tenure systems which include different forms of customary or communal and statutory forms of land rights (72). The state retains residual 'ultimate ownership', while private tenure is granted through rights of usufruct through lease entitlements that can range between 14 through 99 years (12, 73).

The Bayesian network (BN): A BN is a probabilistic graphical model which encodes the joint probability distribution of a set of random variables. It comprises two components: (i) a qualitative component, which is the directed acyclic graph (DAG) that represent the interdependencies between variables (i.e., nodes) , and (ii) a quantitative component with conditional probability tables (CPT) that quantify the strength of the conditional dependencies between the variables using a set of parameters (66, 74). Each node in the DAG denotes a variable which is an attribute, feature, or hypothesis about an uncertain event with a set of state values. These state values are mutually exclusive, collectively exhaustive, and typically discrete. The directed arrows pointing from parent (i.e., predictor) to child (i.e., target) nodes capture cause-effect relationships and/or statistical correlations (75). The CPTs quantify the strength of the influence of the predictor variables on the target variables, given the combination of the discrete states of all the predictor variables. Mathematically, the conditional interdependencies among the variables are derived using Bayes' theorem (76), which links the probability of hypothesis H conditional on evidence E as $P(H|E) = P(E|H) \times P(H)/P(E)$. The joint probability distribution over a set of variables $X_1, X_2, \dots, X_n$ in a BN can be factorized as $P(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{pa(i)})$, where $X_{pa(i)}$ indicates the set of parent variables of X_i (77). The sensitivity of a variable with respect to other variables in a BN is measured using variance reduction or entropy reduction metrics (78).

BNs can incorporate multiple levels of randomness to combine qualitative and quantitative data and also capture the uncertainty and variation inherent in observation and parameter estimation (67, 68, 79). These attributes were particularly amenable to modelling the uncertainties inherent in decision making, given that we worked with a small interview dataset with some missing values. We used Netica™ Version 5.24 software to implement the BN, which includes a graphical interface and built-in algorithms for Bayesian inference.

Variables: The BN included 14 predictor variables, two outcome variables, and an output typology. Table S2 lists the variables, their definitions, the states each variable take, and their thresholds.

All predictor variables were drawn from interview and additional firm data, which included 37 individual investments. The key predictor variables comprised one observed variable (type of production) and 13 latent variables, i.e., variables not directly observed, but calculated using the observed (investor track record, and a set of investment location selection criteria). The parameters for each latent predictor variable was calculated from the observed variables, outside of the net using MS Excel™ (Table S2, under 'Node states: Thresholds', explains how each variable was calculated). These observed and calculated parameters were used to populate the CPTs of the predictor variables. CPTs take values over the interval $[0, 1]$ (77).

The key outcome variable indices, i.e., resource frontier and agglomeration economies, were derived using four spatial variables. For each investment site ($n=121$), we derived a frontier index using spatial data on population density and the area of unconverted land and an agglomeration economies index using spatial data on market activity and field size (Table S3 lists data sources). We selected datasets that characterized the investment conditions as close as possible to the period when the investment decisions were made. Since BNs generally require discretizing continuous data, we used a mix of descriptive statistics and published standards to decide on the number of bins and bin-widths for the spatial variables (Table S2 lists the bins and bin-widths for spatial data). To derive the indices, we elicited a prior conditional probability for each parent state combination of the

two indices, based on the definitions characterizing these indices (see section Model in the main text for definitions) and our expertise in the domain (80). To elaborate this with an example from the agglomeration economies index, we take the case of an investment location that records a large field size and low-level market activity (i.e., the parent states). For this parent state combination, each of the *low*, *medium*, and *high* child states of the agglomeration economies index, were assigned a probability score of 0.6, 0.4, and 0, respectively. These elicited probability scores indicate that for a given investment location with a large field size and low market activity, the likelihood of the occurrence of agglomeration economies is more likely to be *low* with a 60 % chance, less likely to be *medium* with a 40% chance, and unlikely to be *high*. Table S4 presents the elicited conditional probabilities of the resource frontier index and S5 that of the agglomeration economies index. The conditional probability scores were arrived at by revising and deliberating the scores until there was agreement among all the authors. These outcome variable indices were then parameterized using spatial data. The probability distributions of the indices also accounted for the uncertainty in measurement errors. For parameter estimation, we used the built-in expectation-maximization (EM) algorithm in Netica™ (81).

To facilitate interpretation, we derived an output typology by tabulating the resource frontier and agglomeration economies in a matrix. The typology was calibrated using area-weighted probability scores along the low, through medium, through high gradients of frontier and agglomeration economies indices (Table S6 and Fig. 2 in the main text).

The BN outputs and sensitivity scores were imported to R (69) for further analysis and plotting (70).

Model validation: We arrived at the final BN through an iterative and interactive process involving revisions, verification, and validation (77). At each revision we verified the directional influence and tested the plausibility of the results. A series of final validation exercises was carried out with 3 selected investors among the sampled investors and a group of researchers who work in the region. In these final validation exercises, we asked the participants to explore the DAG visually and assess the conceptual framework and the plausibility of the directional influences by taking different investment conditions as examples. We also assessed the plausibility of the overall results, against participants' expertise in the sector or in the region.

To assess the influence of each selected variable on the other variables quantitatively, we conducted a series of sensitivity tests. A sensitivity test estimates the mutual dependence between two variables, by quantifying the amount of information an observed variable yields about another. It is an entropy reduction measure, which quantifies how much knowing one variable reduces the uncertainty regarding the other (55).

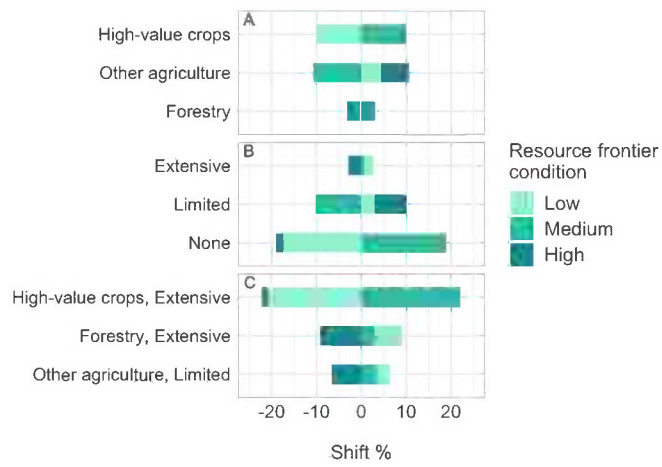


Figure S1. Shifts in the likelihood of resource frontier state conditional on different A) production types, B) investor track records, and C) the aggregate of production types and track records.

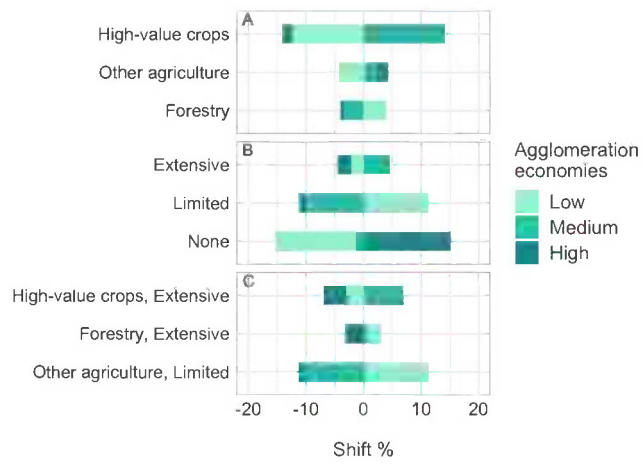


Figure S2. Shifts in the likelihood of agglomeration economies state conditional on different A) production types, B) investor track records, and c) the aggregate of production types and track records.

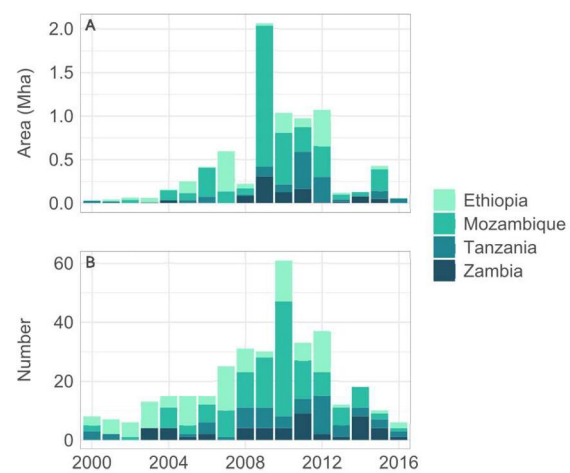


Figure S3. Land based investment trends in Ethiopia, Mozambique, Tanzania, and Zambia between 2000 and 2016, A) area of investments in Mha and B) number of investments (source: 4).

Table S1. *Economic, agriculture, and social development context in the sample countries between 2008-2011 (Adapted from 71, 82, 83-85).*

Country	GDP (USD billion)	Agriculture, value added (% of GDP)	Rural population (%)	Average smallholder farm size (ha)	Forested area in million ha (% of land cover)
Ethiopia	29.93	41.4	82.7	0.6	12.4 (11)
Mozambique	11.09	26.9	69.2	0.9	27.0 (34)
Tanzania	32.01	25.6	71.9	1.1	24.7 (26)
Zambia	20.27	9.4	60.6	0.9	22.4 (30)

Table S2. Description of the investment determinants and their use in the BN. The table lists the observed variables, as described by the investors, the latent variables calculated from the observed which are included as nodes in the BN, node states, and node-state thresholds. (L) denotes a latent variable and (I) denotes an index.

BN nodes	Observed variables	Node states: Thresholds
Predictor variables: Investor profile		
1. Investor track record (L)	Aggregate effect of product market control, skill set, and regional experience	<i>none</i> : all the child/observable variables in none state, <i>limited</i> : one child variable is in none state or both child variables in limited state, <i>extensive</i> : at least one child variable is in extensive state and others in limited or extensive state
a. Product market control (L)	Degree of local and export market control	<i>none</i> : no market control, <i>limited</i> : only local market control, <i>extensive</i> : at least some export market control
i. Export market control (L)	Export market for the invested crop is already established	<i>no</i> : no established export product market, <i>yes</i> : export product market already established
ii. Local market control (L)	Local market for the invested crop is already established	<i>no</i> : no established local product market, <i>yes</i> : local product market already established
b. Skill set (L)	Aggregate effect of farming or forestry experience in the same or similar types of production and other farming or forestry experience	<i>none</i> : all the child variables in no state, <i>limited</i> : all the child variable combinations except none and extensive states, <i>extensive</i> : child variables farming or forestry experience in the same or similar types of production AND regional experience are in yes state
i. Farming or forestry experience in similar types of production (L)	Previous farming or forestry experience in the same or similar types of production as the target production of the investment under study	<i>no</i> : no farming or forestry experience in the same or similar types of production, <i>yes</i> : has farming or forestry experience in the same or similar types of production
ii. Farming or forestry experience in other types of production (L)	Previous farming or forestry experience in types of production other than the target production of the investment under study	<i>no</i> : no farming or forestry experience, <i>yes</i> : has farming or forestry experience in other types of production
iii. Regional experience (L)	Previous experience in farming, forestry, trading, or in any commercial activity in Southern and Eastern Africa	<i>no</i> : no commercial experience in the region, <i>yes</i> : previous commercial experience in the region in farming, forestry, trading, or any other commercial activity
2. Types of production	Type of production	<i>high-value food crops</i> : high-value deciduous fruits and nuts, <i>forestry</i> : logging and plantation forestry, <i>other</i> : other agriculture excluding high value food crops
Predictor variables: Investment selection criteria (aggregate effect of the priorities assigned by the investor to the observable variables);		
1. Agroecology (L)	Irrigation, Altitude, Soils, Other agroecological conditions	<i>low</i> : less than two of the child variables are in high state, <i>high</i> : at least two of the child variables are in high state
<i>Note: Frequency of agroecology outside of the BN measures at least one positive observed variable in the group</i>		
2. Infrastructure and logistics (L)	Access to utility services, Means of transport, Access to processing and storage	<i>low</i> : all the child variables are in low state, <i>high</i> : at least one child variable is in high state
a. Means of transport (L)	Transport logistics, Infrastructure	"
3. Land accessibility (L)	Extent of land, Land conflicts, Population density, Land policy, Brownfields	"
4. Market proximity (L)	Proximity to local markets and distribution hubs, Proximity to international markets	"
5. Labor supply (L)	Labor skill set, Labor availability, Cost of labor	"
6. Policy environment (L)	Agriculture policy, Trade policy, Investment policy	"
7. Strategic reasons (L)	Vertical integration, Competitive advantage, Diversification, Upstream-and downstream balance	"
8. Economic and financial conditions (L)	The proportion of agricultural population to GDP, Long-term investment, Cashflow, Availability of money	"
9. Operations-related reasons (L)	Operational cost, Strong local partners, In situ experience, Technical capacity	"
10. Socio-economic and environmental impact (L)	Environmental impact, Socio-economic impact	"
11. Market drivers (L)	Product price, Demand, Traditional growing areas, Guaranteed market, Timing of the markets	"
12. Other inputs supply (L)	Input supplies, GMOs, Access to forestry(wood), Mechanization	"
Governance	Political stability, Bureaucracy and red tape, Rule of law	Excluded from the BN due to the very low frequency in citing
Pioneering spirit	Pioneering spirit	Excluded from the BN due to the very low frequency in citing
Outcome variables		

1. Resource frontier (I)	Calibrated based on population density and unconverted land	<i>low, medium, and high: see Table S4 for thresholds</i>
a. Population density	People per square kilometer	<i>low: <25/km², medium: 25-50/km², high: >50/km²</i>
b. Unconverted land	Proportion of land unconverted to agriculture but potentially suitable (excluding deserts and bare lands)	<i>low: <10%, medium: 10-20%, high: >20%</i>
2. Agglomeration economies (I)	Calibrated based on field size and market activity	<i>low, medium, and high: See Table S5 for thresholds</i>
a. Field size	Enclosed agricultural areas, including annual and perennial crops	<i>low: <0.64 ha, medium: 0.64 -16ha, high: >16ha</i>
b. Market activity	Travel time to influential markets	<i>low: <0.1, medium: 0.1-0.3, high: >0.3</i>
Output typology		
Investment location (I)	Typology based on resource frontier and agglomeration economies indices	<i>populated smallholder land, subsistence frontier, emerging commercial frontier, and emerging markets: See Table S6 and Fig. 2 in the main text for thresholds</i>

Table S3. *Descriptions and sources of spatial data.*

Variable	Years of data sets	Description	Source
Land cover	1992-2015	Continuous land cover changes at 300 m resolution	Li et al. 2018 (86)
Population density	2000 and 2010	Mosaiced 1km resolution global data sets using sub-national census-based population estimates interpolated at 100m using Random Forest estimation	Lloyd et al. 2019 (87, 88)
Field size	2017	Enclosed agricultural areas, including annual and perennial crops mapped at 30m resolution	Lesiv et al. 2019 (89)
Market influence	1979-2010	Travel time accounting for infrastructure and terrain to influential markets weighted by GDP PPP	Verburg et al. 2011 (90)

Table S4. Elicited probability table used in calibrating the priors for the resource frontier index. The values represent an expert derived probability score of how likely the different states of the child node (i.e., the resource frontier variable) are, given the different states of the two parent nodes (i.e., population density and unconverted land). For e.g., the first line of the table reads, if population density is low and unconverted land is low, there is a 60 % chance that resource frontier is low, a 40 % chance that it is medium, and no chance that it is high.

Population density	Unconverted land	Resource Frontier		
		low	medium	high
low	low	0.6	0.4	0
low	medium	0	0.6	0.4
low	high	0	0	1
medium	low	0.8	0.2	0
medium	medium	0	0.8	0.2
medium	high	0	0.4	0.6
high	low	1	0	0
high	medium	0.2	0.8	0
high	high	0.1	0.6	0.3

Table S5. Elicited probability table used in calibrating the priors for the agglomeration economies index. The values represent an expert derived probability score of how likely the different states of the child node (i.e., the agglomeration economies variable) are, given the different states of the two parent nodes (i.e., field size and market activity). For e.g., the first line of the table reads, if field size is large and market activity is low, there is a 60 % chance that agglomeration economy is low, 40 % chance that it is medium, and no chance that it is high.

Field size	Market activity	Agglomeration economies		
		low	medium	high
large	low	0.6	0.4	0
large	medium	0	0.4	0.6
large	high	0	0	1
medium	low	0.8	0.2	0
medium	medium	0	0.6	0.4
medium	high	0	0.2	0.8
small	low	1	0	0
small	medium	0.4	0.6	0
small	high	0.3	0.6	0.1

Table S6. Elicited probability table used in calibrating investment location typology. The values represent a probability score weighted by area to classify how likely the different investment locations are, along the low, through medium, through high gradients of resource frontier and agglomeration indices. See Fig. 2 in the main text for a graphical representation of the scores.

Resource frontier	Agglomeration economies	Investment location			
		Smallholder populated land	Subsistence frontier	Emerging commercial frontier	Established markets
low	low	1	0	0	0
low	medium	0.2	0	0	0.8
low	high	0	0	0	1
medium	low	0.2	0.8	0	0
medium	medium	0.04	0.16	0.64	0.16
medium	high	0	0	0.8	0.2
high	low	0	1	0	0
high	medium	0	0.2	0.8	0
high	high	0	0	1	0