Food Security: Yield Gap

JM Beddow, TM Hurley, and PG Pardey, University of Minnesota, InSTePP Center, Saint Paul, MN, USA
JM Alston, University of California, Davis, CA, USA
© 2014 Elsevier Inc. All rights reserved.

Glossary

**Allocative efficiency** Occurs when the quantities of variable inputs are chosen such that net value is maximized.

**Allocatively efficient yield** The yield that maximizes the net value of production at a particular location and point in time given existing technology and input constraints.

**Maximum limiting yield** The maximum yield that can be achieved at a particular location and point in time given the best existing technology for that location and time, and nonlimiting inputs other than the amount of land.

**Partial factor productivity** The ratio of output to the amount of a single input (e.g., land) used to produce it.

**Production possibilities frontier** The maximum amount of output that can be produced for each quantity of input.

**Technical efficiency** Occurs when the maximum amount of output is produced for a given amount of inputs.

**Total factor productivity** The ratio of output to an aggregation of all inputs used to produce it.

Introduction

In the past half-century, agricultural science achieved a great deal. From 3.1 billion in 1961, the world’s population more than doubled to just over 7.0 billion in 2012 (an average compound rate of growth of 1.6% per year). Over the same period, total production of cereals grew faster than population (from 877 million MT in 1961 to 2546 million MT in 2012, or approximately 2.1% per year), and this increase resulted largely from unprecedented increases in crop yields (FAOSTAT, 2013; Pardey et al., 2012). (Cereals include wheat, rice, maize, and other agronomic crops that are primarily used for their edible grain.) The fact that the Malthusian nightmare has not been realized in our lifetime is attributable in large part to improvements in agricultural productivity achieved through technological change enabled by investments in agricultural R&D (Alston et al., 2013).

Looking forward, however, some early warning signs suggest that the period of global agricultural abundance may be coming to an end, if it has not ended already. Certainly, after decades of rapid (but slowing) decline, in real (inflation-adjusted) terms farm commodity prices have not trended down appreciably since 2000 (and have recently risen), and in recent times proportional crop yield growth rates have been slowing (Alston et al., 2009). The prospect of continuing population growth and rising per capita incomes, combined with additional demand for biofuels, implies a significant increase in global demand for farm output over the decades to come. These prospects have contributed to a revival of concerns over the future capacity of the world to feed a growing population at affordable prices, giving rise to questions about the prospects for agricultural production and, in particular, crop yields. How and where will this additional output be produced with ever-tightener supplies of arable land and water, and in the face of a changing climate?

The issue of ‘yield gaps’ comes up in this context, in particular. The concept of yield gaps in crops can be applied with equal force to livestock. Productivity or ‘yield’ gains in livestock are often delineated in terms of increases in pounds of meat produced per animal, or milk per cow, or eggs per laying hen. This is akin to tracking the amount of grain produced per plant, which is typically conceived as a harvest index (i.e., the proportion of the plant’s biomass that is realized in grain rather than straw and other plant parts) rather than crop yield per se. A more natural concept of ‘livestock yield’ that parallels crop yields is the amount of meat, eggs, or milk produced per unit of feed or energy or some other input.) One concept of a yield gap is the difference between yields achieved by farmers in high-income countries and their counterparts in low-income countries. Can we bring up the yield of poor farmers in, perhaps, marginal areas, and by closing this yield gap reduce poverty while at the same time addressing the world food problem? (For example, a 2011 New York Times article by Justin Gillis defined the yield gap as “[…]the enormous gulf between the crop yields obtained by the most successful farmers and the least successful,” noting that US maize yields are often five times those of small farmers in Africa. A recent Time magazine article (Walsh, 2012) used the term to describe “…parts of the planet where agricultural yield is lagging,” with reference to “under-performing regions.” This notion of yield gaps is also present in some NGO, government, and scholarly publications, such as reports from the US Department of Agriculture (e.g., Rosen and Shapouri, 2012) and the US Agency for International Development (e.g., Collier, 2012) who use a similar concept, noting the yield gap between grain yields in sub-Saharan Africa and the rest of the world.) Another concept of a yield gap is the difference between the commercial yield achieved by farmers and the experimental yield achieved by researchers. Davidson and Martin (1965), for example, report that average experiment station yields can be as much as 75% higher than average farm yields, depending on the crop and location. Can we close this gap? Should we try to do so? A third type of yield gap is the difference between experimental or commercial yields and some concept of a biological maximum potential yield. Is this gap shrinking in a way that limits our scope for increasing experimental or commercial yields in the future?

Many people are interested in closing yield gaps primarily as they relate to food security – in the sense of producing enough food at affordable prices to meet the caloric and other...
nutritive requirements of a growing population—especially by increasing yields for subsistence farmers. But this is only one dimension of yield gaps, and closing yield gaps for subsistence farmers is only one of the many approaches for improving food security. Indeed, UN projections indicate that in the next 5–10 years the majority of the developing world’s population will be urbanized, shifting attention to urban poverty and the food insecurity status of urban consumers versus subsistence producers (United Nations, 2012). Hence, rather than closing yield gaps by increasing the lowest yields, the more important concept for increasing global food security may be to increase total production by lifting the entire distribution of yields, even if that might mean making gaps larger in some cases.

In this article the issue of yield gaps is explored, primarily from an economic perspective, giving due regard to food security concerns but not at the expense of the broader considerations. Yield is an object of choice for a farmer or a researcher, determined by decisions made about technology, inputs, and management, conditioned by uncontrolled elements in the natural environment. From this perspective, yield gaps reflect differences in choices, as well as other differences in the circumstances of production that cannot be fully controlled, and attention appropriately focuses on the sources of these differences and what they imply. In the Section Economic Conception of Yields and Yield Gaps, some economic concepts are introduced. These concepts are applied to show why yield gaps might exist. The Section Measurement of Yields and Yield Gaps reviews the variety of ways in which researchers have attempted to quantify yields and yield gaps, with particular attention paid to the challenges facing such an exercise. Next, an interpretation of yield gaps in the context of global food security and agricultural science and technology policy is provided. The article ends with some concluding remarks regarding the virtue of framing the potential for global food security in terms of yield gaps.

**Economic Conception of Yields and Yield Gaps**

Crop yield and yield gaps are purely agronomic measures, and provide information on the outcome of a season-long production process. (There are, of course, interseasonal factors that affect yields, such as soil nutrient or water carryover, or the timing of cropping operations.) But crops involve more than agronomy: they are cultivated by farmers who make decisions about what, when, and where to plant, and the quality and quantity of inputs used to foster crop growth. Economics offers a number of conceptual and empirical tools for considering the production process in more detail, allowing one to think not only about yields and yield gaps, but also why yield gaps might exist, and whether it might be advantageous to close them. Thus, an economic perspective can add significant insight and nuance to consideration of yield and yield gaps.

Several concepts are useful for understanding yields and yield gaps from an economic perspective. First is the relationship between what is produced (i.e., the output, such as the amount of maize, wheat, or rice grain), what is used to produce it (i.e., the quantities of inputs, such as land, labor, fertilizer, water from rain or irrigation, solar radiation, pesticides, seed, and services from oxen and tractors), and how it is produced (i.e., the technology of production, which prescribes the different ways inputs can be combined to produce output). This output, input, and technology relationship characterizes the ‘production possibilities.’ Figure 1(a) illustrates hypothetical production possibilities for maize as more or less fertilizer is used holding other inputs constant. (Any other variable input besides fertilizer could be used to illustrate these concepts, e.g., labor, seed, irrigation water, or agricultural chemicals.) The shaded area in the figure reflects all the different combinations of maize and fertilizer that are possible given the available technology. The solid curve denoting the upper boundary of these production possibilities is known as the production possibilities frontier (PPF).

Technically, efficient production occurs where the greatest possible amount of output is produced from a given quantity of inputs or, equivalently, the least possible amount of inputs is used to produce a particular amount of output. Combinations of maize and fertilizer below the PPF (e.g., point a) are technically inefficient because more maize can be produced with the same amount of fertilizer (point b) or the same amount of maize can be produced with less fertilizer (point c). Alternatively, combinations of maize and fertilizer on the PPF...
points b and c are technically efficient because it is impossible to produce more maize with the same amount of fertilizer or the same amount of maize with less fertilizer.

Technical efficiency is one of two types of economic efficiency that are relevant in this context. If production is not technically efficient, then resources are being wasted. But as the PPF in Figure 1(a) illustrates, many (possibly infinite) combinations of maize and fertilizer avoid this wastefulness, and therefore technically efficient farmers or scientists could choose many different input–output combinations. For example, a scientist might set up an experiment to identify the maximum yield for a given maize variety by using an exorbitant amount of fertilizer to produce output at point e. But this ignores the second type of efficiency of interest to economists, allocative efficiency. Allocative efficiency is accomplished when inputs and outputs are combined to achieve their highest possible net value. In standard economics texts, profit is used as the relevant measure of net value, so allocative efficiency implies profit maximization. However, farmers often have objectives other than simple profit maximization. Other benefits and costs that might be viewed as important include food security for subsistence farmers, the risk of adverse health consequences from using pesticides for cash grain farmers, the risk of losses of production to pests and diseases, or business risk arising from market volatilities. Regardless of how a farmer perceives the net value of production activities, the same analytical approach can be used with appropriate adjustments to the measures used.

The solid curve in Figure 1(b) illustrates the maximum net value from maize production for a range of quantities of fertilizer per hectare. Points a–e in this figure correspond to points a–e in Figure 1(a). The highest point on this curve is at point b, which is where production is both technically and allocatively efficient. All other input–output combinations are allocatively inefficient. Net value is lower at point a than at point b because at point b more maize is produced with the same amount of fertilizer; similarly, net value at point a is less than at point c because at point c the same amount of maize is produced with less fertilizer. Point a is technically inefficient, and technical efficiency is a necessary (but not sufficient) condition for allocative efficiency. Points c–e are technically efficient but allocatively inefficient. Points on the PPF to the left of point b (point c) are allocatively inefficient because too little fertilizer is used, whereas points on the PPF to the right of point b (points d and e) are allocatively inefficient because too much fertilizer is used given the relative cost of fertilizer and value of maize. An equivalent way to determine allocatively efficient combinations of maize and fertilizer is to find the point on the PPF where its slope equals the ratio of the unit cost (or price) of fertilizer to the unit value (or price) of maize, which is illustrated in Figure 1(a) at point b.

Maize cannot be produced exclusively with fertilizer. Maize seed, a place to plant the seed and other inputs are also required. The relationship illustrated in Figure 1(a) embodies an assumption that quantities of all other inputs being used to produce maize do not change as the amount of fertilizer changes. This assumption explains why the PPF in Figure 1(a) becomes flatter as the amount of fertilizer increases and eventually begins to decline. Holding the amount of seed and land constant, plant growth can be severely nutrient limited without any fertilizer, so little will be produced. As the amount of fertilizer increases, crop growth becomes more vigorous and more is produced. (This is a hypothetical example. Of course, application of fertilizer does not always increase crop yields. For example, Silesht et al. (2010) found that the addition of enough fertilizer could actually reduce maize yields in some parts of sub-Saharan Africa in some years.) Eventually, with enough fertilizer, growth will no longer be nutrient limited and the amount produced will be as great as possible given other available inputs. Continuing to increase fertilizer beyond this point can become detrimental, leading to a decrease in output. This flattening of the PPF and its eventual decline as the variable input increases is known as diminishing marginal returns – each additional kilogram of fertilizer increases the amount of maize produced by less than what the previous kilogram did. If the amount of seed or land used to produce maize is doubled, the relationship between the amount of maize produced and fertilizer changes, which is illustrated in Figure 1(a) by the higher dashed curve. Similar shifts in the PPF could result from changes in quantities of other inputs, such as labor, or the addition of irrigation or some other change in technology, such as switching from landrace to improved seed varieties. Upward shifts in the PPF could also result from changing where maize is produced to a place with more productive soils or more favorable weather.

Considering these additional inputs, the farmer is free to vary more than one, and potentially all of them. Those inputs that the farmer can freely vary are referred to as variable inputs; those that the farmer cannot freely vary are referred to as fixed inputs. In the ultimate long run, all inputs are variable, but over shorter planning horizons, some inputs are better treated as fixed. Thus, within a season, hired labor, irrigation water, fertilizer, and pesticides are typically regarded as variable inputs, whereas capital inputs (such as the available machinery), land, and its attributes (terrain, rainfall, and solar radiation) are regarded as fixed. In this context, allocative efficiency more generally refers to farmers choosing all variable input quantities, not just the quantity of fertilizer, such that net value is as high as possible. Economists use the net value function to represent the highest net value that could be earned from freely choosing the quantities of all variable inputs. These inputs are chosen in response to the unit cost of inputs and unit value of outputs, and the quantities of fixed inputs. (These ideas apply both to commercial or subsistence farmers. Even if farmers do not directly participate in output (or input) markets, they do make optimizing decisions based on the opportunity cost (or shadow prices) of inputs and outputs. As Schultz (1979) observed in his Nobel Prize lecture: “Farmers the world over, in dealing with costs, returns and risks, are calculating economic agents. Within their small, individual, allocative domain they are fine-tuning entrepreneurs, tuning so subtly that many experts fail to recognize how efficient they are.”)

The net value function shows how the maximum net value varies as farmers choose allocatively efficient combinations of variable inputs and output. As the technology of production is embodied in the net value function, it can be used to derive the allocatively efficient quantity of output given the unit costs of variable inputs, the unit value of output, and the quantities of any fixed inputs. Likewise, the net value function can be
used to derive the demand for land (when it is a variable input) as a function of the same variables. Thus, both the allocatively efficient quantities of output produced and land used can be expressed as functions of the unit cost of inputs and the unit value of output, technology and the quantities of fixed inputs, and therefore the allocatively efficient yield can also be derived as the ratio of these quantities. (In practice, when some production decisions are made farmers do not know the ultimate realization of some variables, such as rainfall, solar radiation, and output prices. For example, the farmer typically does not know the final output price when deciding on the amount of fertilizer to apply or the area to plant. Thus, decisions are based on expectations of likely realizations and distribution of these uncertain variables.)

The important point is that even in the more general multi-input case, the allocatively efficient yield depends on the unit cost of inputs and the unit value of output, the quantities of fixed inputs, technology, and the quantities of other factors over which producers have no choice in a given cropping season, and all these factors vary across farms and over time. As a result, even if farmers are making technically efficient, or even allocatively efficient use of land and other inputs, ‘yield gaps’ could be observed across farms insofar as input unit costs, output unit values, or the quantities of other fixed inputs also vary, dictating different choices. Enduring ‘gaps’ in yields between any two farmers, such as one in the U.S mid-West and one in Malawi, may reflect persistent differences in their constraints and opportunities as indicated by differences in their economic or agroecological circumstances, reflected, for example, in input unit costs, output unit value, pest pressure, access to public infrastructure, or weather.

**Yield as a Productivity Measure**

Yield, defined as the output of a crop (e.g., amount of grain or other crop product) per unit of land used to produce it (e.g., metric tons of maize per hectare), is one of the most common measures of crop productivity. (In Figure 1(a) there is a one-to-one correspondence between the quantity of maize output (as plotted) and maize yields, that is output per unit area, because in this example all other inputs, including land area, are fixed – only the rate of fertilizer used on that given area varies. Similarly, and for the same reason, there is a one-to-one correspondence between the amount of fertilizer used (as plotted) and the rate of fertilizer application, that is the amount of fertilizer used per unit area. Yield is considered a ‘partial productivity’ measure because land is not the only input used in production. Partial productivity could also be measured in terms of crop output per unit of any of the other inputs used to produce it, such as fertilizer, labor, seed, or irrigation. Different partial productivity measures can lead to different conclusions. For example, averaging 2010 production at the continental scale, US farmers produced approximately 9.6 MT ha\(^{-1}\) of maize using 152 kg ha\(^{-1}\) of nitrogen fertilizer and sub-Saharan African (SSA) farmers produced 1.9 MT ha\(^{-1}\) of maize using approximately 9.7 kg ha\(^{-1}\) of nitrogen such that 61 kg of maize per kg of nitrogen was produced in the United States whereas 194 kg of maize per kg of nitrogen was produced in sub-Saharan Africa. (United States values were derived using area and output data from FAOSTAT (2013) along with USDA fertilizer use data (USDA, ERS 2011). SSA values were estimated using area and output data from FAOSTAT (2013), an estimate of average SSA nutrient application on maize (Smale et al., 2011), and applying typical NPK shares from Alexandratos and Bruinsma (2012).) By these measures, US farmers appear more productive than SSA farmers in terms of land, but not as productive in terms of nitrogen, which highlights the perils of naive interpretation of partial productivity measures like yield; namely, productivity with respect to any single input (e.g., land or fertilizer) could easily lead to inaccurate conclusions about the overall productivity of a system.

Furthermore, a difference in crop yield between countries or regions is not, in itself, sufficient for drawing definitive conclusions about productivity. The 2011 average Australian wheat yield (2.0 MT ha\(^{-1}\)) was less than one-third of the national average wheat yield of France (6.5 MT ha\(^{-1}\)), but this difference mainly reflected relatively low land quality and rainfall in the growing environments of Australia rather than any technical or allocative inefficiency in Australian wheat production. (This yield difference might also reflect the influence of differences in incentives for producers provided by European Union subsidies and other policy differences. The 2011 yields are from FAOSTAT (2013).) Even within a country, crop yields differ substantially over space: The 2007 US Census of Agriculture reports that nonirrigated maize yields were 2.5 times higher in Illinois than in Alabama. Although these yield differences can reflect regional differences in the environment, they can also reflect differences in socio-economic circumstances such as opportunities for off-farm employment of farm family members and other dimensions of input costs, educational attainment, family structure, cultural norms, social and public infrastructure and related services, the nature of markets for farm outputs, and other determinants of the unit value of farm output. Such differences exist in market-driven agricultural systems as well as in subsistence agricultural systems.

Ideally, analyses of production, yields, or yield gaps would consider the production from the entire crop rotation of an enterprise or region and not focus on individual crops (Siebert et al., 2010). Additionally, taking account of inputs other than land would allow for a more complete accounting of productivity and the sources of output growth. One way to take account of crop rotations, cropping intensity and multiple inputs is to apply total factor productivity (TFP) metrics to measure the relationship between all outputs and the amount and quality of all inputs used in production, rather than measure yield, which is a partial factor productivity (PFP) metric considering only the ratio of output to the amount of a single input (e.g., land) used to produce it. (Multifactor productivity (MFP) metrics include only some inputs and outputs, whereas TFP metrics ostensibly include all inputs and outputs. Some argue that TFP is a misnomer insofar as it is impossible to capture and properly measure all of the inputs to production (Alston et al., 2010).) But even an increase in TFP does not guarantee that either yield or output has increased. Although TFP is closely related to the notions of efficiency discussed in this article, changes in TFP for a region over time can also reflect changes in the economic scale of
farm operations or changes in technology. Nonetheless, changes in TFP give a more complete accounting of the sources of growth than do changes in yields or yield gaps.

Total output equals total land under production multiplied by the corresponding yield such that increases in production can be thought of in terms of extensification, increasing the amount of land under production, and intensification, increasing the yield per unit of land. (Focusing on output per unit of land puts undue emphasis on land as a limiting factor in economic growth. For example, Schultz (1951) pointed to the declining economic importance of agricultural land; not only because the agricultural share of economic output has declined, but also because “... the value added by land ... declined relative to all inputs used in farming” (p. 735). Although there are prospects for extensification (Pardey et al., 2014), scientists and policymakers often emphasize intensification options as a means to increase crop production because, among other reasons, it is supposed that increasing the land area under production might exact a higher environmental cost than does increasing yields (Garnett et al., 2013). Indeed, approximately 76% of the increase in aggregate crop production worldwide from 1980 to 2010 resulted from intensification, with only 24% attributable to extensification. (This figure is an approximation and is intended for illustrative purposes. The measure of total output is derived by aggregating crops by weight based on data from FAOSTAT (2013).) Thus, it seems, both the historical importance of and policy preference for intensification have led many researchers to focus on the limits to intensification in terms of yield gaps and the potential for closing these gaps as a means for feeding an increasing world population.

Yields, Yield Limits, and Yield Gaps

Because yield measures output per unit of land, it embodies the assumption that production is land constrained, making it subject to diminishing marginal returns and some absolute limit. The gap between this limiting yield and the observed yield bounds how much intensification can potentially increase production. Many conceptualizations and strategies for quantifying the limiting yield have emerged in the yield gap literature with a notable lack of standardization. The quantification of observed yields has also varied. Despite this lack of consistency, key ideas emerging from the literature can be parsimoniously framed from an economic perspective using four quantities:

- Actual yield ($Y^a$) – the measured yield at a particular location and point in time.
- Allocatively efficient yield ($Y^{AE}$) – the yield that maximizes the net value at a particular location and point in time given existing technology and input constraints.
- Limiting yield ($Y^l$) – the maximum yield that can be achieved at a particular location and point in time given existing technology and input constraints.
- Maximum limiting yield ($Y^{ML}$) – the maximum yield that can be achieved at a particular location and point in time given the best existing technology for that location and time, and nonlimiting inputs other than the amount of land.

The terminology used here intentionally differs from that used in previous studies to reduce the confusion that would come from having multiple definitions for the same terms. The terms are also defined relative to the economic concepts previously developed and illustrated in Figure 1 to further promote clarity. These four quantities make it possible to carefully define the yield gap as well as to identify distinct reasons for its existence.

The maximum limiting yield is the most that can be produced assuming land is the only constraining factor, given existing technology. This is illustrated in Figure 2 as $Y^{ML}$ where the dashed curve represents the PPF for fertilizer and maize yield assuming the best available technology of the time is used, and inputs other than fertilizer and land are not limiting yields. The maximum limiting yield occurs where this dashed PPF reaches a maximum because fertilizer is also no longer limiting. It is important to remember that the maximum limiting yield is both location- and time-specific. As new technologies are developed over time the maximum limiting yield will also change, resulting in shifts in the dashed PPF.

The limiting yield, illustrated in Figure 2 as $Y^l$, falls short of this maximum limiting yield because of constraints on technology or inputs other than land. For example, farmers may not have access to pesticides, certain improved seed, or other inputs or technologies, owing to a range of socioeconomic and institutional factors. Therefore, the limiting yield occurs at the maximum of the PPF that reflects these technology and other input constraints as illustrated by the solid curve. An upward shift in this solid PPF can be accomplished by alleviating various constraints, though, by definition, the solid PPF could never exceed the dashed PPF. The difference between the maximum limiting yield and limiting yield results in one type of yield gap: $Y^{ML} − Y^l$.

The actual yield, illustrated by $Y^a$ in Figure 2, typically falls short of the limiting yield. This can occur because input use is technically efficient, but not high enough to achieve the
limiting yield (point a). It can also occur even if enough of the input is used to reach the limiting yield, but input use is technically inefficient (point b), or because not enough of the input is used to reach the limiting yield and what is used is technically inefficient (point c). The difference between the limiting yield and actual yield results in a yield gap: Y\textsuperscript{l} – Y\textsuperscript{a}.

The gap between limiting and actual yields can be decomposed into the gap between limiting and allocatively efficient yields and the gap between allocatively efficient and actual yields: Y\textsuperscript{l} – Y\textsuperscript{a} = (Y\textsuperscript{l} – Y\textsuperscript{AE}) + (Y\textsuperscript{AE} – Y\textsuperscript{a}). The allocatively efficient yield occurs where the slope of the PPF with existing technology and input constraints equals the ratio of the unit cost of fertilizer to the unit value of maize, represented by point d in Figure 2. A gap between limiting and allocatively efficient yields is inevitable whenever inputs are costly to use or farmers have a limited ability to secure the input (e.g., because of poorly functioning credit markets or lack of public infrastructure). For example, if the unit cost of fertilizer in Figure 2 were zero, then the fertilizer-to-maize unit cost-to-value ratio would also be zero, such that the limiting and allocatively efficient yields would coincide (at Y\textsuperscript{l}). A gap between allocatively efficient and actual yields is not inevitable.

Indeed, some economists argue it is nonexistent because the actual yield ultimately reflects the choices of farmers based on their own objectives and perceived values and costs, so it must be allocatively efficient. (That is, Y\textsuperscript{l}, corresponds to the output of maize that maximizes the value of production to farmers (point b in Figure 1(b)). The core of the argument is that any difference between Y\textsuperscript{l} and Y\textsuperscript{AE} results because the analysis has failed to properly incorporate the constraints faced by the farmer, the farmer’s technology or the farmer’s values.) Others argue that a gap between allocatively efficient and actual yields can and does exist for many farmers. Typically, this gap is presumed to be positive as is the case with the difference in yield between points d and a, b, or c illustrated in Figure 2, but this need not be the case if, for example, the observed yield actually corresponded to point e. Moreover, such gaps arise from either allocative inefficiencies (e.g., the difference in yields between points d and a) or technical inefficiencies (e.g., the difference in yields between points d and c or b).

Production at point f in Figure 2 further illustrates the drawbacks of using partial productivity measures like yield and yield gaps as a focus of policy discussions on food security. The yield corresponding to point f is identical to the allocatively efficient yield corresponding to point d. However, production at point f is technically and allocatively inefficient because too much fertilizer is used. Therefore, it is possible for production to be inefficient even if there were no gap between the allocatively efficient and actual yields.

The relationships illustrated in Figure 2 show how the overall gap between the maximum limiting and actual yields can be decomposed into the gaps between maximum limiting and limiting yields, limiting and allocatively efficient yields, and allocatively efficient and actual yields: Y\textsuperscript{ML} – Y\textsuperscript{a} = (Y\textsuperscript{ML} – Y\textsuperscript{l}) + (Y\textsuperscript{l} – Y\textsuperscript{AE}) + (Y\textsuperscript{AE} – Y\textsuperscript{a}). How each of these gaps can be closed and whether it should be closed requires very different policy considerations.

Closing the gap between maximum limiting yield and limiting yield requires addressing the factors that are hindering access to better technology and inputs. In sub-Saharan Africa, for example, in many places the provision and adoption of improved technologies that would allow higher yields and lower costs of production for maize have been hampered by a lack of public infrastructure or by incomplete or nonexistent markets for credit, fertilizer, or other inputs, or by poor intellectual property rights applicable to agricultural technologies. In contrast, in the United States and other high-income countries, the comparable infrastructure, markets, and institutions function comparatively well and market-based incentives are much more effective (although that has not always been so, most notably if US crop yield performance is considered in a long-run historical context). If yields in sub-Saharan Africa could be lifted toward their US counterparts by improving markets for inputs or building better infrastructure, farmer (and consumer) welfare would be improved, but other ways of improving welfare may be more economical.

The gap between limiting and allocatively efficient yields can be closed by policies that reduce the unit cost of inputs or increase the unit value of output to farmers – whether these are measured using market prices or implicitly in terms of the opportunity cost for farmers, such as the value of their time or the value of at-home consumption of their products. But, if the market prices of outputs and inputs accurately capture their social value and costs, policies that interfere in the market may hurt other groups more than they benefit farmers. Thus, such policies might only be warranted if they are aimed at correcting some market distortion. (For example, consider the environmental costs (damages) imposed on society through the use of certain chemicals in agricultural production. If farmers are not liable for such damages, they may use more of the chemicals than they would if they were liable for environmental damages. A policy that lowers the price of such inputs might increase these external costs.)

Closing the gap between allocatively efficient and actual yields, if one exists, requires policies that promote the technical and allocative efficiency of farmers, so their input and output choices are less wasteful and capture as much of the value of production as possible, which could be of benefit to farmers and society as a whole. Such policies might include funding of educational extension programs and dissemination of market information to aid farmers in selecting efficient input mixes, or reducing distortions in market prices and the incentives they convey to farmers, such as, for example, in high-income countries that encourage the overuse of fertilizer and other agricultural chemicals by subsidizing output and failing to tax agricultural pollution.

Measurement of Yields and Yield Gaps

The quantification of yield gaps requires the measurement of a limiting yield as a counterfactual point of comparison for the actual yield. (Crop yields are usually defined in terms of the primary product of the crop. So, for example, maize output is typically measured in tons of grain, not in tons of total biomass produced. Traxler and Byerlee (1993) analyzed the uptake of semidwarf wheat varieties in the context of the demand for the joint (straw and grain) products from growing this crop.) The theoretical maximum yield based on crop physiology independent of realized
genetic improvements attracted interest in the 1960s and 1970s in relation to maximum limiting yields, especially from Dutch researchers, such as de Wit (1967) and Linneman et al. (1979) (as cited in Plucknett, 1995). However, although the theoretical maximum yield is referenced in the yield gap literature, it has not been widely adopted to either conceptualize or quantify the maximum limiting yield. Instead, the notion of potential yield, defined by Evans (1993, p. 26) as “the yield of a cultivar when grown in environments to which it is adapted, with nutrients and water nonlimiting, and with pests, diseases, weeds, lodging and other stresses effectively controlled,” has been used more consistently to conceptualize the notion of a maximum limiting yield. A feature of this definition that distinguishes it from the theoretically maximum yield of a crop species is that it is defined relative to the best adapted cultivar for a particular environment. As land is heterogeneous across space and time in terms of soil and weather, and cultivars have continuously improved and become more highly adapted to specific environments, potential yield is a spatially and temporally varying concept. Although Evans-type definitions are most common, it is not clear that they are actually quantifiable; as Cassman (1999) notes, such potential could not be measured in the field because it is impossible to completely mute all crop stress.

Potential yield as defined by Evans is location- and time-specific, which makes it most appropriate for quantifying yield gaps at the scale of a farm (or even farm field) in a particular year. Other notions of limiting yield for comparison to actual yield have cropped up in studies that attempt to quantify yield gaps at a landscape scale. These landscape studies rely on empirically based definitions of limiting yields, though they still maintain some notion of time- and location-specificity through the use of reference years and climate homologs. Although the empirical nature of these landscape definitions makes them easier to quantify, they are conceptually different from definitions like that of Evans. To better understand some of the key differences between perspectives on yield gaps at the farm level and at broader spatial scales, it is useful to briefly review the different strategies that have been employed in these two strands of literature. (Another way yield gaps have been measured is as the difference in yields between countries or regions (Trueblood and Arnade, 2001; Gillis, 2011; Walsh, 2012; Rosen and Shapouri, 2012; Collier, 2012) or production systems (De Ponti et al., 2012; Yang et al., 2008; Paez, 1973). These strategies for yield gap measurement are distinct from the more common field/farm and landscape scale studies because they ignore the location specificity of the production environment.)

**Farm Yield Gap Measurement**

The early literature on yield gaps relied heavily on measures of yields from intensively managed experiments on farmers’ fields or experiment stations to quantify the maximum limiting yield, though researchers recognized that the limiting yields achievable by farmers in practice would likely differ from these experimental yields because of differences in environmental, socioeconomic, or other constraints faced by farmers, but not researchers (Davidson et al., 1967; Gomez, 1977). (It is common in the yield gap literature to view economics as a ‘constraint.’ This reveals an agronomic point of view in which the counterfactual ideal yield is the limiting yield, even though the limiting yield is unlikely to be economic.) This distinction between the yield achievable in experiments and the yield achievable on the farm led Gomez (1977) and subsequent researchers to define different types of yield gaps; for example, Yield Gap I equal to the difference between the achievable and experimental yields and Yield Gap II equal to the difference between the achievable and actual farm yields. The International Rice Agroeconomic Network, an International Rice Research Institute (IRRI)-led consortium, measured achievable yields by conducting on-farm experiments to reduce the gap associated with potential environmental differences between farms and experiment stations, and thus the achievable yield roughly corresponds to the farm-level limiting yield defined in this article. The early multidisciplinary work at IRRI showed exceptional bio-economic clarity in its framing of yield gaps. This work includes Herdt and Wickham (1975), IRRI (1977 and 1979), De Datta et al. (1978), Herdt and Mandac (1981) and others. Our economic conceptualization of yield gaps draws from and repositions these foundational works.

The lines drawn between limiting and allocatively efficient yields and between allocatively efficient and actual yields are motivated by distinctions like those between Yield Gaps I and II proposed by Gomez (1977). However, these early yield gap distinctions did not always clearly delineate between gaps resulting from technical inefficiencies, rational economic behavior, and more fundamental social, cultural, institutional, or economic constraints, making it difficult to understand the practical and policy implications of alternative gaps.

Herdt (1988), as cited in Pingali and Heisey (1999), found that experiment station yields decreased markedly when the objective of the experiment was changed from yield maximization to profit maximization, which led Pingali and Heisey (1999) to argue that differences between farm yields and experiment station yields designed to maximize profits provide a measure of the “exploitable yield potential” (p. 21). This notion of exploitable yield potential is analogous to allocatively efficient yields, assuming farmers are primarily interested in profit maximization. Alternatively, if farmers are also concerned with risk avoidance as argued by Herdt and Wickham (1975), the analogy between exploitable yield potential and allocatively efficient yield begins to wither.

Duvick and Cassman (1999) assert that contest winning yields provide another and better approximation of farm-level limiting yields, whereas Specht et al. (1999) warned that contest winning yields “... arise from favorable confluences of genotype, management, soil type, rainfall, weather, etc.” (p. 1568) that are unlikely to be scalable across large production areas. Crop growth models offer another way to quantify the maximum limiting yield (e.g., Becker et al., 2003) that has not been as widely adopted because critics have worried about how well such yields capture what farmers can achieve (Lobell et al., 2009), the need to frequently update the models to reliably capture a changing genetic landscape (Fischer and Edmeades, 2010), and the potential for errors and uncertainties in the modeled yields to exceed the actual yield gap (Neumann et al., 2010).
Yield Gap Measurement at Landscape Scales

More recently, the literature has turned to assessments of yield gaps at broader spatial scales than the farm level, including regional and global assessments; these ‘landscape scale’ assessments generally are undertaken with a view to better targeting productivity-enhancing interventions and investments. Although the details of these assessments vary across studies, the general approach is to represent actual yields using global estimates such as those found in Monfreda et al. (2008). In these datasets, the world is partitioned using a grid of pixels (i.e., a raster), with estimates of the actual average yield for each pixel in which the crop is likely to be grown. These global methods also differ in how the limiting yield is estimated for comparison to the actual yield. Using raster data on weather (moisture and temperature), Licker et al. (2010) divided the world into 100 climate zones such that each zone ostensibly has a similar temperature and moisture profile (measured using growing degree days and a soil moisture index). The authors then constructed a cumulative distribution of yields across the constituent pixels in each zone and took the 90th percentile (e.g., the minimum yield of the top ten percent of yielding pixels in each zone) as the limiting yield of the zone given its temperature and moisture profile. (The choice of the 90th percentile by Licker et al. (2010) was apparently somewhat arbitrary, although the authors expressed concern that higher values in the yield dataset might be “erroneous or overestimated” (p. 774). Others have used the 95th percentile and one could argue for use of the maximum yield for a climate zone as the limiting yield for that zone.) The yield gap of each pixel was then calculated as the difference between the limiting yield in the corresponding zone and the estimated actual yield in the pixel. Foley et al. (2011) and Mueller et al. (2012) similarly define the limiting yield for each pixel as the 95th percentile of the yield distribution for pixels with similar rainfall and temperature regimes. It is worth emphasizing that defining limiting yields as an arbitrary percentile of a yield distribution is conceptually different from defining limiting yields based on simulated or experimental yields. As the yields in these distributions are actually observed, they will tend to generate conservatively small estimates of limiting yields. (This is especially true when the yield distribution is made up of pixel-level yields because pixel yields are already a spatial average yield (and in some cases, a temporal average as well).)

The coarse measures of temperature and rainfall used in climate homolog methods do not fully describe the growing environment. Indeed, intraseasonal patterns of rainfall and temperature, and the occurrence of weather events relative to the lifecycle of the crop are extremely important in explaining yields (Beddow et al., 2012). As the methods abstract from the intraseasonal extremes of temperature and drought that generate crop stress, they do not generate estimates of potential yield that are in the spirit of Evans (1993). Further, the binning procedures – whereby pixels are grouped into homologs based entirely on some selected agroecological attributes – do not embed spatially sensitive information on input costs or output values and other socioeconomic factors. Thus, both the limiting and allocatively efficient yields may be different for each pixel even if pixels are in the same climate homologs.

This strand of literature does not tend to employ concepts analogous to the maximum limiting yield, though in some sense the maximum yield observed within the distribution of yields in a climate homolog might be interpreted as such. The distinction of allocatively efficient yields is also not typically quantified. Neumann et al. (2010), who estimate a stochastic production frontier based on global pixelated yield data, provide an exception to some extent. This stochastic production frontier method incorporates monthly temperature deviations, monthly precipitation, solar radiation, and a measure of soil fertility as inputs in an attempt to quantify the source of technical inefficiencies. Therefore, the method does not explicitly quantify allocatively efficient yields, but it does measure part of the gap between allocatively efficient and actual yields. However, because the set of inputs employed by the authors is limited, what appear to be technical inefficiencies may actually reflect differences in unmeasured inputs related to farmers’ management choices or other environmental factors, which could bias the results.

Practical Challenges to Measuring Yield

To quantify yield gaps well requires access to accurate and meaningfully comparable measures of actual yields; whether they are used directly in forming estimates of gaps in yields or indirectly as a basis for calibrating modeled crop yields. However, estimating crop yields that are statistically accurate (or ‘true’ in ISO parlance) and precise is difficult. Survey-based methods are typically used to estimate average yields (which are usually derived from estimates of planted or harvested area and corresponding estimates of grain production) at county, state, or national scales, but the nature, accuracy, and comparability of these surveys vary markedly over time, across countries, and among crops.

For example, in the United States, objective measures of average maize yields at county scales and above begin by establishing an ‘area frame,’ from which a probability-proportional-to-size sampling of farm fields is taken. Two parallel, 15 ft. sections of rows within each sampled maize field are then surveyed by trained enumerators who collect the information required to estimate the number of ears per acre and the (standardized moisture content) grain weight per ear that in turn are used to estimate total crop production and average yields (USDA-NASS, 2012). Remotely sensed data coupled with an array of modeling methods are increasingly being used in conjunction with ground-truthed data to generate (georeferenced) crop area, yield, and production landscapes (Bailey and Boryan, 2010).

In other, often developing-country contexts, crop cut or farmer estimates have been the primary means by which crop production, area, and yields are estimated. (‘Crop cut’ refers to a set of methods for estimating crop quality and yield by harvesting or sampling small portions of fields. Procedures differ, but in general the goal is to scale up the yield of the sampled area to estimate a yield for the entire field or farm.) These estimates are prone to substantial sampling biases and measurement error. This is especially so in the complex
cropping systems characterizing many smallholder operations in low-income countries. The inability to assure a spatially representative sampling of farms, perhaps stratified by size, production systems (e.g., irrigated vs. rainfed), and other relevant attributes, is just the first of many sources of measurement error. Smallholder farms are often subdivided into plots, wherein crops are grown in pure stands as well as multi- or intercropped systems involving two or more crops planted on the same plot in a season, and in many instances involving at least two and in places three cropping seasons per year (Dalrymple, 1971; Poate and Casley, 1985; Fermont and Benson, 2011). In these circumstances, subplot crop cutting methods are often used to estimate yields (and production), introducing a host of potential estimation errors. For example, Poate and Casley (1985) suggest crop cutting methods tend to overestimate yield via a combination of ‘edge effects’ (plants that lie fractionally outside the subplot are included in the count), ‘border bias’ (location methods may over- or underestimate the boundaries of a plot), and nonrandom location of the subplot (enumerators tend to avoid bare or sparsely populated parts of the plot) (Also see FAO, (1982) and Murphy et al., (1991)). The timing of the crop cuts matters too; especially for cropping systems in which individual plants within a particular crop stand, or different crops within each (sub-)plot, mature at different times.

Crop cuts generate estimates of the quantity of production, which carry with them problems of standardization by form – fresh or processed, shelled or unshelled, with or without stalk, polished or unpolished, in the cob or seeds only, or some other form – standardization by moisture content, and standardization by type (e.g., Durum vs. winter or spring wheat). To report yields (or to upscale crop-cut or farmer estimates of plot-level yields to estimate total production) at larger spatial scales also requires commensurate measures of planted or harvested area. A host of conventional area measurement techniques are used (e.g., Poate and Casley, 1985; Poate, 1988), increasingly complemented by handheld PDA devices (Keita et al., 2010). None of these area estimation approaches is without error. Newer GPS aided devices are imprecise on very small plots (say less than 0.01 acres), but are more accurate on larger plot sizes, where they reveal a seemingly systematic tendency for farmers to overestimate the size of small plots (by upwards of 90%) and underestimate the size of large plots (by upwards of 59%) (Carletto et al., 2013).

The comparability of yields taken at a point (say on-station yields, or on-farm experimental yields taken from one or a few farms) versus the estimated average yield in the surrounding or adjacent region is also problematic. Crop yields are typically much more variable at the field level than at the regional level (e.g., Lobell et al., 2007) and farm yields are often not symmetrically distributed even within a field (e.g., Hurley et al., 2004), so that a regional average yield may not represent the yield of a typical farm, and thus the calculated yield gap may not be representative of the region being studied. That point estimates of limiting yield (e.g., from an experiment station or highest yielding farm) are essentially incomparable with regional or national average yields was noted by Barker (1979), who deemed such comparisons “... newsworthy (but scientifically worthless)...” (p. 10).

### Interpretation of Yield Gaps

Meaningfully measuring the magnitude of a yield gap is one challenge. Another challenge is drawing actionable insights with meaningful scientific, farm management, economic, or policy implications.

### Are Yields Gaps Good, Bad, or Irrelevant?

Two somewhat contradictory views of the importance of yield gaps to food security might be taken. One view is that yield gaps are necessary for improvements in farm yield, and as the yield gap decreases, because the actual yield is approaching the limiting yield, it is increasingly difficult for farmers to improve yields such that yield growth rates slow (Grassini et al., 2011; Cassman, 1999; Pingali and Heisey, 1999; Cassman et al., 2003). The natural conclusion is that small yield gaps are a bad omen for future food prospects, and that a robust way to address food security concerns is to increase yield gaps by raising the limiting yield through breeding programs for example. However, the evidence to support a relationship between smaller yield gaps and slower yield growth is not entirely clear (Evans and Fischer, 1999).

A second view is that yield gaps are indicative of potential problems that could be addressed by changing farmer behavior, natural environments or markets. From this viewpoint, ‘closing’ the yield gap is a way forward to sustainably increase agricultural output (Mueller et al., 2012; Foley et al., 2011). In this context, yield gaps show where output gains might be realized by tackling the biotic, abiotic, and socioeconomic constraints to yields. For example, Herdt and Wickham (1975) first discuss constraints that lead to the gap and then suggest research and investment as a way to relax these constraints. Still, whereas yield gaps might show where improvement is technologically possible, this does not necessarily imply improvement is economically or socially feasible or desirable.

Closing yield gaps is not a necessary condition for increasing output via agricultural intensification. A historical example from the United States serves to illustrate. The improvement in maize output over the past century in the United States is an example of a successful intensification effort, especially from the perspective of increasing global food and feed supplies. Consider the period 1889 through 1954. In the earlier years conditions in US agriculture were comparable to current conditions in many low-income regions, with smaller farms, poor rural infrastructure, comparatively limited use of off-farm inputs, and so on (e.g., Gardner, 2002). But the average maize yield in the United States grew significantly. It increased from just over 1.8 MT ha$^{-1}$ in 1899 to approximately 2.5 MT ha$^{-1}$ in 1954. Over the entire period 1889–2007, the average US maize yield increased to 9.3 MT ha$^{-1}$ such that output sextupled even though the amount of US harvested land devoted to maize increased by only approximately 20%. Using county-level annual yields, the authors approximated the methods of the global yield gap studies (e.g., Foley et al., 2011 and Mueller et al., 2012) and measured yield gaps as the difference between the yield for a particular county in a particular year and the 95th percentile of the yield distribution for that year. (These yield gaps (G) were calculated...
for each county, i, and year, t, as $G_{i,t} = \max (Y^L_{i,t} - Y^A_{i,t}, 0)$, where $Y^L_{i,t}$ is the 95th percentile of the county yield distribution for time t and $Y^A_{i,t}$ is the reported (‘actual’) yield for the corresponding county and year. The area-weighted mean gap, $\sum G_{i,t} a_{i,t} / \sum a_{i,t}$ was used as the national yield gap, which was expressed as a percentage of $Y^L_{t}$. It is worth noting that slight variations in the way in which the yield gap is calculated or presented can change the results. One could argue effectively for use of unweighted, area-weighted or output-weighted yield gaps, gaps as a percentage of farm yield, calculating the limiting yield using various percentiles of the yield distribution, and so on. In the present data, the absolute yield gap increased between 1889 and 2007, whereas the yield gap expressed as a percentage of $Y^L_{t}$ decreased (from ~31% to 19%). The gap decreased substantially more (from 80% to 29%) when expressed as a percentage of farm yields.) Although the national yield improved markedly during both periods, this measure of yield gaps increased between 1889 and 1954, and decreased between 1954 and 2007.

More particularly, increases in regional yields do not necessarily imply decreases in the region’s yield gap. Figure 3 shows the spatial distribution of US maize yields for 1889, 1954, and 2007, along with the cutoff for the limiting yield (estimated as the 95th percentile of the county yield distribution). The average yield increased between 1889 and 1954, but the larger yield increases tended to be seen at the high end of the distribution, thus increasing the national yield gap. By 2007, the distribution had become much more symmetric, decreasing the yield gap. Thus, changes in the distribution of yields can affect yield gaps, even causing the yield gap to increase while average yields are improving. Indeed, a region’s yield might be distributed such that the yield gap could change

between two years even if the same amount of output were produced in both years using the same amount of land (and thus the regional yield would not change). (Many distributions exhibit this phenomenon. For example, consider a simple case in which, in a given year, half of a country’s hectares had a yield of 7 MT ha$^{-1}$ and the other half yielded 8 MT ha$^{-1}$. The national yield in that year was 7.5 MT ha$^{-1}$. Suppose that the country’s yield distribution changed the following year, such that half of the hectares yielded 5 MT ha$^{-1}$ and the other half yielded 10 MT ha$^{-1}$. Again, the national average yield was 7.5 MT ha$^{-1}$. Using the 95th percentile of the yield distribution as the limiting yield, the first time period had a yield gap of 6.25%, whereas the second year had a yield gap of 25% even though the same amount of output was produced on the same amount of land.)

### Lack of Standardization and Potential for Manipulation

Yield gaps often entail the comparison of some observed yield with a counterfactual limiting yield. The preceding discussion noted several ways in which the limiting yield might be estimated. Because the measure of the limiting yield can differ across studies, ostensibly similar ‘yield gap’ studies may actually be measuring very different things. Thus, care must be taken in interpreting yield gaps – namely, it is important to understand exactly what is being measured. It has been shown empirically that different methods of estimating the limiting yield can produce markedly different results. For example, Singh et al. (2009) estimate state-level Indian yield gaps for several crops using both simulated yields and experimental yields as alternative measures of the limiting yield. In most

![Figure 3](image-url)  
**Figure 3** County-level maize yield distribution, 1889, 1954, and 2007. The limiting yield for each year was calculated as the 95th percentile of the unweighted county-level yields for all counties that harvested maize in the corresponding year. This limiting yield is shown as a vertical line for each distribution (with the value indicated at the top), in a color corresponding to the color of the distribution. Created by the authors using data from Beddow, J.M., 2012. A Bio-Economic Assessment of the Spatial Dynamics of U.S. Corn Production and Yields. Ph.D. Dissertation. Department of Applied Economics, University of Minnesota.
cases, the limiting yield implied by crop simulation models was markedly higher than the limiting yield implied by experimental results. Aggarwal et al. (2008) present a similar study including yield gaps derived using three methods to estimate the limiting yield: simulation models, experiments, and on-farm demonstrations.

Lack of standardization in the measurement of yields must also be carefully considered in the interpretation of yield gaps. This is particularly true when comparing yield gaps across farm versus landscape analyses, but is also true for comparisons across farm-level analyses or across landscape-level analyses. For example, global analyses of yield gaps rely on aggregate estimates of yields over geo-politically defined statistical regions. The mere process of this aggregation will tend to mute the spatial variability of yields and the size of yield gaps when compared with more disaggregate measures such as those constructed based on farm- or field-level yield observations. This can be illustrated using pixelated estimates of global maize yields on a 5 arc min (approximately 10 × 10 km) grid. (The production, area and yield data used for this illustration come from the Spatial Production Allocation Model 2000 v3.0.6 (You et al. 2000).) Table 1 reports descriptive statistics for the global distribution of yield gaps calculated in two ways. The first distribution is based on the difference between the 95th percentile yield estimate for all pixelated yields within an agroecological zone (data for the agroecological zones are from Sebastian (2006)) and the pixelated yield estimate, which reflects yield gaps derived from disaggregate data and is to some extent analogous to using farm- or field-level yield estimates in a landscape-scale study of yield gaps. (Yield gap distributions were calculated using area weights as described in the Section Are Yields Gaps Good, Bad or Irrelevant?) The second distribution is based on the difference between the 95th percentile yield estimate for the country average yields in an agroecological zone and the country average yield for a pixel, which reflects yield gaps derived using methods analogous to those typically employed in landscape scale assessments of yield gaps. With the more disaggregated yield estimates, the mean, median, maximum, and standard deviation of the yield gap distribution are 65.2, 148, 44.7, and 8.5% higher, respectively, than with the country aggregated data. Therefore, simple differences in spatial aggregation across studies can result in noncomparable yield gap estimates.

The lack of standardization in the measurement of actual and limiting yields can lead to some confusion, particularly in the policy arena. Sumberg (2012) argues strongly that the yield gaps used in policy arguments are often specified such that they are as large as possible or that they justify the analyst’s preferred policies, concluding with a warning that yield gaps are “seldom what they appear” (p. 517).

### Yield Gaps with Multiple Enterprises

Diversification is a common strategy in farming used to manage risk and to better utilize lumpy assets and farm family labor resources for which different farm enterprises have different seasonal demands. It is accomplished by growing multiple crops, raising livestock in addition to crops, or being employed off as well as on the farm. With multiple enterprises, the allocation of scarce inputs to one enterprise inevitably affects the output in other enterprises. Van Ittersum et al. (2013) note that agronomically optimal sowing dates might not be economically optimal when multiple crops are planted each year (e.g., the optimal harvest date of one crop might conflict with the optimal planting date of a follow-on crop). Von Braun (1988) demonstrated that there may be complex interactions in multicrop systems, such that adoption of a yield-increasing technology for one crop may decrease yields of other crops Von Braun (1988) reported on the effects of an improved rice production scheme in The Gambia, finding that, on average, 390 kg of cereals and 400 kg of groundnuts are foregone for each additional ton of rice. This result was influenced both by the expected substitution of labor into the now more productive rice production along with a complex household and farming structure. Von Braun warns that “[i]n the more complex the household structures and production organization in agriculture are, the less straightforward the predictions are on how technological change may impact on nutritional improvement (p. 1095).” The importance of such tradeoffs is not easily accounted for when focusing exclusively on individual crop yields and yield gaps.

Figure 4 illustrates how yield gaps emerge from a farmer’s choice of activities. In the illustrated case, a farmer is allocating his (limited) time between maize production and off-farm work that pays an hourly wage. (The off-farm production example used here could easily be replaced by the production of an additional on-farm output, such as a second (or more) crop(s), with no change in the yield-gap implications of the example.) The shaded area shows the production possibilities for maize yields and off-farm work, given existing technology

<table>
<thead>
<tr>
<th>Disaggregated pixelated yield gaps by AEZ (MT ha⁻¹)</th>
<th>Country aggregated yield gaps by AEZ (MT ha⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.81</td>
</tr>
<tr>
<td>Median</td>
<td>3.38</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.58</td>
</tr>
<tr>
<td>Maximum</td>
<td>13.27</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Table 1** Descriptive statistics for disaggregated pixelated and aggregated country average yield gaps by agroecological zones (AEZ)

*Note: Disaggregated pixelated yield gaps were constructed as the difference between the 95th percentile of pixelated yields in an agroecological zone (AEZ) and the pixelated yield. Country aggregated yield gaps were constructed as the difference between the 95th percentile of the country average yield in an agroecological zone and the average yield of each country’s constituent pixels.*

and input constraints. The upper boundary of this region denoted by the solid curve is the PPF. Point a is inefficient because more maize can be produced without decreasing off-farm work (and thus, off-farm income) or more off-farm income could be acquired without decreasing maize production. Allocative efficiency requires equality between the value of an additional ton of maize and the value of the off-farm work that would be given up to produce the additional maize, which occurs where the slope of the PPF equals the negative of the ratio of the wage to the unit value of maize (point b). The negative sign reflects the fact that some maize must be given up to do more off-farm work when production is technically efficient.

The limiting yield, $Y_L$, occurs at point c where off-farm work is at a minimum so the farmer can devote as much time as necessary to the production of maize to maximize yield. The actual yield, $Y_A$, could be technically, but not allocatively, efficient (point d). Alternatively, it could be technically and allocatively inefficient (point a). The allocatively efficient yield corresponding to point b maximizes the sum of the value of maize production and off-farm work. The gap between the limiting and allocatively efficient yield is once again inevitable if off-farm work is valuable to the farmer, so trying to close it is not economical. Thus, yield gaps can emerge from economical choices over what to produce (e.g., maize or off-farm work) as well as over how it is produced (e.g., with more or less fertilizer). The gap between the allocatively efficient and actual yield is technically or allocatively inefficient because what the farmer is choosing to produce is wasteful given the value of maize and the wage rate.

Conclusion

Making progress on food security requires more than a biological view. Yields are lower or higher depending on the unit costs of inputs, the unit value of output, and other socioeconomic constraints faced by producers. An important economic constraint, particularly in low-income countries, is the inadequate state of transportation infrastructure. Poor road and rail networks increase the cost of getting output to markets and inputs to farms, and thus change the relative unit costs of inputs and the unit values of output at the farm gate, with (often negative) consequences for the allocatively efficient yield. Other constraints also hinder farmers: poor storage facilities require that output be sold rather than stored when market prices are low; trade restrictions can reduce domestic prices; and missing or poorly functioning markets for credit, insurance and labor all serve to reduce a farmer’s allocatively efficient yield.

Yield gaps are the wrong metric if the primary concern is food security, which is fundamentally about the capacity of agriculture to supply affordable food efficiently and sustainably to subsistence farmers and nonfarm consumers alike. Closing yield gaps is neither necessary nor sufficient for improving food security. Indeed, as demonstrated above, a region’s yield and output could improve even though yield gaps are increasing. Further, targeting the closure of yield gaps as a policy in order to feed more people may have unintended consequences. For example, idling either less-productive land (thus decreasing the gap directly) or idling the most productive land (thus reducing the limiting yield) would close the yield gap, but would also reduce total output. Output is the primary determinant of the availability and affordability of food, and efforts to increase output will tend to improve the lot of nonfarmers as well as farmers. Thus, in many cases, policies might best be focused on increasing the output of a region or the world as a whole – such as investments that enhance farmer productivity and participation in markets – rather than on simply decreasing yield gaps, which is not guaranteed to improve food security and often will be uneconomic.

References


Relevant Website

http://www.InSTePP.umn.edu

International Science and Technology Practice and Policy Center, University of Minnesota.